CERLIS
Centro di Ricerca sui Linguaggi Specialistici
Research Centre on Languages for Specific Purposes

Cécile Desoutter, Dorothee Heller
& Michele Sala (eds)

Corpora in specialized communication
Korpora in der Fachkommunikation
Les corpus dans la communication spécialisée

4
CELSB

CERLIS Series

Series Editor: Maurizio Gotti

ISSN 2532-2559

Editorial Board

Ulisse Belotti
Maria Vittoria Calvi
Luisa Chierichetti
Cécile Desoutter
Marina Dossena
Giovanni Garofalo
Davide Simone Giannoni
Dorothee Heller
Stefania Maci
Michele Sala

Each volume of the series is subjected to a double peer-reviewing process.

CERLIS Series
Volume 4

Cécile Desoutter, Dorothee Heller & Michele Sala (eds)

Corpora in specialized communication
Korpora in der Fachkommunikation
Les corpus dans la communication spécialisée

CELSB
Bergamo

This ebook is published in Open Access under a Creative Commons License Attribution-Noncommercial-No Derivative Works (CC BY-NC-ND 3.0).
You are free to share - copy, distribute and transmit - the work under the following conditions:
You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
You may not use this work for commercial purposes.
You may not alter, transform, or build upon this work.

CERLIS SERIES Vol. 4

CERLIS
Centro di Ricerca sui Linguaggi Specialistici
Research Centre on Languages for Specific Purposes
University of Bergamo
www.unibg.it/cerlis

CORPORA IN SPECIALIZED COMMUNICATION
KORPORA IN DER FACHKOMMUNIKATION
LES CORPUS DANS LA COMMUNICATION SPÉCIALISÉE
Cécile Desoutter, Dorothee Heller & Michele Sala (eds)
ISBN 978-88-89804-25-4

© CELSB 2013
Published in Italy by CELSB Libreria Universitaria
Via Pignolo, 113 - 24121, Bergamo, Italy

Indice

MICHELE SALA / DOROTHEE HELLER / CÉCILE DESOUTTER
Introduzione ……...11

I corpora in contesti accademici

ALESSANDRA MOLINO
1. Compiling a Stratified Corpus for a Cross-cultural Study of Academic Writing: Methodological Challenges and Research Opportunities ……..27

PATRIZIA ANESA
2. Avoiding Plagiarism and Self-plagiarism through the Use of Corpora ……..55

GABRIELLA CAROBBIO / DOROTHEE HELLER / CLAUDIA DI MAIO
3. Zur Verwendung von Frageformulierungen im Korpus *euroWiss* …….....75

ANDREA ABEL / AIVARS GLAZNIEKS
4. „Ich weiß zwar nicht, was mich noch erwartet, doch...“ – Der Einsatz von Korpora zur Analyse textspezifischer Konstruktionen des konzessiven Argumentierens bei Schreibnovizen …….....101

I corpora in contesti pedagogici

DENISE MILIZIA
5. Phrasal Verbs and Phrasal Units: Political Corpora within the Walls of the Classroom …….……135

CARMEN ARGONDIZZO / ASSUNTA CARUSO / IDA RUFFOLO
6. The Use of Specialised Corpora: From Research to Pedagogy ……...165

ALESSANDRA LOMBARDI / SILVIA MOLETTA
7. Von der Hochschule in die Berufswelt und wieder zurück. Berufsbezogene Korpusarbeit im Unterricht *Deutsch als Fachsprache* …….....189

NATACHA S.A. NIEMANTS
8. L'utilisation de corpus d'entretiens cliniques (français / italien) dans la didactique de l'interprétation en milieu médical ….....209

I corpora in contesti legali

MARCELLO SOFFRITTI
9. Konjunktiv in deutschsprachigen Gesetzbüchern …….....239

DORIS HÖHMANN
10. Zur Untersuchung erweiterter Nominalgruppen mit Hilfe von Concgrams. Eine sprachvergleichende Studie zum deutschen und italienischen Umweltrecht …….....267

CHIARA PREITE / SILVIA CACCHIANI
11. Traduire la normativité dans les arrêts de la Cour de Justice de l'Union européenne : le cas des dispositifs en français et anglais …….....297

MARIE-PIERRE ESCOUBAS-BENVENISTE
12. Predicati giuridici e schemi argomentali nelle sentenze della Corte. Approccio bilingue francese-italiano ……..323

I corpora in contesti professionali

MICAELA ROSSI
13. Définition de nouvelles terminologies et communautés de professionnels : analyse de corpus en ligne dans le domaine de la dégustation du vin …….....359

ERIK CASTELLO
14. Exploring Existential and Locative Constructions in a Learner and in an Expert corpus of Promotional Tourist Texts ……...385

EUGENIA DAL FOVO
15. The Language of Interpreters on Television: Characteristics, Tendencies And Idiosyncrasies ……....411

CÉCILE DESOUTTER
16. La prise en compte linguistique des femmes dans les discours électoraux : une étude sur corpus …….....435

DANIO MALDUSSI
17. Anisomorphisme et relation de converse à l'épreuve des corpus spécialisés : le couple "créance"/ "crédit" par opposition à "credito" …….....465

MICHELE SALA / DOROTHEE HELLER / CÉCILE DESOUTTER

Introduzione

I contributi di questo volume raccolgono una parte degli interventi presentati al convegno CERLIS (*I corpora nella ricerca e nella didattica dei linguaggi specialistica*) tenutosi il 21-22 giugno 2012 presso l'Università degli Studi di Bergamo. La prospettiva plurilingue adottata dal convegno è rispecchiata dalla composizione del presente volume, che propone ricerche che testimoniano la crescente rilevanza assunta dai corpora per lo studio e per la trasmissione di saperi utili alla comunicazione specialistica. In questa sede si adotta il termine corpus nella sua accezione più ampia e cioè riferito a ogni raccolta sistematica di testi scritti, trascritti da situazioni orali, o registrazioni al contempo autentici e rappresentativi della lingua usata in ambiti particolari e leggibili elettronicamente. L'analisi di tali corpora, appositamente realizzati, permette, anche con il supporto di strumenti messi a disposizione negli ultimi anni dalla linguistica dei corpora, di riconoscere, localizzare e computare vari aspetti della lingua. Queste collezioni di materiali sono diventate nel corso dei decenni, a partire dagli anni 70 del secolo scorso, risorse insostituibili per la ricerca accademica, sia quella di matrice deduttiva che induttiva: nel primo caso, la disponibilità di materiali autentici permette di testare ipotesi e verificare intuizioni e, nel secondo, consente di identificare particolarità linguistiche e frequenze d'uso che altrimenti rimarrebbero difficili da individuare.

In anni più recenti i corpora hanno ricoperto un ruolo sempre maggiore anche in ambito didattico, e vengono sempre più spesso utilizzati in contesti pedagogici, durante corsi di lingua o di linguaggi professionali perché gli studenti abbiano modo di osservare frequenze, funzioni testuali e aspetti pragmatici di collocazioni, nessi fraseologici e costrutti sintattici o, per converso, l'idiosincrasia di altre strutture pur consentite dalla grammatica e dall'uso. Tramite simili risorse il ricercatore può facilmente approfondire l'analisi degli aspetti sintattici

e pragmatici della comunicazione specializzata, ad esempio verificando il nesso tra determinate occorrenze e il contesto della loro collocazione e, al tempo stesso, lo studente ha modo di consolidare la propria conoscenza e competenza relativa alla lingua in uso piuttosto che quella descritta (o prescritta) nei manuali. Entrambi questi aspetti, ricerca e didattica con supporto dei corpora, sono presi in considerazione in questo volume, che è distinto in quattro sezioni, ognuna delle quali finalizzata ad esaminare l'utilizzo dei corpora in contesti particolari come quello accademico, quello pedagogico, quello legale e quello professionale.

Sezione 1. I corpora in contesti accademici

Il primo capitolo, 'Compiling a Stratified Corpus for the Cross-Cultural Study of Academic Writing: Methodological Challenges and Research Opportunities' di ALESSANDRA MOLINO, apre la raccolta esaminando un'importante questione di metodo nella compilazione dei corpora, ossia il campionamento stratificato. Negli studi di retorica contrastiva, la compilazione di un corpus bilanciato è un requisito metodologico fondamentale per garantire comparabilità e il controllo delle variabili contestuali che possono influenzare l'uso della lingua; nel caso del discorso accademico, queste variabili corrispondono all'ambito disciplinare, al genere, alla macrostruttura testuale, ai destinatari e al livello di esperienza dello scrivente. La compilazione di tali corpora rappresenta tuttavia una sfida metodologica quando le lingue a confronto non hanno la stessa diffusione e lo stesso peso all'interno di un determinato ambito, come è il caso dell'inglese e dell'italiano quali lingue della comunicazione scientifica e della pubblicazione accademica. Un corpus bilanciato necessita quindi dell'identificazione di tecniche di campionamento che ne permettano un efficace impiego per analisi di carattere contrastivo. A questo scopo Molino propone il campionamento stratificato anziché quello casuale semplice, poiché la stratificazione (uniforme) consente di raccogliere campioni simili per dimensioni a seconda del fattore contestuale che si intende controllare. In tal modo è possibile limitare il rischio che un fattore prevalga su un altro per un difetto di campionatura piuttosto che per la sua effettiva occorrenza. L'obiettivo

di questo capitolo è verificare se e in quali termini la tecnica di campionamento stratificato sia implementabile nel confronto tra lingue con frequenze d'uso diverse tramite la costruzione di un corpus di saggi accademici in inglese e in italiano nell'ambito della linguistica, all'interno dei quali viene preso in esame l'uso di forme inclusive come il pronome personale soggetto *we* in inglese e le prime persone plurali delle forme verbali in italiano.

Nel secondo capitolo, 'Avoiding Plagiarism and Self-plagiarism through the Use of Corpora', PATRIZIA ANESA tratta una delle applicazioni più pratiche dei corpora, ossia il loro utilizzo al fine di verificare casi di plagio, un fenomeno che è divenuto sempre più complesso, soprattutto in tempi recenti a causa dell'immediata disponibilità di ipertesto offerta da Internet. Focalizzandosi sui testi accademici, lo studio offre una classificazione delle diverse forme di plagio (intenzionale vs. non intenzionale; esterno o auto-plagio; esatto, quasi-esatto o dissimulato) e si concentra sulla discussione dei principali metodi di rilevazione (parametri stilometrici per analisi di tipo endoforico, cioè interna al testo in esame; comparazione per analisi esoforiche, ovvero tra più testi), tramite strumenti quali raccolte di testi, archivi, banche dati e il web oppure tramite la compilazione di corpora creati *ad hoc*. Lo studio presenta poi i possibili usi e le funzionalità di *PlagiarismFinder* (*PF*), un programma specificamente sviluppato per la rilevazione esterna sulla base di raccolte di testi appositamente create. Questo strumento permette di identificare e localizzare porzioni di testo o nessi di parole che ricorrono *verbatim* tra un determinato testo e il corpus di controllo. Tramite una simile procedura, quindi, l'uso di corpora di riferimento permette non solo di individuare casi di plagio da parte di studenti nell'elaborazione di testi scritti, ma è di estrema utilità anche al ricercatore nella stesura dei propri testi, al fine di rilevare e controllare eventuali casi di auto-plagio, auto-citazioni involontarie e non segnalate come tali non tanto per superficialità o disattenzione, ma spesso dovute a fenomeni di criptomnesia, per cui un ricordo (o, in questo caso, la replica di un concetto e spesso della sua esatta rappresentazione linguistica) viene percepito come un creazione originale e come tale testualizzato.

Il discorso accademico tedesco è oggetto di due contributi che ne prendono in considerazione sia le convenzioni di scrittura che la

dimensione orale, anche in una prospettiva comparativa. Aspetti di didattica accademica vengono presi in esame da GABRIELLA CAROBBIO, DOROTHEE HELLER e CLAUDIA DI MAIO nel capitolo 'Zur Verwendung von Frageformulierungen im Korpus *euroWiss*'. Le autrici illustrano gli esiti di un'analisi condotta nell'ambito di un progetto internazionale di ricerca (*euroWiss*) incentrato su un confronto delle modalità di insegnamento universitario osservabili in contesti paragonabili. Dalla disanima del corpus di videografie di lezioni registrate in atenei italiani e tedeschi emerge il ruolo centrale dell'utilizzo di domande da parte del docente, domande che spesso non vengono formulate per sollecitare una risposta da parte degli studenti. L'analisi, di tipo qualitativo, si basa sulle categorie sviluppate nell'ambito della pragmatica funzionale. I risultati mettono a fuoco le funzionalità di tali domande in contesti didattici caratterizzati da monologicità, quali la lezione frontale, nonché le interrelazioni tra questo genere di domande con altre forme illocutive finalizzate a guidare l'attenzione dei partecipanti alle lezioni e a promuovere l'elaborazione del sapere trasmesso. Tale tipologia di domande viene utilizzata dal docente per esempio per introdurre spiegazioni, ulteriori delucidazioni e fasi riassuntive, agevolando così la comprensione degli studenti. Dal punto di vista comparatistico, l'analisi evidenzia inoltre come il corpus tedesco sia caratterizzato da *Forschungsfragen*, ossia dalla formulazione di domande di ricerca rilevanti per l'ambito scientifico-disciplinare – un aspetto che nel corpus italiano sembra essere meno evidente.

Il contributo di ANDREA ABEL e AIVARS GLAZNIEKS, 'Ich weiß zwar nicht, was mich noch erwartet, doch ... " – Der Einsatz von Korpora zur Analyse textspezifischer Konstruktionen des konzessiven Argumentierens bei Schreibnovizen', sposta l'attenzione sull'interfaccia scuola-università per presentare gli esiti di un'analisi pilota che si inserisce in un progetto di più ampie dimensioni. L'indagine muove dal presupposto che la scuola abbia un ruolo centrale per preparare gli studenti alla scrittura (pre-)accademica. Da qui la scelta degli autori di esaminare un corpus di testi prodotti da studenti (tedeschi, austriaci e altoatesini) per mettere in luce le loro competenze argomentative. Particolare attenzione viene rivolta alle modalità di espressione della concessività in considerazione del ruolo rilevante che essa assume nel testo scientifico. L'indagine viene condotta con il supporto del

software *CQP* e considera in particolare l'uso di costruzioni avversative (*zwar* + elementi correlati) e l'uso di marcature modali (quali *freilich, natürlich, sicher, sicherlich, zweifellos*), ma include altresì un'analisi di tipo mesostrutturale di passaggi argomentativi in contesto. L'analisi evidenzia le possibilità ma anche i limiti di un approccio puramente quantitativo. Si rende pertanto indispensabile un'integrazione di tipo qualitativo per consentire un'osservazione più approfondita dei dati ricavati dall'analisi sistematica del corpus preso in esame.

Sezione 2. I corpora in contesti pedagogici

La seconda sezione si apre col capitolo 'Phrasal Verbs and Phrasal Units: Political Corpora within the Walls of the Classroom' di DENISE MILIZIA nel quale l'autrice prospetta la possibilità di un efficace insegnamento di peculiarità linguistiche e fraseologiche tramite il supporto di corpora. Nello specifico, l'insegnamento del quale tratta il capitolo è rivolto a studenti della facoltà di Scienze Politiche dell'Università di Bari e il corpus raccoglie discorsi, dichiarazioni, conferenze stampa e interviste rilasciate da politici americani e italiani al governo e all'opposizione. L'attenzione è concentrata su quelle unità composte da più parole, in particolar modo sui *phrasal verbs*, ossia quei verbi composti da più elementi il cui significato non si riduce alla sintesi di quelli delle parti costituenti. Lo studio inizia con l'analisi dei verbi lessicali più frequenti in politica, per proseguire poi con i verbi composti da due, tre e quattro elementi. Il riscontro reso possibile dal materiale autentico consente di verificare − tramite il supporto di software progettati per la lettura di corpora (*WordSmith Tools*; *ConcGram*) − le frequenze d'uso e le collocazioni oltre che di misurare la preferenza accordata dai parlanti nativi a espressioni frasali piuttosto che a forme sinonimiche costituite da un solo verbo (ad esempio, nel corpus di riferimento sono state trovate 43 frequenze di *turned down the Lisbon Treaty* contro solo otto occorrenze di *rejected the Lisbon Treaty*). Tale supporto offre anche agli studenti la possibilità di misurare l'efficacia e l'impatto dei verbi frasali all'interno del proprio cotesto d'occorrenza, a dissipare eventuali timori circa la difficoltà nel comprenderli e assimilarli, oltre che ad

evidenziare il loro funzionamento cognitivo come unità olistiche di senso piuttosto che frammentate nelle varie parti che li compongono.

Il secondo capitolo della sezione, 'The Use of Specialised Corpora: From Research to Pedagogy' di CARMEN ARGONDIZZO, ASSUNTA CARUSO e IDA RUFFOLO, rappresenta una coerente prosecuzione del tipo di approccio proposto da Milizia, con una particolare enfasi data agli aspetti pragmatici della lingua. Nello specifico, la ricerca si focalizza sull'analisi di due tipi di corpora specialistici, uno di testi turistici promozionali (*TPT corpus*) e un secondo di discorsi politici e articoli di giornale riguardanti il processo di pace del Medio Oriente (*Roadmap corpus*). Il primo aspetto della ricerca, tramite l'analisi del *TPT corpus* che raccoglie esempi di linguaggio ambientale, turistico e pubblicitario, è quello di esaminare la frequenza del termine *nature* e, tramite l'uso di software che permettono di analizzare collocazione e cotesto, osservare qual è la funzione pragmatica e la valenza semantica attribuita a questo referente (estetica, ecologica, economica, turistica, etc.). La seconda parte del capitolo mira invece a dimostrare come il processo di pace del Medio Oriente possa essere interpretato in modo diverso dalle diverse parti politiche coinvolte sulla base di diversi fattori socio-politici e delle differenti ideologie. La ricerca condotta e le attività ad essa correlate dimostrano come i corpora possano essere risorse inestimabili sia per il docente che per lo studente, e strumenti di elaborazione di materiali che aiutano non solo a sviluppare consapevolezza linguistica, ma formano gli studenti a una lettura critica dei testi tramite la presa di coscienza delle possibili discrasie tra ciò che l'autore 'dice' e quello che 'sottintende'.

In seguito alle varie riforme succedutesi dopo il processo di Bologna, l'università dedica una maggiore attenzione ad una prospettiva professionalizzante della didattica affiancata da una trasmissione mirata di competenze nei linguaggi specialistici. In quest'ottica ALESSANDRA LOMBARDI e SILVIA MOLETTA, nel capitolo **'**Von der Hochschule in die Berufswelt und wieder zurück. Berufsbezogene Korpusarbeit im Unterricht Deutsch als Fachsprache', illustrano come si può integrare la didattica in aula con un lavoro su corpora finalizzato ad acquisire le conoscenze linguistiche settoriali che spesso vengono richieste agli studenti in fase di tirocinio

formativo. Le riflessioni delle autrici si sviluppano sullo sfondo di un'indagine sistematica di reports elaborati dagli studenti nell'ambito dei loro tirocini curriculari che ha consentito una rilevazione dei compiti assegnati ai tirocinanti e alle competenze linguistiche correlate da cui è emerso un notevole *gap* tra le richieste da parte delle aziende e la preparazione degli studenti. Da qui ha preso avvio il progetto *Fachsprachen lernen mit Korpora*, promosso dall'Università Cattolica di Brescia con lo scopo di mettere gli studenti in grado di comporre un corpus funzionale *ad hoc* e di utilizzarlo quale supporto per affrontare compiti che si presentano con regolarità già nelle prime esperienze lavorative, ad esempio la traduzione di testi settoriali. Il contributo riporta l'esperienza di una studentessa-lavoratrice che ha applicato le conoscenze e il metodo acquisiti nell'ambito delle attività didattiche correlate al progetto compilando un corpus parallelo (tedesco-italiano) di testi pertinenti al linguaggio contrattuale.

La sezione si chiude con il capitolo di NATACHA NIEMANTS 'L'utilisation de corpus d'entretiens cliniques (français/italien) dans la didactique de l'interprétation en milieu médical'. L'autrice parte dal presupposto che il crescente numero di pazienti stranieri aventi accesso ai servizi sanitari pubblici rende cruciale la partecipazione degli interpreti ai colloqui clinici. Si basa poi sull'analisi conversazionale di corpora di colloqui, reali e simulati, per dimostrare che la pratica professionale è spesso scollegata dai principi della formazione universitaria dei futuri interpreti. Considera in particolare che il *role playing* non è in grado, da solo, di prepararli alle attività e alle responsabilità della vita reale. Nel suo contributo, l'autrice presenta in una prima parte i due corpora alla base delle sue osservazioni. L'uno è costituito da colloqui simulati in cui due docenti e uno studente interpretano rispettivamente il ruolo del medico italofono, del paziente francofono e dell'interprete che li aiuta a capirsi reciprocamente. L'altro è composto da colloqui clinici reali, in cui veri medici comunicano con veri pazienti in ospedali belgi e italiani. Niemants evoca le scelte con le quali lo studioso deve confrontarsi nella trascrizione delle interazioni e presenta il software EXMARaLDA che permette di ottenere sia uno *spoken corpus* (trascrizioni di dati audio o video), sia un corpus multimodale (*speech corpus*) nel quale trascrizione, file audio e video fanno parte del

corpus stesso. In una seconda parte, espone alcuni risultati della ricerca sui due corpora, puntualizzando come l'analisi qualitativa di colloqui reali e simulati possa arricchire la didattica dell'interpretazione in contesto medico, incoraggiando il lettore a riflettere sulle possibilità e i limiti delle simulazioni.

Sezione 3. I corpora in contesti legali

La sezione si apre con due contributi sul tedesco giuridico, che si collocano nell'ambito della riflessione sulle caratteristiche del discorso giuridico-normativo, anche in ottica comparativa, e sulle scelte metodologiche per un'analisi (semi)automatica di corpora di grandi dimensioni.

Tali questioni riguardanti l'interrogazione dei corpora vengono tematizzate da MARCELLO SOFFRITTI ('Konjunktiv in deutsch-sprachigen Gesetzbüchern') in merito all'utilizzo del congiuntivo nel discorso legislativo. Si tratta di uno studio pilota che muove dall'assunto che il *congiuntivo irrealis* ricorra nelle parti non performative del testo normativo e si propone di esplorare combinazioni e funzioni del *Konjunktiv* nei codici e ordinamenti in lingua tedesca. L'analisi offre un confronto tra testi di diversa provenienza (austriaca, svizzera, tedesca e altoatesina). Per rilevare i contesti e le circostanze d'uso, così come le varie restrizioni, l'autore ricorre ad una *cluster analysis*, individuando contestualmente le situazioni in cui gli strumenti di un'analisi (semi)automatica necessitano di un'integrazione manuale. Dall'analisi delle collocazioni ricorrenti e dei *cluster* emergono importanti indicazioni da mettere in relazione con l'ipotesi di partenza, anche alla luce del fatto che l'uso del congiuntivo nel contesto legislativo rappresenta ancora un tema poco esplorato dalla linguistica giuridica. Le considerazioni conclusive sottolineano la necessità di confrontare i risultati ricavati dall'analisi di corpora disciplinari con dati provenienti da corpora di grandi dimensioni, ma non necessariamente di carattere specialistico.

Il lavoro di DORIS HÖHMANN ('Zur Untersuchung erweiterter Nominalgruppen mit Hilfe von Concgrams. Eine sprachvergleichende Studie zum deutschen und italienischen Umweltrecht') è incentrato sui gruppi nominali, ricorrenti nel linguaggio giuridico e spesso

caratterizzati da notevole complessità. Il corpus preso in esame è costituito da norme di diritto ambientale in Italia e in Germania. Indagando i testi in merito all'estensione di gruppi nominali, l'autrice evidenzia che gli elementi che formano delle unità da un punto di vista semantico spesso non si presentano in una successione lineare, ma sono diversamente dislocati da un punto di visto sintattico. Questo vale in particolare per il tedesco giuridico. Il lavoro mette in risalto i vantaggi di un'analisi di *ConcGrams* che considera gli elementi di contesto e le possibilità di combinazione tra i costituenti di unità lessicali complesse. L'ausilio di questo strumento di analisi consente non solo di mettere in evidenza le interrelazioni tra elementi contestuali e le variazioni linguistiche presenti nei gruppi nominali presi in esame ma anche di individuare sottili sfaccettature dei riferimenti semantici e sintattici. Si evince quindi che uno studio *corpus based* non può limitarsi ad un'indagine quantitativa, ma deve essere affiancato da un'analisi qualitativa.

Gli altri due capitoli di questa sezione vertono sulle sentenze della Corte europea. CHIARA PREITE e SILVIA CACCHIANI, con 'Traduire la normativité dans les arrêts de la Cour de Justice de l'Union européenne : le cas des dispositifs en français et anglais', affrontano il tema della rete deontica. Il loro contributo si apre con una panoramica delle nozioni legate alla performatività del linguaggio giuridico e alla modalità deontica, sia nella filosofia del linguaggio e del diritto, sia nella linguistica. In seguito, le due autrici individuano, in un corpus di sentenze della Corte di Giustizia dell'Unione europea, gli elementi che veicolano questa normatività in lingua francese (lingua di procedura) per poi osservare la loro resa linguistica nelle traduzioni in inglese. Questo censimento permette di evidenziare – in entrambe le lingue – la grande varietà di realizzazione linguistica dell'imperatività nelle sentenze della Corte, le quali si distinguono da altre sentenze per il loro potere sia decisionale sia obbligatorio. Lo studio contrastivo delle due versioni linguistiche fa inoltre risaltare il parallelismo o al contrario la divergenza espressiva tra inglese e francese.

Anche MARIE-PIERRE ESCOUBAS-BENVENISTE ('Predicati giuridici e schemi argomentali nelle sentenze della Corte. Approccio bilingue francese-italiano') parte da un corpus bilingue di sentenze della Corte di Giustizia dell'Unione Europea, allo scopo di indagare le

relazioni di sinonimia di unità lessicali complesse: i predicati giuridici. Il corpus bilingue è costituito, da una parte, da sentenze per le quali la lingua di procedura è l'italiano e, dall'altra parte, dalla versione delle stesse in francese, lingua di lavoro della Corte. L'autrice intende dimostrare quanto la descrizione della predicazione per mezzo di schemi argomentali e di classi di oggetti predicativi sia funzionale per una descrizione contrastiva delle forme della lingua delle sentenze. Questo approccio oltre ad offrire un metodo di risoluzione razionale dei fenomeni di polisemia linguistica dei termini, rende possibile l'identificazione degli usi predicativi sinonimici nell'altra lingua e delle condizioni linguistiche di tale sinonimia. Il metodo adottato permette di conseguenza di apportare un complemento lessicografico fondamentale alla descrizione delle unità linguistiche della lingua giuridica, nonché di integrare la descrizione dei predicati della lingua generale e della lingua specialistica. Secondo l'autrice, una tale descrizione risulta indispensabile per lo sviluppo di una lessicografia terminologica, la quale condiziona a sua volta l'attuazione di una didattica ragionata della lingua e della traduzione in ambito giuridico.

Sezione 4. I corpora in contesti professionali

La sezione si apre con un contributo di MICAELA ROSSI sulla terminologia usata nella degustazione del vino: 'Définition de nouvelles terminologies et communautés de professionnels: analyse de corpus en ligne dans le domaine de la dégustation du vin'. Il lavoro si basa su un corpus composto da interazioni (in italiano o in francese) provenienti da forum specializzati nell'ambito della degustazione del vino. Il capitolo analizza in particolare i commenti metalinguistici e metaterminologici degli utenti. Il discorso dei professionisti sulla terminologia che loro stessi usano permette di fatto di verificare la tracciabilità dei processi di creazione e di implementazione dei termini. In linea con gli studi sulla definizione naturale, Rossi esamina quindi nel suo corpus i commenti relativi alla definizione dei termini e dei concetti specializzati, concentrandosi in particolare sulla fase di negoziazione collettiva del senso e dei termini nelle comunità professionali osservate. Questo studio permette di evidenziare in particolare l'intensa attività metalinguistica degli utenti dei siti

specializzati in degustazione del vino. I loro scambi sui forum rivelano un loro bisogno di normalizzazione in materia di terminologia enologica oltre che una coscienza metalinguistica e un desiderio di *parlare* del vino e di esprime le proprie sensazioni con le parole giuste. La microanalisi del corpus selezionato evidenzia così i limiti di un approccio strettamente quantitativo, fondato sulle statistiche delle occorrenze. Inoltre conferma che il termine è un "prodotto sociale" e che occorre, di conseguenza, definire la terminologia sulla base di un accordo tra specialisti.

Il capitolo 'Exploring Existential and Locative Constructions in a Learner and in an Expert corpus of Promotional Tourist Texts' di ERIK CASTELLO si propone di analizzare le caratteristiche delle costruzioni esistenziali e locative usate in un corpus di testi turistici promozionali in lingua inglese (circa 150.000 parole) ad opera di esperti di madrelingua inglese e un *learner corpus* (circa 40.000 parole) composto di testi scritti da studenti universitari italiani con la finalità di promuovere la città di Padova. Dopo una puntuale rassegna di recenti studi condotti sia in Italia che all'estero sulle caratteristiche morfosintattiche, lessicali, fraseologiche e testuali dei testi turistici promozionali − le quali sembrano fare della lingua in essi utilizzata un linguaggio specialistico − viene fatta seguire un'introduzione teorica alle costruzioni esistenziali in inglese e la discussione dettagliata dei risultati dell'analisi contrastiva dei due corpora. Come prevedibile, tali costruzioni vengono usate frequentemente sia nei testi turistici professionali che in quelli degli apprendenti, ma le loro realizzazioni lessico-grammaticali risultano essere differenti. É emersa, in particolare, la presenza di almeno sei *pattern* esistenziali e locativi ricorrenti che contraddistinguono i due corpora. I risultati dell'analisi qui proposta non hanno solo implicazioni per l'insegnamento della scrittura dei testi turistici ma, indirettamente, anche per la definizione puntuale di un importante aspetto lessico-grammaticale distintivo delle pubblicazioni turistiche, ovvero l'espressione linguistica dell'esistenza di luoghi e attrazioni e della loro localizzazione nello spazio.

Il capitolo successivo, 'The Language of Interpreters on Television: Characteristics, Tendencies and Idiosyncrasies' di EUGENIA DAL FOVO, tratta l'utilità di corpora di linguaggio televisivo al fine di stabilire se e in quali termini è possibile parlare di linguaggio

della traduzione. La televisione è uno dei mezzi di comunicazione che maggiormente influenza la società ed il suo sviluppo. Gli aspetti relazionali e le complesse strutture partecipative e organizzative che sottendono i meccanismi propri della televisione rendono il testo televisivo molto più che un semplice copione, delineando i tratti di un'attività linguistica che è prima di tutto pratica del discorso. Il comportamento discorsivo e traduttivo degli interpreti che operano in televisione – o di chi se ne assume il ruolo (es. conduttore, giornalista, ecc.) – e la valutazione delle rispettive *performance* in termini qualitativi tracciano i contorni di un modo di usare la lingua proprio dell'ambiente televisivo, che a sua volta genera aspettative da parte degli utenti/spettatori e norme (più o meno) prescrittive relative non soltanto al comportamento traduttivo dell'interprete, ma della sua figura professionale in generale. Tali variabili mutano a seconda delle modalità interazionali e dei modelli discorsivi alla base di una determinata trasmissione. Con questi assunti il capitolo si propone di studiare le specificità della lingua prodotta dagli interpreti in televisione indipendentemente dalla tematica trattata – sicuramente più vicina al concetto di lingua specialistica. Le peculiarità linguistiche vengono poi analizzate contrastivamente grazie all'utilizzo di corpora comparabili costituiti da item omogenei per tipologia di interazione e modelli discorsivi ma prodotti direttamente in italiano in situazioni simili, allo scopo di far emergere tratti caratterizzanti in base ai quali si può legittimamente ipotizzare l'esistenza di una lingua dell'interpretazione.

Nel penultimo capitolo, 'La prise en compte linguistique des femmes dans les discours électoraux', CÉCILE DESOUTTER indaga su quanto e come le donne vengono prese linguisticamente in considerazione in alcuni discorsi elettorali. In questo contesto, l'analisi di corpus permette di evidenziare la varietà degli usi del francese nei limiti permessi dalla cultura e dalla lingua stessa. Di fatto, mentre le autorità amministrative in Svizzera e in Canada promuovono l'uso di formulazioni non sessiste (femminilizzazione o neutralizzazione del discorso), in Francia, al contrario, l'*Académie Française* ha sempre ribadito che il maschile funge anche da genere non marcato, permettendo quindi di riferirsi tanto alle donne quanto agli uomini. Il contributo indaga il discorso elettorale con lo scopo di

osservare se, pur in mancanza di direttive redazionali, vengano attivate anche in Francia strategie volte a evitare l'uso del maschile generico. Due corpora vengono presi in considerazione: le dichiarazioni elettorali dei candidati e delle candidate alle elezioni presidenziali del 2012 e i volantini elettorali pubblicati dai sindacati in occasione delle elezioni ai *Conseils de Prud'hommes* nel 2008. L'approccio accosta osservazioni e riflessioni di carattere sociolinguistico e linguistico. In particolare, vengono evidenziate e analizzate in entrambi i corpora le formulazioni che portano alla specializzazione o alla neutralizzazione dei generi, e la loro presenza in funzione del sesso o dell'appartenenza politica di chi si candida. Vengono inoltre evidenziati gli effetti della femminilizzazione sintattica sulla coesione e la coerenza dei testi presi in considerazione. L'eterogeneità delle marche della femminilizzazione e la loro presenza saltuaria, anche nello stesso testo, portano l'autrice a considerare che, nel contesto studiato, le formulazioni non discriminatorie tendono a costituire una strategia più che un sistema.

La sezione si chiude con il contributo di DANIO MALDUSSI ('Anisomorphisme et relation de converse à l'épreuve des corpus spécialisés: le couple "créance / crédit" par opposition à "credito"'). Tramite due corpora paragonabili e rappresentativi in francese e in italiano, il capitolo affronta la questione di un anisomorfismo tipico della lingua dell'economia e della finanza: la coppia di parole francesi *créance / crédit* in opposizione alla parola italiana "credito". Il contesto preso in considerazione è quello del sovraindebitamento del consumatore (*surendettement des ménages*). Per l'autore del contributo, il valore aggiunto dell'approccio *bottom-up* guidato dal corpus risiede sia nella possibilità di identificare e analizzare strategie di produzione del testo diverse tra una lingua e l'altra, sia nella possibilità di individuare strategie di traduzione adeguate al testo. Nel caso specifico, Maldussi dimostra che l'analisi su corpora permette inoltre di arricchire la comprensione del rapporto di antonimia tra *créance* e *dette* presentato dai dizionari monolingui tradizionali e di precisare il perimetro della relazione tra *crédit* e *dette*. Nel contesto preso in considerazione, l'analisi *corpus driven* rende inoltre più accessibile il funzionamento di una pratica finanziaria di grande attualità quale la ristrutturazione dei crediti, una tra le misure adottate

oggi per fare fronte al problema sempre crescente e preoccupante del sovraindebitamento del consumatore.

I corpora in contesti accademici

ALESSANDRA MOLINO

1. Compiling a Stratified Corpus for the Cross-cultural Study of Academic Writing: Methodological Challenges and Research Opportunities

1. Introduction

This chapter explores a crucial methodological issue in cross-cultural research: the challenge of comparability. Comparability has always been a highly debated topic in Contrastive Rhetoric (CR), particularly when, in the second half of the 1980s, the focus moved from the analysis of second language writing to comparisons of authentic professional texts in English and in other languages (Leki 1991). As this approach developed, researchers became increasingly aware of the necessity to compare instances of the same text type across languages. Thanks to studies in text and genre analysis, which provided sounder bases to identify equivalent texts, and the contribution of corpus linguistics, which allowed researchers to collect corpora for quantitative descriptive analyses, CR moved significantly beyond anecdotal evidence or speculative conclusions on culturally-determined writing features (Connor 2002).

In order to make CR research more rigorous and consistent, Mauranen (2001) and Yakhontova (2006) have argued for an approach to rhetorical differences aiming at providing empirical evidence of culture-specific patterns rather than seeking explanations for dissimilarities. Such a focus on quantitative descriptive analysis preserves the issue of comparability at the top of the agenda of cross-cultural research. Among the challenges that need to be faced are those of corpus design and the control of significant contextual

variables. As Moreno (2008: 34-37) observes, we need to control contextual factors statistically by carefully designing the corpus for analysis and choosing an effective approach to sampling. Moreno suggests adopting stratified sampling, as this technique enables researchers to control the impact of extra-linguistic factors that may interfere with the variable of writing culture (e.g. topic, text superstructure, gender).

The aim of this chapter is to test the feasibility of stratified sampling in a special situation in cross-cultural research: the comparison of two languages with unequal chances of being used for the same purposes, such as English and Italian for research and publication. Given the predominant role of English as an international language for academic communication, this situation characterises other European contexts too, where local languages risk losing domain to English (see for instance Pérez-Llantada *et al.* (2011) for the Spanish academic context).

One of the main difficulties that may hinder the use of stratified sampling in such contexts is the different availability of data to investigate relevant contextual factors and compile maximally similar corpora. This chapter addresses this issue by trying to answer the following questions:

- Can stratified sampling be applied to the collection of corpora from target populations of different size?
- What are the practical constraints that influence ideal corpus design?
- How can those difficulties be overcome?
- What consequences do possible compromises on corpus design have on the validity of results?
- Is quantitative descriptive research still meaningful when maximal comparability is difficult to achieve?

In order to answer these questions, the technique of stratified sampling will be applied to the collection of a corpus of Linguistics research articles (RAs) in English and Italian. The difficulties encountered during the compilation will be discussed so as to examine the degree to which stratification is possible in this context. The usefulness of

corpus stratification will be illustrated through a case study of 'reader engagement' (Hyland 2005a) as signalled by the use of inclusive authorial references in subject position.

This chapter is organized as follows. By resorting to Connor/ Moreno's (2005) notion of *tertium comparationis*, Section 2 discusses the importance of identifying appropriate criteria of comparison during the choice of primary data for analysis. The technique of stratified sampling will be explained in sub-section 2.1. Section 3 describes the comparable corpus of Linguistics RAs in English and Italian and comments on the practical constraints that affected the design of the sampling frame. Section 4 presents the results of the study of reader engagement. Section 5 concludes commenting briefly on the issues of sampling, balance and comparability.

2. Contrastive research and the notion of *tertium comparationis*

The aim of quantitative descriptive cross-cultural investigations is to assess the impact of the independent variable 'writing culture' on the dependent variable 'form and content of the text' (Moreno 2008). The label 'writing culture' refers to the set of norms, values and common practices of a particular community, which are expressed through a given language code. Cross-cultural investigations need appropriate parameters of comparison, or *tertia comparationis* (Connor/Moreno 2005), a requirement that is particularly important when looking for differences, as these become apparent only if matched against a background of likeness being shared cross-culturally.

Connor/Moreno (2005) argue that *tertia comparationis* should be placed at two major phases: 1) when choosing the primary data for comparison and 2) when establishing comparable textual concepts. The choice of primary data, which is the focus of this chapter, involves the identification of the target population and the parameters of the study. In cross-cultural investigations, the parameters to include

in the sampling frame are those extra-linguistic factors that Moreno (2008) calls 'confounding factors' because, if not controlled during the sampling, they may interfere with the effect of the main independent variable under study, i.e. language as a vehicle of culture.

In descriptive studies of academic discourse as well as in cross-cultural analyses of academic writing, the principal contextual factors that have been identified are genre (Swales 1990), discipline (Hyland 2000, Hyland/Bondi 2006), textual macrostructure (Tarone *et al.* 1998), date of publication (Salager-Meyer *et al.* 2003) and size of the audience (Burgess 2002). Moreno (2008: 34) argues that "if the design is able to maintain constant the values of […] confounding factors and manages to include the same proportion of texts representing those values in each sample, we can then say that the two corpora are equivalent to the maximum degree of similarity".

2.1. Stratified sampling techniques

An important issue characterising the debate over corpus design is the nature of the relationship between the variables in the study (Hunston 2008: 157). In other words, corpus designers have to decide whether each sub-corpus should contain a balanced number of texts, or tokens, or whether the amount of text material in each sub-corpus should be proportionate to its significance in the target population. Generally, it is the purpose of the study that determines the choice.

In a corpus designed to be representative, the function of variables is to increase coverage; therefore, the corpus designer may wish not only to collect a large amount of data, but also to allocate text material in proportion to its significance in the target population. For example, in Italy, in the macro area of the Humanities, about 22% of publications are journal articles, 17% is constituted by book chapters, 60% by books and the remaining 1% is constituted by other

minor research products.[1] In a corpus aiming at being representative of this specific target population, sub-corpora should mirror the above proportions. The procedure best suited to this frame is stratified random sampling with proportional allocation. Using this technique, texts are chosen randomly from each sub-group, or *stratum*, but the amount of text material reflects its real distribution in the target population. The advantage of this technique is that it increases representativeness. However, it may not be the best choice when the aim is to compare data sets, as the sub-corpora may be significantly different in size. Therefore, in a corpus designed for research with a very specific focus, the compiler may decide to privilege the attribute of balance over representativeness for the sake of comparability.

When balance is an important goal, a uniform allocation of the sample is advisable rather than its proportional distribution. With uniform allocation, each stratum is assigned the same number of texts irrespective of the size of strata in the target population. Once having established the parameters of the sampling frame, the texts to include in the corpus may be collected randomly or using other techniques.[2] The advantage of uniform allocation is that comparisons between corpora within individual strata are possible, because uniform allocation ensures that no contextual factor predominates over the others. In addition, comparisons across strata within each corpus can also be carried out to assess whether and to what extent different contextual factors influence language use.

---

1 These data were provided by the Comitato Italiano di Valutazione della Ricerca (CIVR, National Committee for the Evaluation of Research) for the period 2001-2003.

2 Dörnyei (2007) observes that in most Applied Linguistics research, probability sampling procedures, such as simple random sampling, are rarely fully achieved. In fact, research in this area often employs non-probability techniques, such as 'quota sampling', which is similar to stratified random sampling "without the 'random' element" (Dörnyei 2007: 98). The first step is usually the design of the sampling frame, followed by the definition of the proportions of the various sub-groups. The samples are then collected, but not necessarily adopting strict random sampling procedures. Rather, compilers meet the quotas by selecting participants they can have access to.

The disadvantage of uniform allocation is that it reduces representativeness, as the smaller strata in the population are promoted and the larger strata are demoted. A corpus designed for the purpose of comparison, therefore, does not provide data reflecting the actual distribution of language features in the target population.

3. Corpus design

This section describes the sampling frame of the comparable corpus of Linguistics RAs in English and Italian used in this study. Table 1, which is inspired in its display by the example provided in Connor/Moreno (2005: 159), illustrates the sampling frame; Table 2 provides quantitative information about the corpus.

Tertium comparationis	*Value of prototypical feature perceived as constant across corpora*	*Number of texts in each independent corpus*
Genre	Research article	32
Discipline	Linguistics	32
Text superstructure	Introduction – Method – Results – Discussion	16
	More variable superstructure (logico-argumentative papers)	16
Gender	Women	16
	Men	16

Table 1. Sampling frame for genre, discipline, text superstructure and gender.

Each independent language corpus is composed of 32 texts, meaning that the corpus as a whole contains 64 RAs. The first parameter is that of genre: only original research papers were collected, excluding other related genres such as research reports or review articles. The reason is that different overall discourse goals may affect rhetorical and linguistic configurations.

	Number of texts	*Total number of words*	*Average text length*
Linguistics RAs in English	32	215,461	6,733
Linguistics RAs in Italian	32	223,229	6,976

Table 2. Comparable corpus of Linguistics RAs in English and Italian.

The second constant feature is discipline. The influence of discipline on discourse practices has been documented in numerous studies (e.g. Hyland 2000, Hyland/Bondi 2006). Therefore, only papers that can be described using the umbrella term of 'Linguistics RAs' were collected. However, the sample was not stratified in terms of sub-discipline (e.g. sociolinguistics, applied linguistics, pragmatics) or topic.

The third *tertium comparationis* is text superstructure. This parameter refers to the overall organization of discourse in macro-rhetorical units. Text superstructure may have a considerable impact on language use. In this sampling frame, RAs were grouped according to two sub-groups. One sub-corpus collects 16 papers conforming to the so-called IMRAD structure (Swales 1990). The IMRAD structure is typical of articles featuring an Introduction, a Method section, and Results and Discussion sections (conflated together or in separate units). All the papers in this sub-corpus present a clearly identifiable section reporting on the methods and procedures, and an equally unmistakable section presenting the research findings. However, the methodologies are varied, ranging from corpus analysis to fieldwork. It can be hypothesised that different types of approach entail different rhetorical strategies in reporting procedures. This aspect will have to be considered when analysing results. The second sub-corpus includes papers with a more mixed configuration. They are predominantly logico-argumentative in nature, following macro patterns such as the Problem-Solution(-Evaluation) pattern or the Situation-Analysis pattern.

Finally, I also controlled the individual factor of gender. The relationship between language and gender has received considerable

attention in the study of language variation. Language has been described as an important means by which gender is enacted (see Ehrlich 2004). For this reason, I included this feature in the design of the corpus, collecting 16 RAs written by women and 16 written by men.

Having described the composition of the comparable corpus, I now turn to some remarks on the difficulties encountered during the sampling. The first complication was the comparatively limited availability of data in Italian. There are various reasons for such an imbalance. The first is that the number of specialized journals in Italian is much more restricted. A cursory glance at lists of publications such as the ERIH (European Reference Index for the Humanities) Initial List of Linguistics journals (2007) will prove ample evidence for this phenomenon.[3] Scrolling down the list, I counted 246 journals with an English title, 58 journals with a French title, 47 journals with a German title, 44 publications with a Spanish title, 25 journals with an Italian title and 10 with a Portuguese title. Due to the different population size, there may be some relevant contextual factors, such as sub-discipline or topic, that do not recur frequently enough in the smallest population to be able to collect an adequately large sample of texts.

In the case of the present corpus of Linguistics RAs, the very dependent variable of language was at risk. It has become common practice for Italian journals to accept papers in other languages (mainly in English). As a result, in numerous latest issues, articles in Italian are rare. My original plan was to collect papers reflecting the current use of language by taking texts from 2008 to the present. However, given the constraint of the sampling frame, this was not possible. Therefore I decided to extend the time span of the study and to go back to 2001 to collect the desired amount of texts.

Although going further back in time may be a solution, the disadvantage is that over a long period of time, rhetorical conventions

---

3 The 2007 list and the 2011 list with revised journal categories are available at: <https://www2.esf.org/asp/ERIH/Foreword/search.asp> (last accessed 26 June 2013).

may change, thus turning the time coordinate into another significant extra-linguistic influence. In order to partially avoid this risk, in the present corpus the bulk of articles was taken from the 2003-2008 period (see Appendix). In so doing, the isolated cases of more recent or more dated works can be compared to the general tendencies of the corpus to verify whether they diverge considerably, thus skewing overall counts.

A second complication related to the dependent variable of language is that while in Italian journals the population of scholars writing in Italian is rather homogenous, as almost all of them are Italian native speakers, the same cannot be said for international journals, where scholars from a variety of first language backgrounds publish in English. Ideally, the value of native-speaker status should be kept constant. In studies of academic writing, the name and affiliation of the writer have often been considered as parameters to control this factor (Hewings *et al*. 2010, Molino 2010). The same was done in the present study. However, in order to respect the sampling frame, some exceptions were made. In the Italian corpus, one article is written by an Italian native speaker based in the UK; on the other hand, in the English corpus, while all articles are written by authors based in Anglo-American universities, some of them are signed by authors who may well not be native speakers of English, judging from their name. In those cases, before including the paper in the corpus, the scholars' academic profile was checked on the Internet and when a mostly Anglo-American profile emerged, the article was included. The assumption is that English being these scholars' professional language, their written productions are likely to encode the values and expectations of the Anglo-American academic writing culture. Nevertheless, special attention should be paid to the quantitative data coming from such texts to verify whether they diverge significantly from the rest of the corpus.

4. A cross-cultural study of reader engagement in English and Italian Linguistics RAs

This section presents a cross-cultural study of reader engagement as signalled by the use of inclusive *we* and *noi* [we] in subject position. I will refer to these engagement devices as 'inclusive authorial references' because while in English the subject is always expressed, in Italian, person and number are generally signalled through verb endings.

4.1. Previous research on reader engagement

In recent years, studies on academic discourse have focused on its interactive nature demonstrating that the often-claimed neutrality of academic writing is not borne out of empirical evidence. Writers construct for themselves appropriate academic *personae* and demonstrate awareness of their audience by considering their readers' expectations, anticipating their objections and providing metadiscourse guidance (Hyland 2005a). Hyland (2005a) refers to the devices used to recognize the presence of readers as 'engagement markers'. The most direct engagement strategy is the use of reader pronouns (*you* and *your*). However, as Hyland notes, these forms are less typical in academic writing than in other more informal registers. In academic texts, readers and writers are often bound together through the use of inclusive *we*, whose overall function is to claim solidarity and membership (Hyland 2005b: 364).

Reader engagement has often been studied in terms of student and expert interaction (Freddi 2005, Hyland 2005b, Mei 2007). Other studies include comparisons across disciplines (Fløttum *et al*. 2006) and between popular and professional science (Hyland 2010). Engagement markers have also been investigated cross-culturally in Spanish and English Business Management RAs (Mur Dueñas 2008, 2011) and, more indirectly as part of an analysis of first person

markers, in native English, native Italian and non-native English medical editorials (Giannoni 2008).

With regard to gender-related traits, there has been little research, probably because of the scarcity of gender-related studies in academic discourse in general, as noted by Sala and D'Angelo (2009). Exceptions are, precisely, Sala and D'Angelo's (2009) study of RAs in English in the fields of Applied Linguistics, Economics and Law, and Tse and Hyland's (2006) investigation of academic book reviews in Philosophy, Biology and Sociology. These studies are relevant to the present investigation of engagement as in both of them, men were found to be more reader-oriented than women, using engagement markers more often.

4.2. Method

Inclusive authorial references were retrieved using the Concord Tool of WordSmith Tools 6.0 (Scott 2011). As for English, it was sufficient to search for all the occurrences of the pronoun *we*, whereas in the case of Italian, all the relevant verb endings (i.e., *-amo*, *-emo*, *-imo*, *-mmo*, *-ssimo*) were retrieved. A disambiguation process followed, aiming at distinguishing between inclusive authorial references and forms of editorial *we/noi* and *pluralis majestatis*. Finally, occurrences were classified and quantified according to semantic referent and discourse function.

4.3. Results

4.3.1. Frequency of use of inclusive authorial references and functional analysis

Table 3 shows the incidence of use of inclusive authorial references in subject position in English and Italian Linguistics RAs. Italian writers appear more eager than their Anglo-American colleagues to use this engagement strategy. This result is not directly comparable to other findings in the literature. However, in a study of metadiscourse in Business Management RAs, Mur Dueñas (2011) found that Spanish

writers, too, use inclusive authorial references more often than Anglo-American scholars. Her interpretation is that Spanish scholars are aware of addressing a fairly small and homogeneous audience, which would make them more inclined to appeal to mutual disciplinary understandings. This hypothesis may also hold for the present comparison; however, in order to gain a deeper understanding of differences, it may be useful to investigate how data are distributed across semantic referents and discourse functions.

	Inclusive authorial references	
	Raw	*Normalised*
Linguistics RAs in English	241	11.16
Linguistics RAs in Italian	358	16.04

Table 3. Occurrences of inclusive authorial references (raw and normalised figures per 10,000 words).

Four main semantic referents were identified as being encapsulated in inclusive forms: general academic referents (i.e. the other members of the disciplinary community), readers, language users and general non-academic referents (e.g. 'we as people/humans'). Overall, I found a correspondence between type of referent and discourse function, namely:

- *Appealing to the disciplinary background.* When academic referents are brought into the discourse, the function of inclusive forms is to appeal to the disciplinary background in terms of communal knowledge, values, objectives, interests, problems and expectations, as can be seen in the following examples:

(1) I have argued that *we* need a more dynamic view of language and of its learning. (AL4)

(2) To examine language socialization within a community of practice framework, it is necessary to look at the micro-interactional level of ongoing practical activity. This means that *we* have to come to understand what people do. (IJAL6)

- *Elaborating an argument.* When writers resort to inclusive forms to elaborate their argument, they ask for a collaborative construction of meaning by readers, as if the discussion were elaborated simultaneously by both participants, as in examples (3) and (4):

(3) If *we* take the first premise of each case to be a (very) simple context C, in C (22), *we* can say that P is relevant in its own right […]. (L2)

(4) *We* can thus explain why (45), the Finnish counterpart to (1) with a null subject, repeated here, is ill-formed […]. (SL4)

- *Providing metadiscourse guidance*. In this case, too, readers are invoked as semantic referents. Writers guide them through the text thanks to previews and reviews, and by indicating where they can find examples and other information (e.g. tables and appendices), as in (5) and (6):

(5) Although later studies showed some variability across L1/L2, and *we* will consider this in more detail towards the end of this paper […]. (AL2)

(6) *We* have also seen in (12), however, that it can at times serve simultaneously to reduce speaker responsibility […]. (L2)

- *Making generalizations.* Writers refer to typical situations using *we* to include language users or non-academic referents in general.

(7) [Participants] were also told that they would probably not detect every contradiction, as *we* humans have been telling stories for a long time […]. (IJAL7)

(8) How does Derrida inform our thinking in Applied Linguistics? He shows how completely *we*, even we linguists, take language for granted […]. (AL5)

Table 4 presents the frequency of occurrence of inclusive authorial references across discourse functions.

Discourse function	*English*	*Italian*
Appealing to the disciplinary background	4.08	1.97
Elaborating an argument	3.85	6.94
Providing metadiscourse guidance	1.49	5.96
Making generalizations	1.76	1.16

Table 4. Inclusive authorial references across functions (normalised figures per 10,000 words).

The results show quite distinct patterns across the two languages. Only the function of *making generalizations* is similar in frequency in English and Italian. Anglo-American linguists seem to use inclusive references mainly to claim disciplinary solidarity and membership, and to engage readers in the elaboration of the argument. Italian writers, on the other hand, make a more modest use of appeals to solidarity, but engage readers to a greater extent, by involving them in the construction of the line of reasoning and providing more metadiscourse guidance.

4.3.2. The role of text superstructure

Tables 5 and 6 provide the same results as Table 4, but broken down in terms of text superstructure.

Discourse function	*IMRAD English*	*IMRAD Italian*
Appealing to the disciplinary background	1.67	1.25
Elaborating an argument	2.04	3.49
Providing metadiscourse guidance	1.07	3.94
Making generalizations	0.79	0.58

Table 5. Inclusive authorial references across functions (IMRAD text superstructure). (Normalised figures per 10,000 words).

Discourse function	*LA English*	*LA Italian*
Appealing to the disciplinary background	2.37	0.72
Elaborating an argument	1.81	3.45
Providing metadiscourse guidance	0.70	2.02
Making generalizations	0.70	0.58

Table 6. Inclusive authorial references across functions (LA text superstructure). (Normalised figures per 10,000 words).

In both Tables 5 and 6, inclusive authorial references in Italian texts tend to correlate with the functions of *elaborating an argument* and *providing metadiscourse guidance*. Therefore, it seems that the association of inclusive authorial references with these two functions is a crosscutting writing habit of the community of Italian linguists, regardless of the macro rhetorical configuration of RAs. Equally, the preference for the association of inclusive authorial references with the functions of *appealing to the disciplinary background* and *elaborating an argument* is found in English irrespective of the sub-type of RA. These results seem to corroborate those obtained in Table 4, indicating that the two writing cultures follow quite distinct patterns of use of inclusive authorial references.

These findings, however, do not imply that text superstructure has no influence on discourse patterns. Tables 5 and 6 indicate that it may affect the relative frequency of functions in texts. This is particularly evident in English IMRAD papers, where the most common use of inclusive authorial references is to elaborate an argument and not to appeal to the disciplinary background, as appeared in Table 4. This difference suggests that if text superstructure is not properly balanced, it may affect overall results.

4.3.3. The role of gender

Tables 7 and 8 show the results of Table 4 broken down in terms of gender distinctions. Contrary to expectations, both Italian and Anglo-American male scholars resort to inclusive authorial references slightly less than women, although the difference in the Italian sub-corpus is more noticeable. As for discourse functions, the most striking data is the marked preference for the function of *appealing to*

the disciplinary background by female Anglo-American writers. Therefore, it would seem that gender affects writing not only in terms of frequency of use of items, but in English, in particular, also in terms of preferred form-function association patterns.

Discourse function	*English (women)*	*Italian (women)*
Appealing to the disciplinary background	3.11	0.81
Elaborating an argument	1.53	3.99
Providing metadiscourse guidance	0.32	3.18
Making generalizations	0.97	0.63
TOT	5.93	8.61

Table 7. Inclusive authorial references across functions in RAs by women (normalised figures per 10,000 words).

Discourse function	*English (men)*	*Italian (men)*
Appealing to the disciplinary background	0.97	1.16
Elaborating an argument	2.32	2.96
Providing metadiscourse guidance	1.44	2.78
Making generalizations	0.51	0.54
TOT	5.24	7.44

Table 8. Inclusive authorial references across functions in RAs by men (normalised figures per 10,000 words).

In order to assess whether gender alone is responsible for this discrepancy between female and male English writers, I considered results obtained for individual texts. Two texts by female writers, i.e. AL4 and AL5 (see Appendix), feature a considerably high use of inclusive *we*, i.e. 30 and 40 occurrences respectively, while the frequencies of the rest of the articles by women range between 1 and 18 hits. This divergence suggests that other factors than gender may account for the high use of inclusive authorial references in AL4 and AL5.

The question, therefore, is why AL4 and AL5 feature such an extensive use of inclusive *we*, particularly for the function of appealing to the disciplinary background (15 and 24 raw occurrences respectively), thus skewing the overall results of Table 4. In the

attempt to answer this question, I will consider the impact of aspects such as text length, superstructure, native-speaker status and date of publication. Text length may be an influencing factor in the case of AL4, which totals 7,631 words; however, it seems not to be relevant for AL5, which counts 6,813 words, thus not diverging significantly from average text length, i.e. 6,733. Text superstructure may also play a role, at least in part. Despite being shorter, AL5 (LA paper) features a higher number of inclusive *we* than AL4 (IMRAD paper), which might suggest that the logico-argumentative nature of AL5 encourages more appeals to communal knowledge. Nevertheless, the frequencies of *we* in AL4, too, are well above the average of the IMRAD papers in the corpus. Attributing the high use of *we* to the supposedly non-native status of writers (see section 3) is equally disputable. While one writer may actually not be a native speaker of English (AL5), web-based searches confirmed that the other is a native speaker (AL4). Finally, the time coordinate does not seem to be a significant factor either, as AL4 was published in 2006 and AL5 in 2008. Therefore, the high use of *we* to appeal to the disciplinary background seems to be related to either individual idiosyncrasies or to other contextual factors that could not be controlled during the sampling process, such as sub-discipline and topic.

This analysis suggests that individual divergences may have a great impact on overall results, not only in terms of frequency but also in terms of distribution patterns. Consequently, it becomes difficult to distinguish the role of idiosyncrasies or factors such as topic from that of writing culture. In the Italian corpus, too, there are articles in which the overall incidence of use of inclusive authorial references is particularly high (i.e. AGI5, 46 occurrences and RID5, 69 hits). For this reason, I decided to subtract AL4 and AL5 from the English corpus and, in order to preserve balance, I also excluded two papers by male writers, one per each type of text superstructure (L5 and SL2), in which the occurrence of use of *we* was considerably high. I did the same in the Italian sub-corpus (AGI5, RID5, RID7, SGI1). The resulting corpus contains 28 RAs in each sub-corpus and Table 9 summarises the results obtained from its analysis.

Discourse function	*English (28 texts)*	*Italian (28)*
Appealing to the disciplinary background	2.23	0.99
Elaborating an argument	2.51	3.85
Providing metadiscourse guidance	0.70	3.40
Making generalizations	0.79	0.76
TOT	6.23	9.00

Table 9. Inclusive authorial references across functions in English and Italian (28 texts). (Normalised figures per 10,000 words).

In English, the discrepancy between the two discourse functions of *appealing to the disciplinary background* and *elaborating an argument* is not only less evident as compared to the data in Table 4, but their reciprocal significance is inverted, with the function of *elaborating an argument* appearing as slightly more frequent. In Italian, on the other hand, the data relating to the articles omitted had an impact on total frequencies but not on the description of preferred patterns of writer-reader interaction, which remain unvaried. The data displayed in Table 9 are more likely to reflect the role of writing culture since the articles that skewed the results have been omitted.

4.4. Summary of cross-cultural findings

This exploratory study of reader engagement indicates that in both English and Italian Linguistics RAs, inclusive authorial references are most often used to engage readers in the construction of the argument, so as to make them appear active and collaborative discourse participants. Anglo-American and Italian linguists, however, differ in the extent to which they resort to inclusive authorial references to appeal to communal background and to provide metadiscourse guidance. While Anglophone scholars prefer to engage their readers to refer to shared knowledge, values, problems and expectations, Italian linguists prefer to involve readers when it comes to the construction of the text, as if sharing the responsibility for its organization, coherence and clarity.

5. Concluding remarks

The analysis of inclusive authorial references reported in this chapter has illustrated how a stratified corpus can be used in the cross-cultural study of academic discourse. The choice of a stratified corpus enables researchers to control the interference of contextual factors. A balanced corpus with uniform allocation also enables analysts to compare corpora within individual strata, as was done in 4.3.2 and 4.3.3, in order to investigate the impact of specific contextual factors on overall results. Finally, a balanced corpus facilitates the task of evaluating the influence of individual texts and uniform allocation enables researchers to fine-tune the corpus so as to arrive at a maximum grade of comparability.

Differences were found between English and Italian Linguistics RAs. Overall, Italian linguists resort to inclusive authorial references more often than Anglo-American scholars. In addition, in the two writing communities inclusive authorial references tend to be used for somewhat dissimilar discursive strategies. The analysis of the results has also shown that some articles were significantly divergent from corpus tendencies, thus skewing overall quantitative data. If on the one hand, those divergences may indicate that within the two communities more marked linguistic patterns are acceptable, on the other hand, they also suggest that in further investigations of academic writing, sub-discipline and topic should be considered as parameters to include in the sampling frame, if enough text material is available.

This issue is relevant for the last question posed in the Introduction: is quantitative descriptive research still meaningful when maximal comparability is difficult to achieve? If a contextual variable is too infrequent to lend itself to be sampled, other methods of inquiry will best suit its investigation, such as case studies or ethnographic analyses. Nevertheless, as Leech observes (2007: 143-144), if our corpus design goals cannot be fully satisfied,

> it is best to recognise that these goals are not an all-or-nothing: there is a scale of representativity, of balancedness, of comparability. We should seek to

> define realistically attainable positions on these scales, rather than abandon them altogether.

In a situation such as the one illustrated in this chapter, where several major contextual factors were controlled, the difficulty of achieving maximal comparability may be overcome if the design criteria are explicitly described so as to be able to identify areas of imbalance and their potential impact on results.

References

Burgess, Sally 2002. Packed Houses and Intimate Gatherings: Audience and Rhetorical Structure. In Flowerdew, John (ed.) *Academic Discourse*. Harlow: Longman, 196-215.

Connor, Ulla 2002. New Directions in Contrastive Rhetoric. *TESOL Quarterly* 36, 493-510.

Connor, Ulla / Moreno, Ana I. 2005. Tertium Comparationis: A Vital Component in Contrastive Research Methodology. In Bruthiaux, Paul / Atkinson, Dwight / Eggington, William G. / Grabe, William M. / Ramanathan, Vaidehi (eds) *Directions in Applied Linguistics: Essays in Honor of Robert B. Kaplan*. Clevendon: Multilingual Matters, 153-164.

Dörnyei, Zoltán 2007. *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.

Ehrlich, Susan 2004. Language and Gender. In Davies Alan / Elder, Catherine (eds) *Handbook of Applied Linguistics*. Oxford: Blackwell, 304-327.

Fløttum, Kjersti / Kinn, Torodd / Dahl, Trine 2006. 'We Now Report On...' Versus 'Let Us Now See How...': Author Roles and Interaction with Readers in Research Articles. In Hyland, Ken / Bondi, Marina (eds) *Academic Discourse Across Disciplines*. Bern: Peter Lang, 203-224.

Freddi, Maria 2005. Arguing Linguistics: Corpus Investigation of One Functional Variety of Academic Discourse. *Journal of English for Academic Purposes* 4, 5-26.

Giannoni, Davide S. 2008. Medical Writing at the Periphery: The Case of Italian Journal Editorials. *Journal of English for Academic Purposes* 7, 97-107.

Hewings, Anne / Lillis, Theresa / Vladimirou, Dimitra 2010. Who's Citing Whose Writings? A Corpus Based Study of Citations as Interpersonal Resource in English Medium National and English Medium International Journals. *Journal of English for Academic Purposes* 9, 102-115.

Hunston, Susan 2008. Collection Strategies and Design Decisions. In Lüdeling, Anke / Kytö, Merja (eds) *Corpus Linguistics: An international Handbook*, Volume 1. Berlin: Mouton de Gruyter, 154-168.

Hyland, Ken 2000. *Disciplinary Discourses: Social Interactions in Academic Writing*. London: Pearson.

Hyland, Ken 2005a. Stance and Engagement: A Model of Interaction in Academic Discourse. *Discourse Studies* 7/2, 173-192.

Hyland, Ken 2005b. Representing Readers in Writing: Student and Expert Practices. *Linguistics and Education* 16, 363-377.

Hyland, Ken 2010. Constructing Proximity: Relating to Readers in Popular and Professional Science. *Journal of English for Academic Purposes* 9, 116-127.

Hyland, Ken / Bondi, Marina (eds) 2006. *Academic Discourse Across Disciplines*. Bern: Peter Lang.

Leech, Geoff 2007. New Resources, or Just Better Old Ones? In Hundt, Marianne / Nesselhauf, Nadja / Biewer, Carolin (eds) *Corpus Linguistics and the Web*. Amsterdam: Rodopi, 134-49.

Leki, Ilona 1991. Twenty-Five Years of Contrastive Rhetoric: Text Analysis and Writing Pedagogies. *TESOL Quarterly* 25/1, 123-143.

Mauranen, Anna 2001. Descriptions or Explanations? Some Methodological Issues in Contrastive Rhetoric. In Hewings, Martin (ed.) *Academic Writing in Context. Implications and Applications*. London: Continuum, 43-54.

Mei, Wu Siew 2007. The Use of Engagement Resources in High- and Low-rated Undergraduate Geography Essays. *Journal of English for Academic Purposes* 6, 254-271.

Molino, Alessandra 2010. Personal and Impersonal Authorial References: A Contrastive Study of English and Italian Linguistics Research Articles. *Journal of English for Academic Purposes* 9/2, 86-101.

Moreno, Ana I. 2008. The Importance of Comparable Corpora in Cross-Cultural Studies. In Connor, Ulla / Nagelhout, Ed / Rozycki, William (eds) *Contrastive Rhetoric: Reaching to Intercultural Rhetoric*. Amsterdam: John Benjamins, 25-41.

Mur Dueñas, Pilar 2008. Analysing Engagement Markers Cross-Culturally: The Case of English and Spanish Business Management Research Articles. In Burgess, Sally / Martín Martín, Pedro (eds) *English as an Additional Language in Research Publication and Communication*. Bern: Peter Lang, 197-213.

Mur Dueñas, Pilar 2011. An Intercultural Analysis of Metadiscourse Features in Research Articles Written in English and in Spanish. *Journal of Pragmatics* 43, 3068–3079.

Pérez-Llantada, Carmen / Ramón, Plo / Ferguson, Gibson R. 2011. 'You Don't Say What You Know, Only What You Can': The Perceptions and Practices of Senior Spanish Academics Regarding Research Dissemination in English. *English for Specific Purposes* 30, 18-30.

Sala, Michele / D'Angelo, Larissa 2009. Gender- and Authority-related Writing Performance in Academic Research Articles. *Textus* 22/1, 59-76.

Salager-Meyer, Françoise / Alcaraz Ariza, María Angeles / Zambrano, Nahirana 2003. The Scimitar, the Dagger, and the Glove: Intercultural Differences in the Rhetoric of Spanish, French and English Medical Discourse 1930-1995. *English for Specific Purposes* 22, 223-247.

Scott, Mike 2011. *WordSmith Tools* (Version 6.0). Liverpool: Lexical Analysis Software.

Swales, John M. 1990. *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Tarone, Elaine / Dwyer, Sharon / Gillette, Susan / Icke, Vincent 1998. On the Use of the Passive and Active Voice in Astrophysics Journal Papers: With Extensions to Other Languages and Other Fields. *English for Specific Purposes* 17/1, 113-132.

Tse, Polly / Hyland, Ken 2006. Gender and Discipline: Exploring Metadiscourse Variation in Academic Book Reviews. In Hyland, Ken / Bondi, Marina (eds) *Academic Discourse Across Disciplines*. Bern: Peter Lang, 177-202.

Yakhontova, Tatyana 2006. Cultural and Disciplinary Variation in Academic Discourse: The Issue of Influencing Factors. *Journal of English for Academic Purposes* 5, 153-167.

Appendix

Italian sub-corpus

Amenta, Luisa 2006. La perifrasi aviri da /a + infinito nel siciliano contemporaneo. Analisi di un campione di dati ALS. *Rivista Italiana di Dialettologia* XXX, 59-73. [RID1]

Arcangeli, Massimo 2007. Il lessico sportivo e ricreativo italiano nelle quattro grandi lingue europee (con qualche incursione anche altrove). *Studi di Lessicografia Italiana* XXIV, 195-247. [SLI1]

Aureli, Massim 2003. Pressione dell'uso sulla norma: le relative non-standard nei giudizi degli utenti. *Studi Italiani di Linguistica Teorica e Applicata* 32, 45-67. [SILTA1]

Benati, Alessandro 2005. Ricreazione nella classe di lingua. Il modello sperimentale. *Rassegna Italiana di Linguistica Applicata* 37/2-3, 53-73. [RILA1]

Chinellato, Paolo 2003. Sintassi dialettologica e afasia non fluente. Analisi linguistica e strategie riabilitative. *Rivista Italiana di Dialettologia* XXVII, 179-196. [RID2]

D'Achille, Paolo 2005. L'invariabilità dei nomi nell'italiano contemporaneo. *Studi di grammatica italiana* XXIV, 189-209. [SGI1]

Dalosio, Michele 2006.Lingua straniera e sviluppo dei processi di memoria del bambino. *Rassegna Italiana di Linguistica Applicata* 38/2-3, 377-396. [RILA2]

De Cesare, Anna-Maria 2005. La frase pseudoscissa in italiano contemporaneo: aspetti semantici, pragmatici e testuali. *Studi di grammatica italiana* XXIV, 293-322. [SGI2]

Fresu, Rita 2006. Gli uomini parlano delle donne, le donne parlano degli uomini. Indagine sociolinguistica in un campione giovanile di area romana e cagliaritana. *Rivista Italiana di Dialettologia* XXX, 23-53. [RID3]

Grandi, Nicola 2005. I verbi valutativi in italiano tra azione e aspetto. *Studi di grammatica italiana* XXIV, 153-188. [SGI3]

Guazzieri, Annavaleria 2008. Conoscenza dell'apprendimento cooperativo da parte degli insegnanti CLIL. *Rassegna Italiana di Linguistica Applicata* 40/1-2, 91-115. [RILA3]

Lisma, Grazia M. 2007. Sul lessico marinaresco dell'Ottocento. *Studi di Lessicografia Italiana* XXIV, 165-194. [SLI2]

Lo Prete, Rosario 2005. Ricerca sullo studio della L2 nella scuola primaria. *Rassegna Italiana di Linguistica Applicata* 37/2-3, 121-138. [RILA4]

Maiocco, Marco 2004. Varietà e intensità del contatto linguistico: inerzia sintattica nel greco antico semitizzato. *Archivio Glottologico Italiano* 89/1, 84-110. [AGI1]

Marazzi, Martino 2007. Preistoria e storia di "Afro-Americano". *Studi di Lessicografia Italiana* XXIV, 249-264. [SLI3]

Mariottini, Laura 2006. La pragmatica della CMC. Strategie di cortesia linguistica nelle interazioni in chat. *Studi Italiani di Linguistica Teorica e Applicata* 35/2, 319-338. [SILTA2]

Masini, Francesca 2006. Diacronia dei verbi sintagmatici in italiano. *Archivio Glottologico Italiano* 91/1, pp. 67-105. [AGI2]

Mauri, Caterina 2006. Combinazione e contrasto: i connettivi, congiuntivi e avversativi nelle lingue d'Europa. *Archivio Glottologico Italiano* 91/2, 166-202. [AGI3]

Menegale, Marcella 2008. Tipi di domande utilizzate durante la lezione frontale partecipata e output degli studenti. *Rassegna Italiana di Linguistica Applicata* 1/2, 117-145. [RILA5]

Miotti, Renzo 2010. Importazione fonologica in spagnolo e italiano. Riflessioni sul consonantismo. *Studi italiani di linguistica teorica e applicata* 39/3, 491-521. [SILTA3]

Patrucco, Elisa 2003. Sul dialetto in internet. *Rivista Italiana di Dialettologia* XXVII, 139-174. [RID4]

Pinto, Immacolata 2004. Alcune osservazioni sul prefisso in- negativo nel sardo e in area romanza. *Rivista Italiana di Dialettologia* XXVIII, 197-218. [RID5]

Rati, M. S. 2004. L'alternanza tra indicativo e congiuntivo nelle proposizioni completive: sondaggi sulla prosa italiana del Due-Trecento. *Studi di Grammatica Italiana* XXIII, 1-59. [SGI4]

Regis, Riccardo 2006. Breve fenomenologia di una locuzione avverbiale: il "solo piu" dell'italiano regionale piemontese. *Studi di Lessicografia Italiana* XXIII, 275-289. [SLI4]

Romagno, Domenica 2006. Gradiente di transitività e codifica dell'oggetto. Dall'accusativo preposizionale al partitivo. *Archivio Glottologico Italiano* 91/2, 203-222. [AGI4]

Rossi, Giovanni 2009. Il discorso scritto interattivo degli SMS. Uno studio pragmatico del "messaggiare". *Rivista Italiana di Dialettologia* XXXIII, p.143-193. [RID6]

Rossi, Lucia 2004. Italia-Europa: una ricognizione di terminologia glottodidattica. *Studi italiani di linguistica teorica e applicata* 33/3, 453-480. [SILTA4]

Ruggiero, Raffaele Savio 2004. Il dialetto tra i giovani a Torino. *Rivista Italiana di Dialettologia* XXVIII, 11-35. [RID7]

Scarpino, Cristina 2008. Il lessico scientifico nel dizionario di John Florio. *Studi di Lessicografia Italiana* XXV, 65-95. [SLI5]

Stefanelli, Rossana 2008. E' un composto nominale. Analizzare con cautela. *Archivio Glottologico Italiano* 93/1, 3-45. [AGI5]

Telve, Stefano 2004. Vicende editoriali e normative della Gramatica ragionata della lingua italiana di Francesco Soave. *Studi di Grammatica Italiana* XXIII, 61-86. [SGI5]

Vietti, Alessandro 2009, Contatto e variazione nell'italiano di immigrati. La formazione di una varietà etnica. *Studi italiani di linguistica teorica e applicat*a 38/1, 29-53. [SILTA5]

English sub-corpus

Anderson, Kate T. 2007. Constructing "otherness": ideologies and differentiating speech style. *International Journal of Applied Linguistics* 17/2, 178-197. [IJAL1]

Barnbrook, Geoff 2005. Usage notes in Johnson's dictionary. *International Journal of Lexicography* 18/2, 189-201. [IJL1]

Bender, Emily M. 2005. On the Boundaries of Linguistic Competence: Matched-Guise Experiments as Evidence of Knowledge of Grammar. *Lingua* 115, 1579–1598. [L1]

Berns, Margie 2005. Expanding on the Expanding Circle: Where do WE Go from Here? *World Englishes*, 24/1, 85-93. [WE1]

Boecxk, Cedric 2004. Long-Distance Agreement in Hindi: Some Theoretical Implications. *Studia Linguistica*, 58/1, 23-36. [SL1]

Ellis Nick C. 2006. Selective Attention and Transfer Phenomena in L2 Acquisition: Contingency, Cue Competition, Salience, Interference, Overshadowing, Blocking, and Perceptual Learning. *Applied Linguistics* 27/2, 164-194. [AL1]

Ellis Rod 2006. Modelling Learning Difficulty and Second Language Proficiency: The Differential Contributions of Implicit and Explicit Knowledge. *Applied Linguistics* 27/3, 431-463. [AL2]

Friginal, Eric 2007. Outsourced call centers and English in the Philippines. *World Englishes* 26/3, 331-345. [WE2]

Hall, Tracy A. 2006. English syllabification as the interaction of markedness constraints. *Studia Linguistica*, 60/1, 1-33. [SL2]

Hellerman, John 2006. Classroom Interactive Practices for Developing L2 Literacy: A Microethnographic Study of Two Beginning Adult Learners of English. *Applied Linguistics* 27/3, 377-404. [AL3]

Hill, Virginia 2006. Stylistic Inversion in Romanian. *Studia Linguistica* 60/2, 156-180. [SL3]

Holmberg, Anders 2007. Null subjects and polarity focus. *Studia Linguistica*, 61/3, 212-236. [SL4]

Hughes, Shaun F. D. 2004. Was there ever a "Māori English"? *World Englishes*, 23/4, 565-584. [WE3]

Hyland, Ken 2008. Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics* 18/1, 41-62. [IJAL2]

Jary, Mark 2008. The Relevance of Complement Choice: A Corpus Study of "Believe". *Lingua* 118, 1-18. [L2]

Larsen-Freeman, Diane 2006. The Emergence of Complexity, Fluency, and Accuracy in the Oral and Written Production of Five Chinese Learners of English. *Applied Linguistics* 27/4, 590-619. [AL4]

Lindeman, Stephanie 2005. Who speaks "broken English"? US undergraduates' perceptions of non-native English. *International Journal of Applied Linguistics* 15/2, 188-212. [IJAL3]

Martin, Elizabeth 2007. "Frenglish" for Sale: Multilingual Discourses for Addressing Today's Global Consumer. *World Englishes*, 26/2, 170-188. [WE4]

McDermott, Anne 2005. Johnson's Definitions of Technical Terms and the Absence of Illustration. *International Journal of Lexicography*, 18/2, 173-187. [IJL2]

McKenzie Robert M. 2008. Social factors and non-native attitudes towards varieties of spoken English: A Japanese case study. *International Journal of Applied Linguistics* 18/1, 63-88. [IJAL4]

Mitchell, Linda C. 2005. Johnson Among The Early Modern Grammarians. *International Journal of Lexicography*, 18/2, 203-216. [IJL3]

Moon, Rosamund 2004. On Specifying Metaphor: An Idea and its Implementation. *International Journal of Lexicography*, 17/2, 195-222. [IJL4]

Morgan, Gary 2006. "Children are Just Lingual": The Development of Phonology in British Sign Language (BSL). *Lingua* 116, 1507-1523. [L3]

Osselton, Noel E. 2006. Usage guidance in early dictionaries of English. *International Journal of Lexicography*, 19/1, 99-105. [IJL5]

Padney, Anita 2004. Culture, Gender, and Identity in Cross-Cultural Personals and Matrimonials. *World Englishes*, 23/3, 403-427. [WE5]

Pereltsvaig, Asya 2007. The Universality Of Dp: A View From Russian. *Studia Linguistica* 61/1, 59-94. [SL5]

Ramanathan, Vaidehi 2008. Redux: A Derridean Exploration of Alzheimer Life Histories. *Applied Linguistics* 29/1, 1-23. [AL5]

Seargent, Philip 2005. "More English than England itself": the simulation of authenticity in foreign language practice in Japan. *International Journal of Applied Linguistics* 15/3, 326-345. [IJAL5]

Smith Jennifer L. 2004. Making Constraints Positional: Toward a Compositional Model of CON. *Lingua* 114, 1433-1464. [L4]

Velde, John R 2005. Unifying Prosody and Syntax for Right- and Left-Edge Coordinate Ellipsis. *Lingua* 115, 483-502. [L5]

Vickers, Caroline H. 2008. Expertise, Language Competencies and the L2 User. *International Journal of Applied Linguistics* 18/3, 237-255. [IJAL6]

Walter, Catherine 2007. First- to second-language reading comprehension: Not transfer, but access. *International Journal of Applied Linguistics* 17/1, 14-37. [IJAL7]

PATRIZIA ANESA

2. Avoiding Plagiarism and Self-plagiarism through the Use of Corpora[1]

Introduction

Plagiarism is an ancient phenomenon that seems to be present in a vast array of eras, regions, disciplines, and fields, and a plethora of tools and procedures have been developed in order to try to contrast it. Indeed, plagiarism assumes different forms and, subsequently, plagiarism detection encompasses a wide range of methodologies that may be applied to different fields in the professional, academic and educational world. In particular, two of the main research trends focus on free text plagiarism (for a survey see Maurer *et al.* 2006) and on source-code plagiarism detection (e.g. Arwin/Tahaghoghi 2006).

This chapter focuses on text plagiarism and in particular on written academic texts. As regards academic writing, the battle against plagiarism seems to become more and more complex as it is often stated that the advent of the Internet has multiplied the opportunities for plagiarism (Howard 2007), making it easier and faster. For instance, if we look at the simplest and most obvious form of plagiarism, i.e. copy-and-paste procedures, we know that they are adopted worldwide by myriads of students. Although this battle seems, to some extent, bound to fail, different tools and strategies are being implemented in order to detect plagiarism efficiently and effectively. Given the nature of the principal contemporary forms of

1 I am deeply indebted to Giuseppe Parrinello for the precious help received in this study, especially as regards Section 3.

plagiarism, it is generally argued that these tools are aimed in particular at the detection of digital plagiarism. Traditional plagiarism and 'digital' plagiarism are no longer clearly distinguishable given the number of publications made available in digital format and of publications which are only available digitally. Consequently, digital forms of plagiarism will most likely continue to be the most widespread.

The first section of the paper offers a taxonomy of plagiarism, showing the different forms assumed by the phenomenon and its main characteristics. The second section is devoted to a discussion of the main approaches to plagiarism detection. Finally, the third section focuses on how corpora or other types of text collections may be used for plagiarism detection and research, and it also presents the potential uses of *PlagiarismFinder* (*PF*), a programme specifically developed for preliminary external detection.

1. Defining plagiarism

Plagiarism may be broadly defined as "the use of another author's words, research, or ideas without proper attribution and citation" (Stepchyshyn/Nelson 2007: 63). Therefore, it involves the inappropriate use of intellectual property (Bloch 2012: 1) and, from a legal perspective, it does not necessarily equal copyright infringement in that it may be based on the duplication of material that is not copyrighted. More precisely, Murray highlights that plagiarism refers to the "use or reuse of words or ideas without *acknowledgment*", whereas copyright infringement is the "use or reuse of words or ideas without *permission*" (Murray 2008: 174).

Plagiarism depends significantly on cultural factors and a monolithic interpretation of plagiarism is not possible. Buranen points out that cultural differences are not the only factor to be used in defining plagiarism, but such differences inevitably contribute to the complexification of the problem (Buranen 1999: 65). Indeed, the

boundaries between rephrasing, paraphrasing, citing, and plagiarizing are often unclear and different cultures may assign different levels of importance to originality and creativity (for a discussion of plagiarism in different eras and cultures see Anderson 1998).

As has been mentioned, plagiarism may concern different fields, countries and languages, and the phenomenon is so widespread that it is impossible to clearly identify which categories of people are more inclined to commit plagiarism. For instance, as regards academic writing, some studies have focused on the incidence of plagiarism in works written by native and non-native speakers (Campbell 1990, Pecorari 2003), but no incontrovertible differences seem to emerge.

Plagiarism is often investigated in connection with other phenomena such as the practice of 'ghost writing' (writing texts that are officially credited to another person) and the use of essay mills. They represent forms of intellectual cheating, but not necessarily of plagiarism, however condemnable these practices may be. For instance, a ghost writer does not generally make use of plagiarism in its original sense. Similarly, essay mills are often sources of plagiarism in that the same essay may be sold to several students, and, therefore, the same text is attributed to different people who present it as their own. However, the original text may be exempt from instances of plagiarism.

The concept of plagiarism is extremely protean and its realization may assume a vast variety of forms. Different parameters could be used in order to identify different types of plagiarism, among which I would suggest the following:

- intentional and unintentional (or accidental) plagiarism;
- other-plagiarism and self-plagiarism;
- exact, near-exact and concealed plagiarism.

The next section will discuss these phenomena in detail.

1.1. Intentionality

Aspiring to adopt criteria that allow us to discern objectively and incontrovertibly between intentional and unintentional plagiarism is a process that is somehow doomed to failure. However, situations in which authors do not clearly acknowledge other people's words or ideas with the obvious aim of presenting them as their own is to be seen as a deliberate act of plagiarism. In contrast, occasional mistakes and inaccuracies should be framed within a different category; not using quotations marks when the source is cited, or, conversely, not citing a source but presenting a quotation correctly may be examples of unintentional plagiarism. This category is often associated with less experienced writers. Borderline cases are of course numerous: for example, a free and inaccurate use of class notes may be seen as a form of unintentional plagiarism linked to lack of experience on the part of a student in assigning citations and indicating sources. Moreover, a student may have genuinely misunderstood that a certain statement pronounced in class was in fact a quotation.

1.2. Other- and self-plagiarism

Text plagiarism in its simplest terms refers to the practice of using somebody else's words as if they were our own. Self-plagiarism, however, may also occur. The American Psychological Association defines self-plagiarism as follows: "Whereas plagiarism refers to the practice of claiming credit for the words, ideas, and concepts of others, self-plagiarism refers to the practice of presenting one's own previously published work as though it were new" (APA 2010: 170).

It is clear that in academic writing, especially at a scholarly level, a specific topic is often dealt with by the same author on numerous occasions, given the high level of specialization scholars tend to reach. Research is intrinsically based on the advancement of knowledge and publications are often conceived as follow-ups to previous works. When reading different publications by the same author, it is, therefore, understandable that the same topic may be introduced in similar ways or that the literature overview may lack a

high level of originality. As the APA underlines, however, the key distinction lies in the lack of acknowledgement regarding previously published works. Not citing the (more or less explicit) contribution that such studies make to the current publication is to be seen as plagiaristic practice. It is neither erroneous nor immoral to start from the same theoretical background or to use the same methodological approach to analyze different sets of data in order to answer a different research question. However, blatantly re-using the same sources, the same data, or the same wording without a clear reference to previously published work is a debatable praxis.

The expression 'self-plagiarism' itself may appear oxymoronic. The debate has often focused on its very nature and on whether or not it should be considered a crime. If we assume that the notion of plagiarism can be related to that of theft, is it plausible to assume the existence of a crime such as self-theft? Moreover, if we observe plagiarism from a philosophical perspective, we seem to be dealing with a kind of Sorites paradox[2]: what is the line that discerns exactly what plagiarism is and what isn't? How many words, ideas, concepts or methods can one 'steal' from himself before it constitutes a crime?

Practices related to self-plagiarism may be labeled in different ways, from more condemning expressions such as 'text recycling' to more laxist and apologetic ones such as 'text re-use' (Clough/ Gaizauskas 2009) or 'text self-borrowing'. Different phenomena, such as duplication, recycling or cryptomnesia, are included in or related to the idea of self-plagiarism and at times overlap. Duplicating or recycling part of one's research may be considered as a form of self-plagiarism, especially when dealing with an exact duplication of theories, materials, data, findings, etc., or when they are recycled without acknowledging their previous use. However, practices based on considerable modification, significant re-framing or different positioning of one's research may not be seen as plagiarism.

---

2 The Sorites Paradox (or Paradox of the Heap) is generally attributed to Eubulides of Miletus and states that it is impossible to establish the exact point of transition between two states along a continuum.

Even more complicated seem phenomena like cryptomnesia. In this case the distinction suggested between intentional and unintentional plagiarism plays a crucial role and assumes complex contours. Cryptomnesia refers to a recollection that is perceived as a new creation. In other words, as van den Berk states, "one remembers something without realizing it is a recollection" (2012: 3) and the process may regard one's own words or somebody else's. The phenomenon has long been investigated (see Jung's approach to cryptomnesia; see inter alia Jung 1970[1905]) and is an example of the complexities underlying the process of defining and detecting plagiarism and self-plagiarism.

1.3. Concealment

A further distinction among different types of plagiarism regards their level of concealment, which can be distinguished between exact, near-exact and concealed plagiarism. This is not to say the first two categories do not imply an attempt to conceal plagiarism on the part of the author, but that there is a less significant presence of obfuscation processes (see Barrón-Cedeño *et al.* 2010: 2). In fact, all forms of plagiarism are based on some form of concealment. As Posner aptly remarks, "concealment is at the heart of plagiarism" (2007: 17), and if the reference to the other source is obvious and unmistakable, it will not constitute plagiarism. Therefore, for example, an allusion should not be seen as plagiarism "because the readership is expected to recognize the allusion" (Posner 2007: 18).

As has been pointed out, the concept of plagiarism is protean, and presenting a clear taxonomy is complex as different categories overlap and may be placed along a continuum. Figure 1 attempts to show some of the different categories of plagiarism:

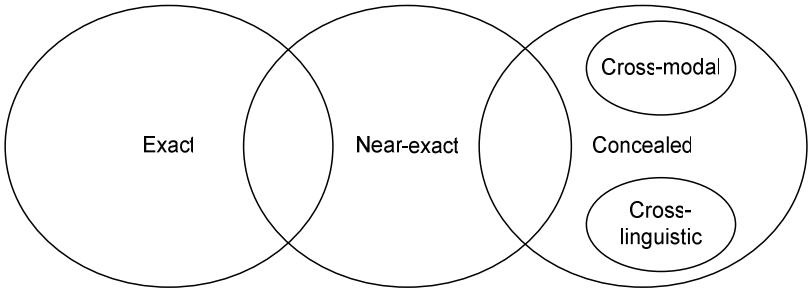


Figure 1. Types of plagiarism.

In its simplest form, exact plagiarism regards a verbatim replication of somebody else's words without an appropriate quotation. Near-exact plagiarism may involve procedures such as modest changes in word order, the use of synonyms, rephrasing strategies, etc. Concealed plagiarism, instead, concerns the duplication of concepts rather than words, in the sense that the author does not reproduce portions of texts or expressions literally but reframes and reformulates certain concepts without citing the original source.

The first type of plagiarism is of course the easiest to detect, in that detection software products can uncomplicatedly detect exact replication of words that are reproduced in a certain sequence. Conversely, as we shall see, processes of obfuscation make automatic detection more complex. Two of the phenomena that may be included in the category of concealed plagiarism are cross-modal and cross-linguistic plagiarism. The former is based on representing the same concept by using a different modality, e.g. a picture instead of words. The latter is a very common process in academic writing. In this case the translation of a text into another language obviously implies that an exact representation of the same words is absent, but the same concepts are presented through a different tongue without acknowledging the original source.

2. Plagiarism detection

Plagiarism detection may be manual or computer-mediated. The two approaches are clearly interdependent and are generally combined in order to achieve reliable findings.

As regards automatic analyses, different types of plagiarism detection software have been developed in recent years. A wide variety of systems is now available and they are ascribable to two main methods: internal (also called intrinsic or endophoric) and external (or extrinsic or exophoric) detection. Figure 2 attempts to rudimentally show the main types of approaches to plagiarism detection.

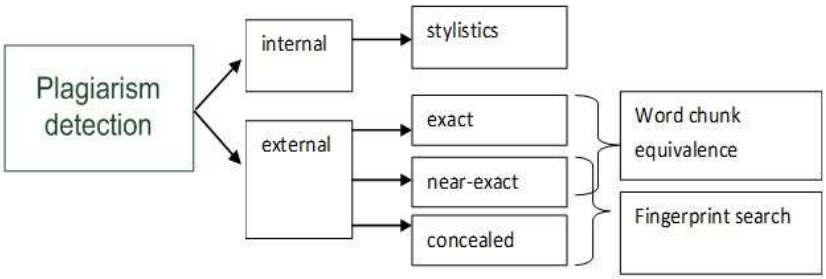


Figure 2. Main plagiarism detection approaches.

The principal internal detection method is based on stylistic analysis. This process aims at identifying chunks of a text which may not have been entirely written by the same author because they are characterized by a style that differs from the predominant one. This approach is often based on statistical evaluations focusing, for example, on lexical, syntactic and textual tendencies. The use of stylometry is now well consolidated and has been developing for

several years (see Holmes 1998). For example, *Stylysis*[3] is a Web application based on a set of measures for stylistic analysis. More specifically, it is based on the computation of features such as Gunning Fog Index, sentence length, average word length, Honore's R function, Yule's K function, and Flesch-Kincaid readability test. The results show the sentences which differ from the general writing style and, thus, may be cases of plagiarism.

Similar procedures are based on the computation of other features. For instance, the parameters listed by Meyer zu Eissen *et al.* (2007: 361) are:

- text statistics;
- syntactic features;
- part of speech features;
- closed-class word sets to count special words;
- structural features.

In other words, intrinsic plagiarism detection is generally used without external reference corpora or collections as it is based on the identification of stylistic variations and discrepancies within a text. However, it should be noted that stylometric tools may also be used in external detection, more specifically as a "preprocessing step to an external plagiarism detection tool" (Stamatatos 2009: 38).

On a final note, internal detection should be combined with manual evaluation in order to achieve more reliable results. Indeed, it is clear that an author may display different stylistic features within the same text for reasons that are not related to plagiarism and therefore false positive results may occur.

Exophoric plagiarism is generally based on text comparison, checking the text under investigation against others (Barrón-Cedeño *et al.* 2010). Automatic systems based on external detection are often

---

3 To learn more about Stylysis (developed by Barrón-Cedeño, Vallés-Balaguer and Rosso) and to access the stylistic analysis tool see: <http://memex2.dsic.upv.es:8080/StylisticAnalysis/en/index.jsp>

designed to detect not only sameness in word sequences but also concealed similarities which are identifiable through different procedures such as fingerprinting[4]. Several systems for the detection of plagiarism in natural languages are now available and they display a vast array of functions (for a comprehensive overview see Maurer *et al.* 2006). Some of the most widely used are:

- Turnitin by iParadigms
- WordCHECK by WordCHECKsystems
- Findsame by Digital integrity
- Eve2 by CaNexus
- CopyCatch by CFL Software Developments

These programmes are based on the comparison between the document being investigated and a collection of documents assumed to be original. Other online tools for plagiarism detection are available at:

- www.PlagAware.com
- www.PlagScan.com
- www.CheckForPlagiarism.net
- www.PlagiarismDetection.org

3. Using corpora

3.1. Corpora for plagiarism detection and research

Plagiarism detection and research may make use of different kinds of corpora. By and large, a corpus may be defined as "an electronically stored collection of samples of naturally occurring language" (Hunston 2006: 234). More specifically, McEnery and Wilson (2001:

4 Fingerprints can be defined as sets of "integers created by hashing subsets of a document to represent its key content" (El Bachir Menai 2012: 841).

197) specify that the term 'corpus' can refer to: "(i) (loosely) any body of text; (ii) (most commonly) a body of machine-readable text; (iii) (more strictly) a finite collection of machine-readable text, sampled to be maximally representative of a language or variety".

In this chapter I argue that corpora may be fruitfully employed in plagiarism detection and, in particular, in two main ways. First of all, in the case of external plagiarism detection, authentic corpora can be used as monitor collections against which to check the text under investigation. Secondly, corpora can be specifically created to include instances of plagiarism, with the aim of reaching advancements in the understanding of how to improve plagiarism research tools.

As regards the first approach, a text may be checked against collections of locally stored documents (as in the case of *PF*, which will be described in Section 3.2), general or specialized corpora, online databases, or the web. In the case of corpora, the choice of what kind of corpus should be used for plagiarism detection (monolingual/ multilingual, spoken/written, general/specialized, synchronic/diachronic, etc.) depends on the objectives and the resources available. For example, a corpus such as CADIS[5], Corpus of Academic English, compiled at the University of Bergamo, may be used for preliminary text comparisons. This corpus consists of academic texts written in English and is subdivided into four different disciplinary areas (i.e. applied linguistics, economics, law and medicine). According to the topic of the text under investigation, it is possible to limit the search to a certain discipline, therefore optimizing the search time.

In relation to other collections of texts, such as archives, databases or the web, it may be argued that they may not technically be seen as corpora. Indeed, they lack some of the defining characteristics of a corpus, such as representativeness, balance or size. However, as we shall see, for some specific purposes related to plagiarism detection, they may be used as corpora in the loose sense identified by McEnery and Wilson (2001: 197). For example, although not corpora in the traditional senses, databases provided by a university library system may also serve as monitor collections, and

5 For an overview see <http://www.unibg.it/cerlis/cadis>.

the same holds true for the web (for a comprehensive discussion of the issues related to considering the web as a corpus see Kilgarriff/ Grefenstette 2003).

The second way corpora may be used to investigate plagiarism is by specifically creating them for detection purposes. The PAN Plagiarism Corpus[6] is probably the most representative example. It contains documents in which plagiarism has been inserted automatically and manually in order to allow for the evaluation and the assessment of automatic plagiarism detection algorithms (Potthast *et al.* 2010). The PAN corpus contains cases of artificial and simulated plagiarism and its development is based on the principle that the construction of training corpora of this kind "can be automated, and hence be done on a large scale" (Potthast *et al.* 2010: 997). This corpus, being artificially created for the evaluation of automatic plagiarism detection tools (Barrón-Cedeño *et al.* 2010), is widely used for research purposes. It predominantly includes cases of monolingual plagiarism but some cross-linguistic examples are also present.

Working along the same lines as PAN, another collection was compiled to focus specifically on cross-linguistic plagiarism, namely, ECLaPA (Europer Cross-Language Plagiarism Analysis). This is an artificially compiled test collection which includes instances of cross-language plagiarism detection (see Pereira *et al.* 2010).

3.2. PF as an external plagiarism detection tool

If an evaluator (or self-evaluator) is aiming at a preliminary form of plagiarism search, a programme purely based on the identification of word chunk similarity may easily be created. With this objective in mind, we developed *PF*[7] (*PlagiarismFinder*) which works simply on

---

6 The latest version is PAN-PC-11, downloadable at http://www.uni-weimar.de/cms/medien/webis/research/corpora/corpus-pan-pc-11.html#c58437

7 The software is in its testing phase but will soon be available. For information: patrizia.anesa@unibg.it or giuseppe.parrinello@gmail.com.

text similarity detection. It was developed in C#[8] and the reference corpus may be composed of files in different formats[9]. It is primarily intended for external detection where the text is checked against a specific collection of texts or against corpora. There are no limits to the number of files which may be used as reference and the search can be carried out against specific files or entire folders or sub-folders.

Maurer *et al.* (2006: 1059) point out that there are three main approaches to detecting plagiarism:

- comparing a document to a body of texts (which may be stored locally);
- using a search engine;
- carrying out a stylistic analysis.

Following this categorization, I suggest here that the use of *PF* would fall within the first class. However, it can easily be combined with the other two approaches. Indeed, one can also check a part of text that is suspicious by using search engines, and external plagiarism detection can also be integrated with stylistic observations (Maurer *et al.* 2006: 1059).

PF lacks the benefits offered by more sophisticated programs (see Section 2) but may represent a fruitful tool in two main situations:

- to detect self-plagiarism in academic writing;
- to detect plagiarism in student assignments.

Detecting self-plagiarism may seem trivial in that authors should be fully aware of what they have previously published. However, cases of unintentional self-plagiarism are possible. For instance, when an author is working on several related publications, time pressure may lead to the unintentional duplication of sentences. Running a very

8 C# is a modern, object-oriented programming language developed by Microsoft within the .NET initiative.

9 While testing has been mainly based on txt files, processing can be done also using other formats, such as pdf files.

simple programme such as *PF* could prevent such unpleasant situations. Indeed, the text can effortlessly be checked against an author's collection stored locally, showing potential cases of duplication. Subsequent manual intervention can easily disambiguate the results.

The other main potential application of simple software such as *PF* lies in the educational field. It may certainly be argued that several institutions regularly require that their students submit their essays using a computerized system which automatically highlights potential cases of plagiarism. However, when such a system is not available, instructors can carry out a preliminary analysis by creating collections of students' assignments to be used as reference collections.

Locally stored collections are certainly homogenous regarding some of Sinclair's (2005) criteria to be considered in compiling a corpus, namely: mode, type, domain, language(s), location, and date. However, as has been mentioned, a collection of this type is not definable as a corpus in the strict sense (McEnery/Wilson 2001: 197). Therefore, in this case we are not talking about typical corpora but rather about *ad hoc* collections of texts.

Given its purposes, the creation of such collections does not generally require sampling. Moreover, the collections of texts which are used for plagiarism detection are not in line with the principle of representativeness described by Biber (1993: 243) as "the extent to which a sample includes the full range of variability in a population". It is clear that the texts collected for this specific purpose are not intended to fully depict a certain language or language variety. Indeed, in terms of size, an author's collection or a collection of students' assignments obviously cannot be representative of a language variety *in toto*. However, a collection of this type may include the total number of an individual's writings, and therefore it may even be seen as almost exhaustive in representing one's own variety. Similarly, a collection of student assignments may be considered as illustrative of a specific group's variety.

Representativeness should be also taken into account from a diachronic perspective in that "any corpus that is not regularly updated rapidly becomes unrepresentative" (Hunston 2002: 30). In line with the objectives of *PF*, the collection to be used as reference can simply

be updated including new texts (both in the case of self-plagiarism and student plagiarism).

As with internal detection, external plagiarism detection also has some limitations. For instance, the recognition of an identical passage taken from another text could be based simply on a citation which has been correctly reported. However, as stated above, false positives can easily be double-checked manually. Moreover, plagiarism in its concealed form is realized through reformulations, paraphrases and other modification processes. Therefore, a detector based on word chunk identity such as *PF* can certainly serve as a preliminary tool, in that it is free, fast and easy to use. However, for more in-depth analyses it should be combined with more sophisticated tools that are also based, for instance, on fingerprint indexing (see Barrón-Cedeño *et al.* 2010). Another limitation of an approach based on chunk identity is that it does not solve the problem of multilingual plagiarism. However, new methods have been developed to identify instances that involve different languages (e.g. MLPlag, as illustrated in Ceska *et al.* 2008). Such methods would go beyond the scope of *PF* but may be used for more refined detection procedures.

Other issues are related to the fact that, as happens with any automatic detection system, plagiarism concerning material that is only available in a paper format is not detected. Moreover, as mentioned in Section 1, works produced by ghost writers or through a paper mill may also be difficult to identify as unauthentic if they do not display real cases of plagiarism.

On a final note, it should be underlined that *PF* does not discriminate between intentional and unintentional plagiarism, where the latter is caused by an inability to deal with quotations and references or by genuine mistakes which may occur in the process of rephrasing or paraphrasing. However, no software product seems to deal satisfactorily with the issue of intentionality. Therefore, an evaluator's opinion is inevitably necessary.

As mentioned above, the notion of plagiarism, like those of intellectual creativity and originality, assume different contours in different cultures and depend on individual factors. Moreover, the interpretation of plagiarism in an educational context is also based on the concept of teacher reflexivity, intended as the reflexive role of the

teacher in the process of reading and interpreting a text (Sutherland-Smith 2008: 16), which may lead to different perceptions and interpretations of plagiarism.

It should also be noticed that the fight against plagiarism cannot be based solely on detection, but also on prevention. From a technical point of view, the use of passwords or encryptions in digital files can limit the ease of plagiarism practices. Secondly, as regards students, a clear policy focusing on the importance of originality as well as on the measures to be adopted in the case of plagiarism could be used as a deterrent. Indeed, as Emerson (2008: 193) aptly remarks, "the power of detection does not compensate for the breaking of the relationship between student and teacher" and detection should be accompanied by adequate prevention policies.

4. Conclusions

The overarching purpose of this chapter was twofold: to discuss potential systems to detect phenomena of plagiarism and self-plagiarism in writing and, more specifically, to observe how corpora could be used in this perspective. Firstly, I observed that different kinds of authentic corpora could be employed for text-external detection. Moreover, artificially created plagiarism corpora can be used in order to gain further knowledge in the area of automated plagiarism detection research and for the evaluation of current detection tools.

As regards plagiarism detection, I have argued for a multi-methodological approach which combines automatic and manual systems. In particular, in terms of automatic detection, this chapter highlights the necessity to check documents against self-plagiarism by using a locally stored collection of one's texts, in order to avoid possible cases of accidental self-plagiarism. I also observed the usefulness of attempting to detect other-plagiarism by using corpora or other collections not technically definable as corpora, such as archives or databases.

More specifically, I presented a simple detection program, *PF*, which may be used to check the document under investigation against a locally stored collection. When dealing with plagiarism detection, a web retrieval scenario is certainly more exhaustive, but, as has been mentioned, *PF* is merely intended as a preliminary step in the detection process. Moreover, if the aim is to avoid self-plagiarism, an author's collection can represent an exhaustive monitor collection. Similarly, a collection of assignments may be sufficient if the objective is to carry out a preliminary analysis of student assignments to prevent intra-class or inter-class plagiarism.

Ethical issues lie at the heart of plagiarism. Some scholars have argued against the use of detection software following the principle that it is unethical in that, essentially, it makes money from a crime. However, a simple tool like *PF* is completely free and therefore issues of this kind are minimized, and, despite its limitations, it represents a useful resource for preliminary detection.

To sum up, we can argue that no plagiarism detection program can be used in isolation. First of all, a good combination of automatic and manual practices is desirable. Secondly, in terms of automatic detection, different tools may be used according to the specific objectives and the resources available. Simple word chunk identity detection may be seen as a preliminary step towards the avoidance of self- or other-plagiarism in student assignments. However, when dealing with the education field, it should be noted that the attempt to limit plagiarism among students cannot be based solely on detection systems but should start with a good teaching approach that focuses on the consequences and the moral issues related to this practice. Similarly, in the case of self-plagiarism, automated systems cannot replace a careful and meticulous analysis of one's texts.

Further research in this field should also focus on qualitative analysis involving authors and students whose texts present some forms of plagiarism or self-plagiarism in order to gain a finer understanding of the reasons and the practices lying behind this phenomenon.

References

Anderson, Judy 1998. *Plagiarism, Copyright Violation and Other Thefts of Intellectual Property*. Jefferson: McFarland.

APA 2010. *The Publication Manual of the American Psychological Association*. 6th ed. Washington, D.C.: American Psychological Association.

Arwin, Christian / Tahaghoghi, Seyed M.M. 2006. Plagiarism Detection across Programming Languages. In Estivill-Castro, Vladimir / Dobbie, Gillian (eds) *Proceedings of the 29th Australasian Computer Science Conference* 48, Darlinghurst: Australian Computer Society, 277-286.

Barrón-Cedeño, Alberto / Potthast, Martin / Rosso, Paolo / Stein, Benno 2010. Corpus and Evaluation Measures for Automatic Plagiarism Detection. In Calzolari, Nicoletta / Choukri, Khalid / Maegaard, Bente / Mariani, Joseph / Odijk, Jan / Piperidis, Stelios / Rosner, Mike / Tapias, Daniel (eds) *Proceedings of the International Conference on Language Resources and Evaluation, LERC 2010*. Valletta: European Language Resources Association, 771-774. Retrieved from: http://users.dsic.upv.es/~prosso/resources/BarronEtAl_LREC10.pdf

Biber, Douglas 1993. Representativeness in Corpus Design. *Literary and Linguistic Computing* 8/4, 243-257.

Bloch, Joel 2012. *Plagiarism, Intellectual Property and the Teaching of L2 Writing*. Bristol: Multilingual Matters.

Buranen, Lise 1999. But I wasn't Cheating: Plagiarism and Cross-cultural Mythology. In Buranen, Lise / Roy, Alice (eds) *Perspectives on Plagiarism and Intellectual Property in a Postmodern World*. NY: State University of New York.

Campbell, Cherry 1990. Writing with Others' Words: Using Background Reading Text in Academic Compositions. In Kroll, Barbara (ed.) *Second Language Writing*. Cambridge: Cambridge University Press, 211-230.

Ceska, Zdenek / Toman, Michal / Jezek, Karel 2008. *Multilingual Plagiarism Detection*. Berlin: Springer-Verlag.

Clough, Paul / Gaizauskas, Robert 2009. Corpora and Text Re-use. In Lüdeling, Anke / Kytö, Merja (eds) *Corpus Linguistics: An International Handbook*. Berlin: Mouton de Gruyter. 1249-1271.

El Bachir Menai, Mohamed 2012. Detection of Plagiarism in Arabic Documents. *International Journal of Information Technology and Computer Science* 10, 80-89.

Emerson, Lisa 2008. Plagiarism, a Turnit Trial, and an Experience of Cultural Disorientation. In Eisner, Caroline / Vicinus, Martha (eds) *Originality, Imitation, and Plagiarism.* Ann Arbor: The University of Michigan Press, 183-194.

Holmes, David I. 1998. The Evolution of Stylometry in Humanities Scholarship. *Literary and Linguistic Computing* 13/3, 111-117.

Howard, Rebecca 2007. Understanding Internet Plagiarism. *Computers and Composition* 24/1, 3-15.

Hunston, Susan 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hunston, Susan 2006. Corpus Linguistics. In Keith, Brown (ed.) *The Encyclopedia of Language and Linguistics*. Boston, MA: Elsevier, 234-248.

Jung, Carl Gustav 1970[1905]. Cryptomnesia. *Collected Works*, Vol. 1. London: Routledge and Kegan Paul, 95-106.

Kilgarriff, Adam / Grefenstette, Gregory 2003. Introduction to the special issue on the web as corpus. *Computational linguistics* 29/3, 333-347.

Maurer, Hermann / Kappe, Frank / Zaka, Bilal 2006. Plagiarism: A Survey. *Journal of Universal Computer Science* 12/8, 1050-1084.

McEnery Tony / Wilson Andrew 2001. *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

Meyer zu Eissen, Sven / Stein, Benno / Kulig, Marion 2007. Plagiarism Detection without Reference Collections. In Decker, Reinhold / Lenz, Hans-Joachim (eds) *Advances in Data Analysis*. Berlin: Springer, 359-366.

Murray, Laura J. 2008. Plagiarism and Copyright Infringement: The Costs of Confusion. In Eisner, Caroline / Vicinus, Martha (eds) *Originality, Imitation, and Plagiarism.* Ann Arbor: The University of Michigan Press, 173-182.

Pecorari, Diane 2003. Good and Original: Plagiarism and Patchwriting in Academic Second-language Writing. *Journal of Second Language Writing* 12, 317-345.

Pereira, Rafael C. / Moreira, Viviane / Galante, Renata 2010. A New Approach for Cross-Language Plagiarism Analysis. In Agosti, Maristella / Ferro, Nicola / Peters, Carol / de Rijke, Maarten / Smeaton, Alan (eds) *Proceedings of the CLEF 2010 Conference on Multilingual and Multimodal Information Access Evaluation*. Padua: Springer, 15-26.

Posner, Richard 2007. *The Little Book of Plagiarism*. New York: Pantheon.

Potthast, Martin / Stein, Benno / Barrón-Cedeño, Alberto / Rosso, Paolo 2010. An Evaluation Framework for Plagiarism Detection. *23rd International Conference on Computational Linguistics (COLING 10)*. Retrieved from: http://www.uni-weimar.de/medien/webis/publications/papers/stein_2010p.pdf

Sinclair, John 2005. Corpus and Text - Basic Principles. In Wynne, Martin (ed.) *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, 1-16.

Stamatatos, Efstathios 2009. Intrinsic Plagiarism Detection Using Character n-gram Profiles. In *Proceedings of the 3rd International Workshop on Uncovering Plagiarism, Authorship, and Social Software Misuse*. Retrieved from: http://ceur-ws.org/Vol-502/paper8.pdf.

Stepchyshyn, Vera / Nelson, Robert 2007. *Library Plagiarism Policies*. Chicago: American Library Association.

Sutherland-Smith, Wendy 2008. *Plagiarism, the Internet, and Student Learning: Improving Academic Integrity*. New York: Routledge.

van den Berk, Tjeu 2012. *Jung on Art: The Autonomy of the Creative Drive*. New York: Routledge.

GABRIELLA CAROBBIO / DOROTHEE HELLER / CLAUDIA DI MAIO[1]

3. Zur Verwendung von Frageformulierungen im Korpus *euroWiss*

1. Einleitung

„Linguisten brauchen Korpora und Korpora Linguisten“ – so leitete Ludwig Eichinger den Band der 42. IDS-Jahrestagung zum Thema *Sprachkorpora* ein. Für die Beobachtung sprachlicher Phänomene in Texten und Diskursen bieten korpuslinguistische Tools ein zunehmend differenziertes Spektrum von Möglichkeiten der systematischen Erfassung und Aufbereitung von Daten. Für ihre Deutung bleiben qualitative Analysen jedoch weiterhin unverzichtbar.

Unsere Fragestellung hat sich aus einer qualitativen Analyse des *euroWiss*-Korpus heraus entwickelt, das in Sektion 2 ausführlicher beschrieben wird. Bei der Auswertung videographisch aufgezeichneter universitärer Lehrveranstaltungen wurde der häufige Gebrauch von Frageformulierungen augenfällig. Ihrem funktionalen Einsatz nachzugehen, ist Anliegen des Beitrags.

Fragen stellen eine Form sprachlichen Handelns dar, das die Lehrpraxis in allen Stufen der Schul- und Hochschulausbildung kennzeichnet (Ehlich/Rehbein 1979, 1986; Deckwerth 1983; Nowak/Macht 1996; Petersen/Sommer 1999; Wiesmann 1999, 2001; Bamford 2000). Bemerkenswert bei akademischen Vorlesungen ist, dass Fragen meistens in Sprechhandlungsverkettungen vorkommen, d.h. innerhalb des „monologischen“ Vortrags des Dozenten, der die jeweilige Lehrveranstaltung (LV) hält. Dies führt in vielen Fällen dazu, dass auch die auf die Fragen bezogenen Antworten vom Dozenten selbst gegeben wer-

1 Claudia Di Maio (geb. Zech).

den. Um die Rolle solcher „selbst beantworteten Fragen“ im akademischen Lehr-Lern-Diskurs, die hier unter dem Oberbegriff Frageformulierungen subsumiert werden, soll es im vorliegenden Beitrag gehen[2]. Dabei wird eine komparatistische Perspektive eingenommen, indem Beispiele aus deutschen und italienischen Vorlesungen verschiedener Disziplinen herangezogen werden. Die Analyse beruht auf einem breit angelegten Korpus von Audio- und Videoaufzeichnungen universitärer LV, das im Rahmen des internationalen Forschungsprojektes „*euroWiss* – Linguistische Profilierung einer europäischen Wissenschaftsbildung“ zusammengestellt wurde.

2. *euroWiss*: Projektvorstellung und Untersuchungsdesign

Das Forschungsprojekt *euroWiss* wird seit März 2011 und für die Dauer von drei Jahren im Rahmen der Initiative „Deutsch Plus – Wissenschaft ist mehrsprachig“ von der Volkswagen Stiftung gefördert. Projektbeteiligte sind die Universität Hamburg (Prof. Dr. Angelika Redder, Projektleitung), die Technische Universität Chemnitz (Prof. Dr. Winfried Thielmann), die Università degli Studi di Bergamo (Prof. Dr. Dorothee Heller), in Kooperation mit der Università degli Studi di Modena e Reggio Emilia (Prof. Dr. Antonie Hornung).

Abb. 1. Das Forschungsprojekt *euroWiss* (Logo).

2 Die Inhalte des Beitrags wurden von den Autorinnen gemeinsam erarbeitet, Sektion 1-5 und 8 wurden von Gabriella Carobbio verfasst, Sektion 6 von Dorothee Heller und Sektion 7 von Claudia Di Maio. Wir danken Alessandra

Zentrales Anliegen des *euroWiss*-Projekts ist die Erforschung der Strukturen sprachlichen Handelns in der wissensvermittelnden Hochschulkommunikation, insbesondere der binneneuropäischen Differenzen zwischen simultaner und sukzessiver Erprobung und Praktizierung diskursiver Wissensvermittlung[3] in universitären LV. Das Projekt geht unter anderem folgenden Leitfragen nach: Wie wird wissenschaftliches Wissen an deutschen und italienischen Universitäten vermittelt? Wie werden Lernende im Prozess der Wissensvermittlung einbezogen? Lassen sich länder- bzw. disziplinspezifische Unterschiede beobachten? Die Untersuchungen des Materials, das an den jeweiligen Standorten zusammengestellt wurde, sind empirisch basiert und stützen sich auf Videographien (Audio- und Videoaufzeichnungen von LV), Befragungen (Fragebögen und Interviews an Studierende) und studentische Seminararbeiten/Mitschriften.

Bei der Zusammensetzung der Teilkorpora stellt das Korpus an Videographien den größten Teil der Daten. Befragungen und Seminararbeiten/Mitschriften wurden als kleinere Nebenkorpora angelegt, um das Bild zu vervollständigen, das sich aus den Analysen der Videographien ergibt. Fragebögen und Interviews geben Einblicke in die Einschätzung des Studiums seitens der Studierenden und ihre Bewertungen etwaiger Studienerfahrungen im Ausland. Durch die Analyse studentischer Schriftproduktionen wie Mitschriften wird u.a. der Frage nachgegangen, wie mündlich vermitteltes Wissen schriftlich umgesetzt wird.

Alghisi und Elisa Bolpagni für die sorgfältige Mitarbeit bei der Aufnahme, Transkription und Aufbereitung der Daten.

3 S. Redder (2009: 24): „Die diskursive Wissensvermittlung gilt als „typisch deutsch“. Sie ist nicht einfach mit der eristischen Struktur wissenschaftlicher Wissensgewinnung zu vergleichen oder gar zu identifizieren. Vielmehr stellt die diskursive Wissensvermittlung den didaktischen Versuch dar, den Lernprozess so weit wie möglich durch produktive Nutzung des bei den Lernenden bereits vorhandenen Wissens zu gestalten, also Aneignungsstrukturen zu aktivieren, die durchaus mit dem Problemlösungsprozess des Forschens gemeinsame Elemente aufweisen“.

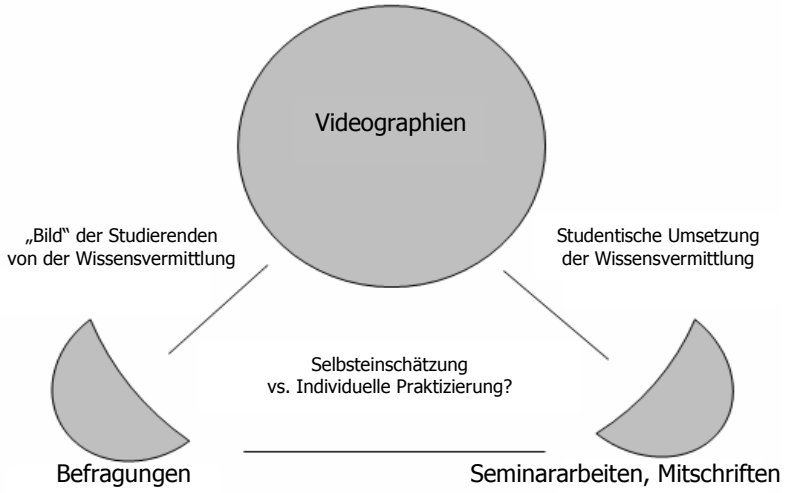


Abb. 2. Untersuchungsdesign von *euroWiss*.

Empirische Grundlage des vorliegenden Beitrags stellt das Videographie-Korpus dar. Es besteht aus Audio- und Videoaufzeichnungen verschiedener Veranstaltungsformen (Vorlesungen, Seminare, Übungen, Praktika und Kolloquien) an deutschen und italienischen Universitäten aus der Zeit vom Sommersemester 2011 bis Sommersemester 2012. Es handelt sich also um authentische Ausschnitte aus dem Alltag der universitären Lehre in beiden Ländern. Es wurden insgesamt ca. 346 Stunden Videomaterial von 225 Sitzungen aus 62 LV zusammengestellt. Von allen LV wurden mindestens drei, zumeist aufeinander folgende, Sitzungen aufgezeichnet und ein Aufnahmeprotokoll erstellt. In allen Fällen konnte zumindest der Lehrende videographiert werden – in Seminaren, Übungen, Kolloquien und Praktika wurden zumeist auch die Studierenden mitgefilmt.

Das Korpus ist transdisziplinär angelegt. Vertreten sind nicht nur die Geisteswissenschaften, sondern auch die Wirtschafts-, Sozial- und Naturwissenschaften. Angestrebt wurde eine Darstellung der verschiedenen Phasen des Studiums mit Daten aus Bachelor-, Master- und Promotionsstudiengängen.

Nach der Datenerhebung und -aufbereitung wurden alle Videoaufzeichnungen phasiert und der Ablauf jeder aufgenommenen Sitzung in Hinblick auf Turn-Taking-Phänomene und Modalitäten der

Wissensvermittlung beschrieben (Planungshoheit der LV bei Lehrenden bzw. Lernenden; Umsetzung bestimmter sprachlicher Handlungen wie das Referieren, das Problematisieren, das kooperative Erarbeiten der zu vermittelnden Inhalte). Die Phasierung ermöglicht eine auf die Projektfragestellungen ausgerichtete Übersicht auf das gesamte Korpus und erlaubt die Identifizierung der zu transkribierenden Stellen. Im Anschluss daran erfolgte die Transkription der ausgewählten Videoausschnitte gemäß den HIAT-Transkriptionskonventionen und mit Hilfe der Partitur-Software EXMARaLDA. Den theoretischen und methodischen Rahmen der Transkriptanalyse[4] bildet das Instrumentarium der Funktionalen Pragmatik (s. zum Beispiel: Ehlich/Rehbein 1979; Ehlich 2007; Redder 2008).

3. Schwerpunktsetzung des Beitrags

Um Formen der diskursiven Wissensvermittlung im universitären Lehr-Lern-Diskurs zu ermitteln, wurden im Rahmen des *euroWiss*-Projekts folgende Aspekte als relevant betrachtet:

A. die studentenseitige Beteiligung

B. die dozentenseitige Einbeziehung der Lernenden

C. die Struktur des verbalisierten Fachwissens

Auf die ersten zwei Aspekte (A+B) wurde im Rahmen von Studien zur Hörerinvolvierung in Vorlesungen zur Literaturwissenschaft exemplarisch eingegangen (vgl. Carobbio/Zech 2013; Heller 2014).

4 Verwiesen sei insbesondere auf die Projektbände von Hornung/ Carobbio/ Sorrentino (2014) und Redder/ Thielmann/ Heller (i.E.). Für eine vollständige Übersicht der Vorträge und Publikationen der Projektmitglieder vgl. die Webseite des Projektes <http://www1.slm.uni-hamburg.de/de/forschen/ projekte/ eurowiss.html> (letzter Abruf September 2013).

Bei einem Vergleich einer deutschen Vorlesung zur Romanistik und einer italienischen Vorlesung zur Germanistik wurden verschiedene Phänomene sichtbar, die unterschiedliche Formen der Einbeziehung der Studierenden in den Lehr-Lern-Diskurs zeigen. Darunter fallen nicht nur Hörerbeiträge bei Turn-Taking-Sequenzen, sondern auch dozentenseitige sprachliche Verfahren wie der Einsatz vorgestellter Rede, Rückversicherungssignale und Frageformulierungen. Als besonders auffällig angesichts ihrer facettenreichen Verwendung in Vorlesungen erwiesen sich dozentenseitige „Formulierungen als Frage" ohne Sprecherwechsel, wie etwa in folgenden Beispielen (aus Carobbio/Zech 2013):

(1) „Und weshalb beschäftigen wir uns damit? Es ist natürlich …"

(2) „Di che cosa parla quasi sempre Böll nella sua vastissima opera? ..."

Diese Fragen werden vom jeweiligen Lehrenden im Laufe der Vorlesung gestellt und anschließend beantwortet. Sie unterbrechen also den Redefluss des Dozenten nicht und eröffnen keine Turn-Taking-Sequenz mit den Lernenden. Es bietet sich daher an, genauer zu betrachten, wie solche Frageformulierungen realisiert werden und was sie funktional im Lehr-Lern-Diskurs leisten.

Abb. 3. Identifizierung der Schwerpunktsetzung des Beitrags – qualitatives Vorgehen.

Die Graphik zeigt, wie die Schwerpunktsetzung des vorliegenden Beitrags durch eine qualitativ orientierte Herangehensweise identifiziert wurde. Die Analyse exemplarischer Belege aus dem *euroWiss*-Videographie-Korpus erfolgt unter besonderer Berücksichtigung ihrer Rolle für die Wissensvermittlung. Der Rückgriff auf Korpusdaten ermöglicht es, „sprachliche Empirie zu dokumentieren" (Eichinger 2007: 2) und dabei auch die jeweiligen Handlungskontexte einzubeziehen.

4. Ausgangspunkt: Das Frage-Antwort-Muster (Ehlich/Rehbein 1979)

In einer Handlungskonstellation, bei der Sprecher (S) und Hörer (H) über ungleiche, „asymmetrische" Wissensbestände verfügen, kommt es häufig zur Umsetzung des sogenannten Frage-Antwort-Musters[5] (Ehlich/Rehbein 1979). S kann aufgrund einer ihm vorliegenden Wissenslücke eine Frage[6] an H richten, um sich das ihm fehlende Wissenselement anzueignen. Der Frage geht eine Überlegung bzw. eine Einschätzung von S voraus, ob H über das gewünschte Wissen verfügt und somit diese Wissenslücke schließen kann. Die Frage von S löst einen Suchprozess im mentalen Bereich von H aus, der nach dem erfragten Wissenselement in seinem Wissensbereich sucht und gegebenenfalls die Antwort gibt.

Eine Visualisierung des Frage-Antwort-Musters bietet die folgende Graphik:

---

5 Sprachliche Handlungsmuster sind „Formen von standardisierten Handlungsmöglichkeiten, die im konkreten Handeln aktualisiert und realisiert werden" (Ehlich/Rehbein 1979: 250).

6 Formal zeichnet sich die Sprechhandlung der Frage durch besondere Verfahren der Wortstellung und Intonation aus.

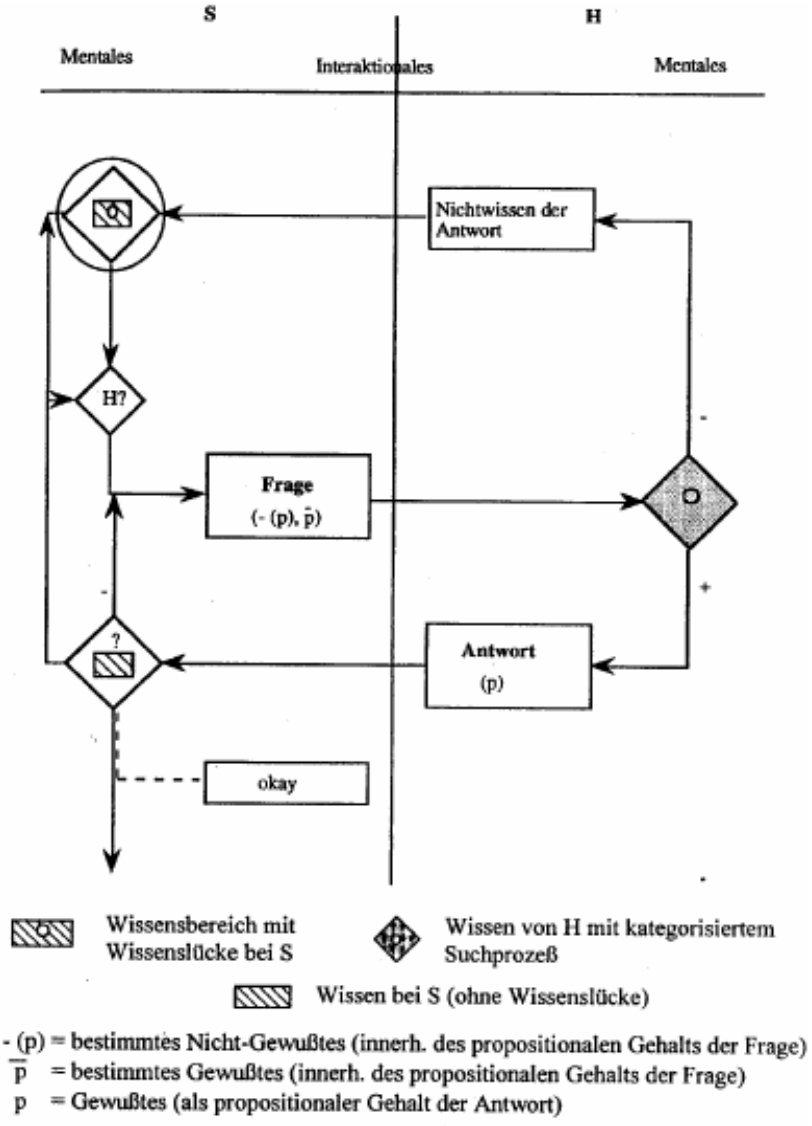


Abb. 4. Das Frage-Antwort-Munster (Redder 1998:76).

Das Frage-Antwort-Muster dient also der Erarbeitung von Wissensdefiziten und der Abgleichung/Synchronisierung des Kenntnisstandes

der an der Kommunikation beteiligten Aktanten. Zu berücksichtigen ist dabei, wie sich die Aktanten im Handlungsmuster als Wissend oder Nicht-Wissend positionieren. Wie oben gezeigt, hat S normalerweise eine Wissenslücke, die von H kompensiert wird.

Bei Lehr-Lern-Diskursen ist genau das oft nicht der Fall. Die handlungstheoretische Forschung hat zeigen können, wie das Mittel der Frage im Schuldiskurs auch anders funktionalisiert werden kann. Ehlich/Rehbein (1986) führen z.B. den Begriff der „Regiefragen" ein, um einen großen Teil der Lehrerfragen zu beschreiben. LehrerInnen im Unterricht stellen Fragen nicht, um sich fehlendes Wissen anzueignen, sondern um die Wissensprozessierung bei ihren Schülerinnen zu steuern[7]. Auch sogenannte „Paraphrasefragen" in Prüfungsgesprächen[8] gehen von keinem Wissensdefizit von S aus, sondern dienen der Verdeutlichung einer Frage bei H.

5. Frageformulierungen in universitären Lehrveranstaltungen: exemplarische Analysen

Formulierungen, die in formaler Hinsicht Merkmale von Fragen aufweisen (z.B. Wortstellung, Intonation, Interpunktion) haben daher illokutiv eine z.T. andere Handlungsqualität bzw. Funktion. Sie dienen nicht dazu, vorhandene Wissenslücken von S zu füllen, sondern S legt vielmehr seinen gesamten Handlungsplan und dessen Ausführung so an, dass die Hörer zu neuem Wissen gelangen.

In vielen Fällen zeigt sich, dass sich die Frageformulierungen mit besonderer Zweckbestimmung und an spezifischen Punkten im

---

7 Ehlich/Rehbein (1986: 81ff.) sprechen in diesem Sinne auch von „Lehrervortrag mit verteilten Rollen".

8 Der Terminus „Paraphrasefragen" wurde von Fierbinţeanu im Rahmen einer von ihr vorgestellten umfangreichen Typologisierung der „Fragen in mündlichen Prüfungsgesprächen an der Hochschule" eingeführt (Vortrag IDT 2013 – Sektion G6 „Wissenschaftssprache Deutsch").

Dozentenvortrag finden, und zwar unmittelbar vor anderen Sprechhandlungen, zum einen:

Vor Begründungen und Erläuterungen (A):

(3) Nun hatte Aristoteles dies so definiert und dann im zweiten Kapitel von De anima zwei gesagt, dass diese Definition nicht zufriedenstellend ist. **Warum nicht**? Weil sie nichts Konkretes erklärt. (Phil_V_1-4, Fl. 26-29)

(4) **Perché** questo c'entra con la lingua? **Perché** in realtà ritorna il discorso del silenzio. (GerSpr_K_5, Fl. 264)

(5) **Was heißt** funktionale Einkommensverteilung? Funktional ist eben der Bezug… (VWL_V_1, Fl. 17)

(6) **Cosa vuol dire** che il manoscritto medievale si presenta come un codice? E infatti si chiamano normalmente codici manoscritti, hm? ((2s)) **Eh vuol dire** semplicemente che… (RomSprGe_V_2, Fl. 63f.)

Solche Fragen bereiten das Terrain für nachfolgende Ausführungen oder Vertiefungen vor und lösen – idealerweise – mentale Suchprozesse bei H aus.

Vor Zusammenfassungen (B):

(7) Was wären jetzt **also** zentrale Aspekte der Frauenbildforschung? ((2s)) Die Grundannahme ist, dass… (RomLit_V_23, Fl. 991f.)

(8) **Allora**, che cos'è la sua estetica dell'umano? (GerLit_V_2, Fl. 439)

In den Beispielen (7) und (8) leitet der Lehrende eine Rekapitulation des Vorhergesagten im Frage-Modus ein und bringt somit seine bisherigen Ausführungen auf den Punkt. Auch die paraoperativen Prozeduren „also" und „allora" signalisieren, dass hier sozusagen die Fäden des Diskurses zusammengezogen werden.

In anderen Fällen werden Frageformulierungen *als Aufforderung zum Mitdenken* funktionalisiert *(C):*

(9) Effettivamente che cosa fa insomma? Guardiamo un attimo alla professione scelta. Che cosa fa un clown che poi è mimo anche lui un po'? **Avete presente**? Non farà mai un Theaterstück di due ore, eh? (GerLit_V_2, Fl. 655f.)

(10) Ich würde jetzt vorschlagen, dass Sie sich – also die Begriff hängen alle irgendwie zusammen –, dass **wir uns kurz n paar Minuten Zeit nehmen** ((1s)) um Assoziationen zu sammeln. **Also** was/was assoziieren Sie mit diesen drei Begriffen? Wie hängen die vielleicht zusammen? (VWL_V_1, Fl. 408ff)

Im deutschen Korpus lässt sich beobachten, dass auf Fragen vom Typus (C) häufig Antworten seitens der Studierenden erfolgen. Fragen dieser Art lösen also einen Hörereinbezug mit Interaktion aus. Die Aufforderungen sind also nicht nur auf ein Mitdenken, sondern auch auf ein Mitmachen ausgerichtet[9]. Charakteristisch ist dabei, dass wie in Beispiel (10) jeweils mehrere Fragen dieses Typus hintereinander geschaltet werden. Während der wiederholte Einsatz dieser Frageformulierungen in den deutschen Vorlesungen auf eine aktive Beteiligung der Studierenden abzielt, lässt sich diese Ausrichtung in den uns vorliegenden Daten aus LV in Italien nichts Vergleichbares beobachten.

Anders als bei rein „rhetorischen" Fragen, die auf einer Synchronizität der mentalen Bereichen der Aktanten operieren (die Antwort ist für beide S und H selbstverständlich), gehen die hier untersuchten Frageformulierungen von einer Diskrepanz zwischen den Wissensbeständen der zwei Akteure aus. Die Wissenslücke liegt nicht bei den Fragenden S vor, sondern bei H. S versucht, durch die Fragen auf eine progressive Annährung bzw. eine Bearbeitung dieser Wissenslücke hinzuarbeiten. Sie versuchen den Rezeptionsprozess und das Verstehen bei H zu steuern und lassen sich daher funktional als *verstehensorganisierende Fragen* charakterisieren[10].

---

9 Aus diesem Grund kann Beispiel (10) möglicherweise auch als Umsetzung des Musters Aufgaben-Stellen/Aufgaben-Lösen betrachtet werden (Ehlich/ Rehbein 1986).

10 Wir greifen hier den Bezeichnungsvorschlag von Angelika Redder auf, den sie im Rahmen einer *euroWiss*-Arbeitssitzung auch in Hinblick auf die Abgrenzung von rhetorischen Fragen entwickelt hat. Ähnlich den Verstehens-

6. Verstehensorganisierende Fragen im Diskurszusammenhang

Der nachfolgende Transkriptausschnitt entstammt einer Lehrveranstaltung aus der Germanistischen Linguistik zum Thema Textlinguistik und Übersetzen. Im Mittelpunkt der aufgezeichneten Sitzung steht die Frage der Evaluierung von Übersetzungen, in Hinblick auf eine textarten- und zielsprachengerechte Übertragung des Ausgangstextes. Die Sitzung hat zum Teil Vorlesungscharakter, wenn die Dozentin beispielsweise unterschiedlich gelungene Übersetzungen touristischer Texte kommentiert. Zum Teil nimmt sie aber auch Seminarcharakter an, wenn die Dozentin die Studierenden Übersetzungsproblematiken selbst an den Texten erarbeiten lässt. Das Bestreben der Dozentin geht auch dahin, durch die Gegenüberstellung unterschiedlicher Übersetzungsvorschläge den Studierenden Unterschiede auf stilistischer Ebene bewusst zu machen und sie dadurch für die Möglichkeiten eines freieren Umgangs mit dem Ausgangstext zu sensibilisieren.

In dem hier wiedergegebenen Transkriptausschnitt greift die Dozentin eine Textstelle aus Hemingways *The old man and the sea* auf, die bereits Gegenstand einer vorausgehenden Sitzung war:

> When there are no hurricanes, the weather of hurricane months is the best of all the year. If there is a hurricane you always see the signs of it in the sky for days ahead, if you are at sea. (The old man and the sea)

Sie diskutiert in diesem Zusammenhang insbesondere die Wiedergabe des Satzes: *If there is a hurricane you always see the signs of it in the sky for days ahead, if you are at sea*, indem sie auf folgende

---

fragen mit „denn“ (Redder 1990: 67), die das „bestimmte Nicht-Gewußte speziell als Verstehensdefizit bearbeiten“, gehen auch verstehensorganisierende Fragen von (möglichen) hörerseitigen Verstehensproblemen aus, die allerdings nicht direkt vom Hörer ausgedrückt, sondern vom Dozenten *antizipierend* bearbeitet werden.

Übersetzungsvarianten Bezug nimmt, die sie dem Vademecum für Übersetzer von Macheiner (1995) entnommen hat:

> Ü1 … sieht man immer Anzeichen davon am Himmel Tage vorher, wenn man auf See ist.
> Ü2… sieht man immer/schon Tage vorher Anzeichen davon am Himmel, wenn man auf See ist.
> Ü3… sieht man schon Tage vorher am Himmel Anzeichen davon, wenn man auf See ist.
> Ü4… sieht man, wenn man auf See ist, schon Tage vorher am Himmel Anzeichen davon.
> Ü5… sieht man auf See schon Tage vorher am Himmel die Anzeichen davon
> Ü6… sieht man das auf See schon Tage vorher am Himmel.
> Ü7… kann man das auf See schon Tage vorher am Himmel sehen.

Bei dem Kommentar der Varianten – vor allem des letzten Vorschlags (Ü7) – knüpft die Dozentin zunächst an Beobachtungen aus einer zurückliegenden Sitzung an.

[224]

Df So, • • • können wir ((2s)) äh • • • zurück zu unseren

[225]

Paraphrasen des äh Hemingway-•-Satzes • • • kommen(?)

[226]

Äh was haben wir bemerkt, • letztes Mal? In dieser letzten

[227]

Paraphrase wird ein • Modalverb hin/ eingefügt, ja? • • und

[228]

dieses Modalverb • haben wir im Originalsatz • nicht, • • • das

[229]

Verb können, hm? **Können (Sie) sich • erinnern? • • Ja.**

[230]
114
Äh aus (Ganzem) gesehen ist der gleiche Inhalt • gewährleistet,

[231]
..
115 [21:12.7]
ja? Also, diese • letzte Übersetzung — wenn auch so

[232]
..
unterschiedlich vom Originalsatz — • • äh • • vermittelt

[233]
..
denselben • • • propositionalen Inhalt wie • der Originalsatz, ok?

[234]
..116
Äh es wurde keine Information verloren, hm? • in der

[235]
..
117
Übersetzung. **Wo liegt also hier das Problem, hm? • •**
118

[236]
..
Äh ((2,5s)) stilistisch gesehen • • ist diese Paraphrase, nach

[237]
..
Monika Doherty, • für einen deutschsprachigen Leser auch • die

[238]
..
119
beste. ◡Also, äh • • äh die Autorin sagt: Das ist die

[239]
..
Paraphrase, die mir am besten, • also die mir viel besser gefällt,

[240]
..
120 [21:56.3]
äh? (wie) die anderen Paraphrasen. **Warum es so ist? • •**

[241]
..
Das kann • sie nur • • äh • • mit Bezug auf den Satzrhythmus

[242]
.. 121
erklären. Also, der Satz klingt • • einem deutschen Ohr einfach

[243]
..
• besser, hm? • • • viel geschickter • als die anderen, ok? 122 Aber

[244]
..
äh • dennoch haben wir im Fall des Modalverbs ein Problem.

[245]
123
((1,5s)) **Wo ist hier das Problem?** • • Äh • • auch ohne

[246]
..
Modalverb, hm? — • • • wie in der Paraphrase • • • Ü sechs,

[247]
..
hm? — • • • ist die Verständlichkeit • • des Originalsatzes

[248]
.. 124
gewährleistet. Also, wir könnten auch ohne Modalverb/ • •

[249]
..
also, ohne Modalverb funktioniert die Übersetzung eigentlich, • •

[250]
..
ja? Das brauchen wir nicht, • • um • • den Text, äh? den Inhalt

[251]
.. 125 [22:54.2]
zu verstehen, ja? **Äh ((2,5s)) also, • • warum/ • • • äh was**

[252]
..
gibt uns das Recht auf dieses zusätzliche Element, • • • wenn

[253]
..
die Verstehbarkeit praktisch schon • garantiert ist? ((3s)) Das

[254]
..
hat • mit der Spezifik für erfolgreiche Sprachverarbeitung im

[255]
..
Englischen und im Deutschen • zu tun, • • • ja?

Der Transkriptauszug enthält sechs Fragen, die das Vorgehen der Dozentin veranschaulichen, durch Frageformulierungen das Verstehen der Studierenden gezielt zu fördern:

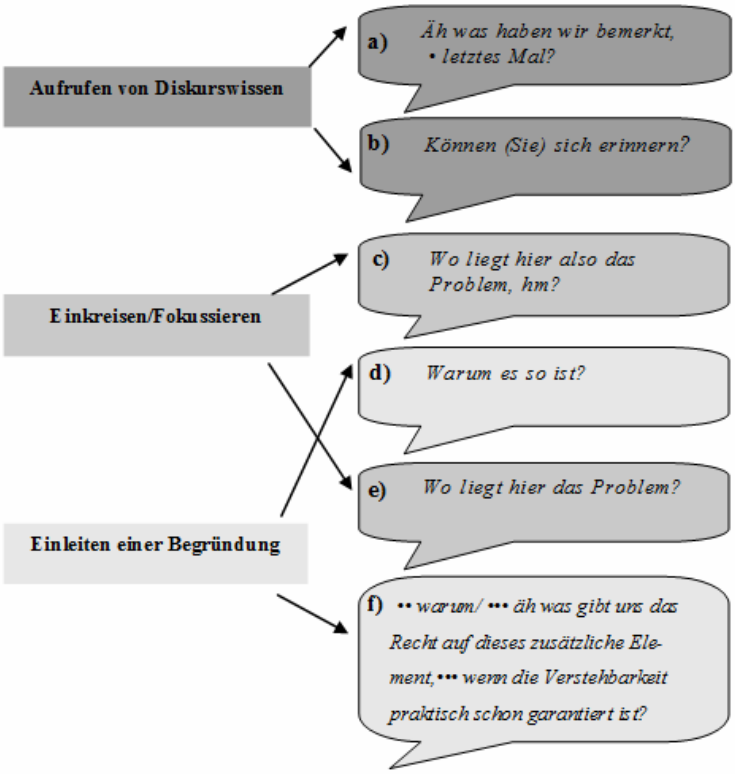


Abb. 5. Funktionen verstehensorganisierender Fragen im Lehr-Lern-Diskurs.

Es werden zunächst Wissenselemente aus dem Vorlesungsdiskurs aufgerufen (a), die aufgrund der vorausgegangenen Sitzungen vorausgesetzt werden können. Im Anschluss daran lenkt die Dozentin die Aufmerksamkeit auf eine Besonderheit von Ü7 (Hinzufügung eines Modalverbs). In einem zweiten Schritt werden die Studierenden wiederum durch eine Frageformulierung (b) aufgefordert, sich weitere Beobachtungen in Erinnerung zu rufen (hier: trotz Hinzufügung des Modalverbs ergeben sich keine Änderungen auf der propositionalen Ebene). Bei diesem Bemühen, schrittweise ein Verstehen der Studierenden zu bewirken bzw. zu fördern, bedient sich die Dozentin nicht nur dem Verfahren der selbstbeantworteten Fragen, die zum Teil

wiederholt werden (c+e), sondern verschiedener Umformulierungen bei deren Beantwortung. Solche Reformulierungen in Form von Erläuterungen und Zusammenfassungen zielen darauf ab, zum einen mögliche Verständnisschwierigkeiten zu beheben, zum anderen die aufgerufenen Wissenselemente zu verfestigen. Sie werden z.T. durch paraoperative Prozeduren (*also*) eingeleitet und von wiederholten Rückversicherungssignale (*ja? ok?*) und Signalen der Sprecher-Hörersteuerung (*hm? äh?*) begleitet.

Nach den beiden Aufforderungen (a+b) zu einem *sich-in-Erinnerung-rufen* bereits erfolgter Beobachtungen bei einem Vergleich von Übersetzungsvarianten leitet die Dozentin mit Frage (c) über zur stilistischen Ebene einer Übersetzung, die für die Evaluierung seitens der Muttersprachler eine wesentliche Rolle spielt.

Frage (d) antizipiert eine denkbare Frage seitens der Studierenden nach einer Begründung der höheren Akzeptanz der Fassung, die sich am weitesten vom der Vorlage entfernt (eine Frage, die allerdings, nicht gestellt wird). Sie wird wiederum von D selbst beantwortet unter Einbeziehung einer weiteren Kategorie (Satzrhythmus).

Die nachfolgende Frage (e) hebt auch durch die vorangehende Reformulierung (*dennoch haben wir im Fall des Modalverbes ein Problem*) sowie durch Reformulierung von Frage (c) heraus, dass es noch andere Kriterien bei der Evaluierung einer Übersetzung gibt[11]. Die Dozentin bereitet auf diese Weise eine Überlegung vor, die durch Frage (f) präzisiert wird, wenn sie die Frage aufwirft, wodurch übersetzerische Eingriffe (hier: die Hinzufügung eines Elements, zu dem im Ausgangstext kein entsprechendes Element vorliegt) ihre Berechtigung erhalten.

Mit Hilfe dieser verstehensorganisierenden Frageformulierungen werden die Kriterien einer empirisch fundierten Übersetzungskri-

---

11 Dass die Dozentin in diesem Zusammenhang den Symbolfeldausdruck Problem verwendet, geschieht m.E. weniger um ein tatsächliches Problem zu benennen, als vielmehr um eine Herausforderung der zielsprachengerechten Translation hervorzuheben, und damit auch eine Frage, die sich dem Übersetzer stellt, wenn er Eingriffe vornimmt und sich dadurch möglicherweise von seinem Ausgangstext entfernt.

tik zunehmend enger fokussiert. An ihnen lassen sich gleichzeitig die einzelnen Handlungsschritte ablesen, mit denen die Dozentin auf ihr Ziel hinarbeitet, deutlich zu machen, dass bei der Beurteilung einer Übersetzung in Hinblick auf ihre Adäquatheit auch Aspekte relevant werden, die mit der kognitiven Verarbeitung von Sprache zusammenhängen. In den Kategorien der funktionalen Pragmatik gefasst, geht es dabei um operative Prozeduren.

Das Beispiel verdeutlicht, dass Dozentin und Studierende über das notwendige Handlungswissen verfügen, dass von den Studierenden keine Antwort auf diese Fragen erwartet wird, sondern von der Dozentin selbst unmittelbar im Anschluss an die Frage (nach)geliefert wird. Dies zeigt sich nicht zuletzt auch daran, dass Antwortversuche seitens der Studierenden unterbleiben und seitens der Dozentin auch nicht eingefordert werden.

7. Forschungsfragen

Neben den verstehensorganisierenden Fragen ist im *euroWiss*-Korpus ein anderer Frage-Typus nachweisbar. Vgl. etwa die folgenden Beispiele:

(11) Und die Frage ist: sind das wirklich Teile der Seele? Nun es ist so, dass… (Phil_V_1-4, Fl. 86)

(12) Wenn Sie Wissenschaftler sind (…) dann fragen Sie: Wozu hat das Lamm diese Teile…. (Phil_V_1-4, Fl. 436ff)

(13) Und solche Dinge wärn natürlich auch interessante Forschungsfragen: Ist es wirklich so? … Oder gibt es nicht auch …? Könnten wir das widerlegen? (RomLit_V_23, Fl. 392ff)

(14) […] in der wir uns sozusagen gefragt haben: Welche Funktion übernimmt das Weibliche in diesen Bildern oder in diesen Texten? (RomLit_V_23, Fl. 957ff)

Die Spezifik solcher Frageformulierungen, die wir im Folgenden Forschungsfragen bezeichnen, scheint darin zu legen, die Frage als

Ausgangspunkt von Forschung zu verdeutlichen, d.h. zu zeigen, wie Forscher vorgegangen sind bzw. welche offenen Fragen es noch gibt. Anders als die oben beschriebenen verstehensorganisierenden Fragen werden diese Frageformulierungen teilweise nur als solche benannt und nicht beantwortet, oft weil es sich eben um echte, derzeit noch unbeantwortete Forschungsfragen handelt.

Forschungsfragen dieser Art tauchen in den uns vorliegenden italienischen Daten kaum auf. Hier sind zwei der wenigen Beispiele wiedergegeben, die im Korpus nachweisbar sind:

(15) Ci si chiede qual è la differenza tra una semplice descrizione narrativa di un fenomeno e un'analisi propriamente scientifica (Soz_V_1; Fl. 1032ff)

(16) Si cercherà di rispondere alla domanda: che cos'è la traduzione (GerSpr_V_4_LingTedLM_HA2)

Diese Daten lassen sich möglicherweise dadurch erklären, dass Forschungsfragen und das In-Frage-Stellen von tradiertem Wissen – soweit es sich aus den Videographien des *euroWiss*-Korpus ergibt – in der Lehre an italienischen Universitäten eher in fortgeschrittenen Phasen des Studiums thematisiert werden.

Auf die Tendenz, dass Forschungsfragen im italienischen Wissenschaftsdiskurs weniger häufig explizit bzw. auf metasprachlicher Ebene verbalisiert werden, deuten auch komparative Untersuchungen von Schrifttexten. So ergab eine Bestandsaufnahme (s. Heller 2012: 122) von verbalen Ausdrücken anhand eines von Fandrych (2002) entwickelten Rasters sprechhandlungsbezeichnender Verben in einem deutsch-italienischen Paralleltextkorpus wissenschaftlicher Rezensionen, dass die semantisch-funktionalen Gruppe *Fragen stellen* im deutschen Teilkorpus sehr viel stärker vertreten war als im italienischen Teilkorpus – sowohl auf der *type*-, als auch auf der *token*-Ebene.

Hier zeichnen sich Unterschiede zum Lehr-Lerndiskurs an deutschen Universitäten ab, die ihre Bestätigung finden, wenn man die Schwierigkeiten betrachtet, die italienischsprachige DaF-Lernende bei der Ausformulierung einer klaren Fragestellung am Anfang ihrer ersten (vor)wissenschaftlichen Arbeiten haben. Der korrekte Einsatz von Kollokationen wie „auf eine Frage eingehen" oder „einer Frage nachgehen" bereitet ihnen häufig große Schwierigkeiten (vgl. Ronco-

roni 2011; Heller 2006). Anhand weiterer Untersuchungen ist zu überprüfen, inwiefern das „Fragen“ bzw. das „Sich Fragen“ eine für die deutschsprachige Wissenschaftskultur spezifische Handlungsqualität bei der Wissensvermittlung besitzt.

8. Fazit

Frageformulierungen in Sprechhandlungsverkettungen des universitären Lehr-Lern-Diskurses werden zwar prosodisch und syntaktisch als Fragen artikuliert, haben aber illokutiv einen anderen Status. Sie suspendieren kurzzeitig den assertiven Redefluss des Dozenten, um Sprechhandlungen wie z.B. Begründungen, Erläuterungen und Zusammenfassungen einzuleiten bzw. die Studierenden aufzufordern, einen bestimmten Gedankengang nachzuvollziehen.

Durch Frageformulierungen sollen Suchprozesse bei H ausgelöst werden, die durch das mentale, automatische Abrufen des Frage-Antwort-Musters aktiviert werden. Da die Frage aber meist vom Lehrenden beantwortet wird, der sie auch gestellt hat, sollen Frageformulierungen mögliche, bei H vorliegende Wissensdefizite bearbeiten. Als Einleitungen zu anderen Sprechhandlungen bereiten Frageformulierungen auch auf das Kommende vor und dienen der Aufrechterhaltung der Aufmerksamkeit.

Die genannten Funktionen charakterisieren die Verwendung von Fragen im Lehrdiskurs, deren Hauptfunktion darin besteht, einen dozentenseitigen Beitrag zum Verständigungshandeln zu leisten und das Verstehen seitens der Studierenden zu steuern und zu befördern. Zur Verdeutlichung wurde hier ein Transkriptausschnitt aus einer italienischen Vorlesung kommentiert. Er veranschaulicht den gezielten Einsatz – seitens der Dozentin – von Fragen, um Wissenselemente aus vorausgegangen Sitzungen in Erinnerung zu rufen und sie mit dem aktuellen Lehrdiskurs zu vernetzen. Desweiteren dienen solche von der Dozentin selbstbeantworteten Fragen der progressiven Fokus-

sierung der zu vermittelnden Wissenselemente sowie der Überleitung zu Begründungen und Erläuterungen.

Während verstehensorganisierende Fragen sowohl in den deutschen und als auch in den italienischen Vorlesungen des *euroWiss*-Videographie-Korpus weit verbreitet sind, lässt sich in Bezug auf das Vorkommen von sogenannten Forschungsfragen ein Unterschied zwischen den dort zusammengeführten deutschen und italienischen Daten feststellen. Deutschsprachige Lehrende greifen häufig auf solche Frageformulierungen zurück und stellen somit das zu vermittelnde Wissen als ‚der Befragung wert' vor. Demgegenüber ist dieses Verfahren in den Lehrveranstaltungen der italienischen Dozenten eher selten zu beobachten. Solche Divergenzen werfen weitere Überlegungen auf, etwa, ob diesem Verfahren eine Spezifik der deutschen Wissenschaftskultur zu Grunde liegt. Ihre Untersuchung bedarf weiterer Studien, bei der auch andere weniger evidente Ausdrucksformen von Forschungsfragen einzubeziehen wären.

Bibliographie

Bamford, Julia 2000. Question and Answer Sequencing in Academic Lectures. In Coulthard, Malcolm / Cotterill, Janet / Rock, Frances (Hgg.) *Dialogue analysis VII: working with dialogue*. Tübingen: Niemeyer, 159-169.

Carobbio, Gabriella / Zech, Claudia 2013. Vorlesungen im Kontrast – zur Vermittlung Fremder Literatur in Italien und Deutschland. In Hans-Bianchi, Barbara / Miglio, Camilla / Pirazzini, Daniela / Vogt, Irene / Zenobi, Luca (Hgg.) *Fremdes wahrnehmen, aufnehmen, annehmen – Studien zur deutschen Sprache und Kultur in Kontaktsituationen*. Frankfurt/M.: Peter Lang, 289-307.

Deckwerth, Michael 1983. Zur Entscheidungsfrage im Unterricht. In Ehlich, Konrad / Rehbein, Jochen (Hgg.) *Kommunikation in Schule und Hochschule*. Tübingen: Narr, 29-38.

Ehlich, Konrad / Rehbein, Jochen 1979. Sprachliche Handlungsmuster. In Soeffner, Hans-Georg (Hg.) *Interpretative Verfahren in den Sozial- und Textwissenschaften*. Stuttgart: Metzler, 243-274.

Ehlich, Konrad / Rehbein, Jochen 1986. *Muster und Institution: Untersuchungen zur schulischen Kommunikation.* Tübingen: Narr.

Ehlich, Konrad 2007. *Sprache und sprachliches Handeln.* Drei Bände. Berlin: de Gruyter.

Eichinger, Ludwig M. 2007. Linguisten brauchen Korpora und Korpora Linguisten – Wege zu wohl dokumentierten und verlässlichen Aussagen über Sprache. In Kallmeyer, Werner / Zifonun, Gisela (Hgg.) *Sprachkorpora: Datenmengen und Erkenntnisfortschritt*, IDS Jahrbuch 2006. Berlin: de Gruyter, 1-8.

Fandrych, Christian 2002. Herausarbeiten vs. Illustrate: Kontraste bei der Versprachlichung von Sprechhandlungen in der englischen und deutschen Wissenschaftssprache. In Ehlich, Konrad (Hg.) *Mehrsprachige Wissenschaft − europäische Perspektiven. Eine Konferenz zum Europäischen Jahr der Sprachen 2001*. München: Institut für Deutsch als Fremdsprache / Transnationale Germanistik 2002.

Fierbinţeanu, Ioana Hermine 2013. *Fragen in mündlichen Prüfungsgesprächen an der Hochschule.* Vortrag gehalten an der XV. IDT-Tagung „Deutsch von innen – Deutsch von außen" (Sektion G6 „Wissenschaftssprache Deutsch"), Bozen 29.7.-3.8.2013.

Heller, Dorothee 2006. *L'autore traccia un quadro...* Beobachtungen zur Versprachlichung wissenschaftlichen Handelns im Deutschen und Italienischen. In Ehlich, Konrad / Heller, Dorothee (Hgg.) *Die Wissenschaft und ihre Sprachen.* Bern: Peter Lang, 63-85.

Heller, Dorothee 2012. *Wissenschaftskommunikation im Vergleich: Fallstudien zum Sprachenpaar Deutsch-Italienisch*. Frankfurt/M.: Peter Lang.

Heller, Dorothee 2014. Dozentenseitige Beiträge zum Verständigungshandeln in einer italienischen Germanistikvorlesung. In Hornung, Antonie / Carobbio, Gabriella / Sorrentino, Daniela (Hgg.) *Diskursive und textuelle Strukturen in der*

Hochschuldidaktik. Deutsch und Italienisch im Vergleich. Münster: Waxmann, 95-116.

Hornung, Antonie / Carobbio, Gabriella / Sorrentino, Daniela (Hgg.) 2014. *Diskursive und textuelle Strukturen in der Hochschuldidaktik – Deutsch und Italienisch im Vergleich.* Münster: Waxmann.

Nowak, Johann / Macht, Konrad 1996. *Die Kunst des Fragens. Theorie und Praxis der Frage als didaktisches Steuerungsinstrument.* Augsburg: Universität Augsburg.

Petersen, Jörg / Sommer, Hartmut 1999. *Die Lehrerfrage im Unterricht. Ein praxisorientiertes Studien- und Arbeitsbuch mit Lernsoftware*. Donauwörth: Auer Verlag.

Redder, Angelika 1990. *Grammatiktheorie und sprachliches Handeln: 'denn' und 'da'*. Tübingen: Niemeyer.

Redder, Angelika 1998. Sprachwissen als handlungspraktisches Bewußtsein – eine funktional-pragmatische Diskussion. *Didaktik Deutsch* 5/1998, 60-76.

Redder, Angelika 2008. Functional Pragmatics. In Antos, Gerd / Ventola, Eija (Hgg.) *Interpersonal Communication.* Berlin: de Gruyter, 133-178.

Redder, Angelika 2009. Sprachliche Wissensbearbeitung in der Hochschulkommunikation. In Lévy-Tödter, Magdalène / Meer, Dorothee (Hgg.) *Hochschulkommunikation in der Diskussion*. Frankfurt/M.: Peter Lang, 17-44.

Redder, Angelika / Thielmann, Winfried / Heller, Dorothee (Hgg.) (i.E.). *Kritik und Eristik in der Hochschullehre – komparative Analysen in Deutschland und Italien.* Heidelberg: Synchron.

Roncoroni, Tiziana 2011. Hausarbeiten – Schreibanforderung an nicht muttersprachliche Studierende. Beobachtungen zur Textstrukturierung und Leserorientierung. In Knorr, Dagmar / Nardi, Antonella (Hgg.) *Fremdsprachliche Textkompetenz entwickeln.* Frankfurt/M.: Peter Lang.

Wiesmann, Bettina 1999. *Mündliche Kommunikation im Studium. Diskursanalysen von Lehrveranstaltungen und Konzeptualisierung der Sprachqualifizierung ausländischer Studienbewerber.* München: iudicium.

Wiesmann, Bettina 2001. „Und was können wir jetzt draus machen? – Studieren in einer diskursiv entwickelnden Lehr-Lern-Kultur. *Jahrbuch Deutsch als Fremdsprache* 27/2001, 251-274.

ANDREA ABEL / AIVARS GLAZNIEKS

4. „Ich weiß zwar nicht, was mich noch erwartet, doch ...“ – Der Einsatz von Korpora zur Analyse textspezifischer Konstruktionen des konzessiven Argumentierens bei Schreibnovizen

1. Einleitung

Hintergrund des Beitrags bildet die sowohl fachintern als auch öffentlich stark diskutierte Sprachkompetenz für die Ausbildung und den Beruf, insbesondere der Bereich der Schreibkompetenzen an der Schnittstelle Oberschule-Universität, und in diesem Zusammenhang eine aktuelle Studie, die sich mit dem Schreiben an eben dieser Schnittstelle beschäftigt. Die Studie nimmt daher die Kompetenzen in den Blick, die es SchülerInnen im Hinblick auf die Anforderungen in der Schule und vor allem im daran anschließenden Studium zu vermitteln gilt. Dazu zählt unter anderem die Argumentationskompetenz als Teil der Schreibkompetenz und dabei die Fähigkeit zur konzessiven Argumentation als ein herausragendes Kennzeichen argumentativer, insbesondere wissenschaftlicher Texte. Konzessivität gehört im Bereich der „Alltäglichen Wissenschaftssprache“ (Ehlich 1999: 9; Ehlich/Graefen 2001: 373) zu einem Spektrum von Sprachmitteln, mit denen sich Lernende vertraut machen müssen.

Die folgenden Beispiele aus den Schülertexten der Studie mögen illustrieren, um welche Art von Phänomenen es im Beitrag gehen soll:

(1) Er mag zwar Recht haben mit seiner Beschreibung von jungen Menschen, aber labil und unsicher zu sein gehört meiner Meinung nach zum Erwachsenwerden dazu. (ID1090)

(2) Natürlich ist es nicht schlecht viele Sachen im Schrank zu haben, trotzdem ist die Sucht nicht bei jedem vorhanden! (ID1005)

(3) Sicher machen Jugendliche in dieser Phase Fehler, laufen allen Trends nach usw doch gehört das nicht zur Jugend dazu? (ID1061)[1]

In den aufgeführten Textausschnitten verwenden die SchreiberInnen zweiteilige Konstruktionen wie *zwar …, aber …* (siehe Beispiel (1)), die auch in Texten von ExpertInnen und StudentInnen „sehr häufig vorzufinden [sind] und […] als prototypische Mittel des konzessiven Argumentierens in der Wissenschaftsdomäne aufgefasst werden [können]." (Steinhoff 2007: 332). Daneben spielen auch Konstruktionen mit Modalwörtern wie *natürlich* oder *sicher* (siehe Beispiele (2) und (3)) eine wichtige Rolle (vgl. Steinhoff 2007: 359). Im Gebrauch solcher Konstruktionen zeigen sich typische Entwicklungsphänomene (vgl. Steinhoff 2007: 360); sie sind wichtig beim Auf- und Ausbau von Schreibkompetenzen.

Die Begriffe *Kompetenz* und *Schreibkompetenz* wurden und werden in vielfältigen Kontexten mit teilweise unterschiedlicher Bedeutung verwendet. Im vorliegenden Beitrag dient ein Kompetenzbegriff als Grundlage, der in der aktuellen Bildungsdiskussion (z. B. PISA etc.) weit verbreitet ist und wonach Kompetenz „eine Disposition [ist], die Personen befähigt, bestimmte Arten von Problemen erfolgreich zu lösen, also konkrete Anforderungssituationen eines bestimmten Typs zu bewältigen" (Klieme *et al.* 2003: 72). Daraus wird, wie in der Schreibkompetenzforschung durchaus üblich, eine pragmatisch konzipierte Schreibkompetenz, in der Sprache in einem sozialen Handlungskontext eine kommunikative Funktion übernimmt, als Teil einer – in der Wissenschaft bislang ungenügend definierten – allgemeinen, umfassenden Kompetenz abgeleitet (vgl. Becker-Mrotzek/ Böttcher 2006: 56).

---

1 Die Schreibweisen der Beispiele entsprechen dem Original. Selbstkorrekturen, Tilgungen o.Ä. werden in den transkribierten Beispielen aus Gründen der Lesbarkeit allerdings nicht wiedergegeben.

> Unter Schreibkompetenz verstehen wir im Folgenden die Fähigkeit zur Produktion von Texten. Dabei handelt es sich im Kern um eine spezifisch sprachliche Fähigkeit von mittlerer Komplexität, die sich aus unterschiedlichen Teilfähigkeiten und Kenntnissen (Dimensionen) zusammensetzt. Konstitutiv ist die Fähigkeit, sprachliche Äußerungen so zu konzipieren, dass sie aus sich heraus verständlich sind, also über Raum und Zeit zerdehnt werden können. Um das zu leisten, müssen sie sich ihren eigenen Kontext, ihre eigene Situation schaffen. Dazu bedarf es des Einsatzes spezifischer lexikalischer, grammatikalischer sowie pragmatischer Mittel. Wir haben es also mit einem besonderen lexikalischen, grammatischen und pragmatischen Wissen zu tun. (Becker-Mrotzek/Böttcher 2006: 57)

Beim schriftlichen Argumentieren, vor allem konzessiver Art, ist es im Besonderen nötig, Fähigkeiten zur Perspektivenübernahme und -integration an den Tag zu legen. Diese Fähigkeiten gelten als besonders anspruchsvoll (vgl. Rezat 2011: 50; Feilke 2010: 218; Leitao 2003: 272).

Konzessive argumentative Strukturen zeigen in elaborierter Form eine hohe Stufe der Schreibentwicklung an (vgl. Rezat 2011: 50; Leitao 2003: 272) und sind auch Indikatoren einer wissenschaftlichen Textkompetenz (vgl. Steinhoff 2007: 124). Es ist nicht Aufgabe der Schule, eine solche zu vermitteln, da wissenschaftliches Schreiben „eine Tätigkeit *sui generis*“ (Dittmann *et al.* 2003: 163) darstellt und Fähigkeiten sowie Kenntnisse inkludiert, die im schulischen Kontext nicht grundlegend sind (vgl. Steinhoff 2007: 2; Tente/Furchner/Ruhmann 1999: 62). Sehr wohl kann jedoch ein erstes Heranführen an wissenschaftliche Texte und Konventionen bereits in der Oberstufe etwa durch wissenschaftspropädeutische Arbeiten stattfinden.

„In pragmatischer Perspektive ist die wissenschaftliche Textkompetenz eine *Sprachgebrauchskompetenz*, eine Kompetenz zum Gebrauch einer fachübergreifenden, aber auch fachgebundenen Sprache“ (Steinhoff 2007: 2). Für die fachübergreifende Nutzung von alltagssprachlichen Ausdrucksmitteln für wissenschaftliche Zwecke hat Ehlich (1993: 33) den Begriff der „Alltäglichen Wissenschaftssprache“ geprägt und meint „die fundamentalen sprachlichen Mittel [...] , derer sich die meisten Wissenschaften gleich oder ähnlich bedienen“. Nicht zuletzt aus dieser Perspektive können Teilbereiche des wissenschaftlichen Schreibens, zu denen das konzessive Argumentie-

ren gehört, auch für den schulischen Kontext als relevant betrachtet werden, in dem gewöhnlich stärker mit dem Begriff der Bildungssprache als dem der „Alltäglichen Wissenschaftssprache“ operiert wird.

Bildungssprache kann sehr allgemein als „die Sprache [bezeichnet werden], in der Bildung in Einrichtungen und Institutionen gestaltet wird“ (Gogolin/Lange 2010: 9). Außerdem „bezeichnet der Begriff [der Bildungssprache] eine spezielle Ausprägung des Deutschen: Je weiter eine Bildungsbiografie fortschreitet, desto mehr unterscheiden sich die sprachlichen Anforderungen in den Bildungseinrichtungen vom Repertoire der alltäglichen mündlichen Sprache“ (Gogolin/Lange 2010: 9). Die Bildungssprache kann auch etwas spezifischer als die Sprache beschrieben werden,

> die überwiegend in den Massenmedien, in Fernsehen, Rundfunk, Tages- und Wochenzeitungen benutzt wird. Sie unterscheidet sich von der Umgangssprache durch die Disziplin des schriftlichen Ausdrucks und durch einen differenzierteren, Fachliches einbeziehenden Wortschatz; andererseits unterscheidet sie sich von Fachsprachen dadurch, dass sie grundsätzlich für alle offensteht, die sich mit den Mitteln der allgemeinen Schulbildung ein Orientierungswissen verschaffen können. […] Sie wird durch die Funktion, Fachwissen in die einheitsstiftenden Alltagsdeutungen einzubringen, definiert. Die Bildungssprache ist ein Medium, durch das Bestandteile der Wissenschaftssprache von der Umgangssprache assimiliert werden. (Habermas 1981: 345-6)

Ein so eingegrenzter Begriff von Bildungssprache scheint dem der „Alltäglichen Wissenschaftssprache“ durchaus ähnlich (für eine wieterführende Diskussion s. Redder: im Druck). Konstruktionen und Routinen der Wissenschaftssprache, wie z. B. *einer Frage nachgehen*, *im Folgenden*, *kann man feststellen* etc. (vgl. Steinhoff 2007: 3) sind als solche keineswegs auf eine Gruppe von akademisch Gebildeten beschränkt, sondern durchdringen große Bereiche des gesellschaftlichen Lebens, so auch den Bereich des schulischen Schreibens (vgl. Ehlich/Valtin/Lütke 2012: 25-27).

Zur Untersuchung sprachlicher Phänomene wie der eben vorgestellten können unterschiedliche Methoden bzw. Methodenkombinationen angewandt werden: Zum einen eignen sich qualitative, manuelle Analysen u. a. mittels Analyseraster (z. B. Nussbaumer/ Sieber 1994), zum anderen kommen quantitative, (semi-) automati-

sche Analysen korpuslinguistisch aufbereiteter Texte mithilfe computerlinguistischer Werkzeuge zum Einsatz (z. B. Schründer-Lenzen/ Henn 2009). Dabei können entweder eine hypothesentestende Vorgangsweise und die Untersuchung textsortenspezifischer Phänomene aufgrund von Vorgaben aus der Literatur oder eine explorative Vorgangsweise und die Untersuchung bestimmter oberflächensprachlicher Realisierungen (z. B. Verben mit Präpositionalphrasen) im Vordergrund stehen, die anschließend wiederum qualitativ ausgewertet werden können (vgl. Abel/Glaznieks im Druck).

In diesem Beitrag soll aufgezeigt werden, ob spezifische, in der einschlägigen Literatur beschriebene, konzessive argumentative Strukturen (u. a. Steinhoff 2007) in Schülertexten überhaupt nachweisbar sind und wie sie in einem Schüleraufsatzkorpus mit computerlinguistischen Methoden systematisch untersucht werden können.

Nachdem die allgemeinen Ziele des Beitrags und durch die einführenden Erläuterungen die Bereiche, in denen er sich bewegt, abgesteckt sind, folgen in Abschnitt 2 Ausführungen zum konzessiven schriftlichen Argumentieren. In Abschnitt 3 werden die Pilotstudie, auf die der Beitrag Bezug nimmt, im Detail vorgestellt und dabei deren spezifische Ziele, aber auch das Gesamtprojekt sowie die Datengrundlage der Analysen, die verwendeten Methoden und damit erzielte Ergebnisse beschrieben. Schließlich zeigt Abschnitt 4 auch Bezüge zu anderen Studien auf und gibt sowohl einen Ausblick auf aus den Analysen ableitbare weitere Arbeitsschritte als auch Anregungen für die Didaktik.

2. Konzessive Argumentation

Konzessives Argumentieren, bei dem gegnerische Argumente eingeräumt werden, gilt, wie eingangs bereits angedeutet, als genuines Mittel der Perspektivenübernahme und ist Ausdruck konzeptionell schriftlicher Argumentationskompetenz (vgl. Schmidlin/Feilke 2005: 11; Rezat 2011: 50). Es ist nicht nur ein Kennzeichen argumentativer

Texte im Allgemeinen (z. B. Erörterung) und wissenschaftlicher Texte im Besonderen (Fokus: „Alltägliche Wissenschaftssprache“, vgl. Ehlich 1993, 1999; Steinhoff 2007: 124), sondern auch des elaborierten Schreibens sowie guter Argumentation insgesamt (vgl. Leitao 2003: 272).

Beim schriftlichen Argumentieren lässt sich der/die SchreiberIn auf eine virtuelle Diskussion mit einem vorgestellten Gegenüber ein. Die Argumentation erfordert „eine aufwändige ‚Simulierung einer Dialogstruktur‘ (Paek 1993: 17). Der möglichen Widerrede der Rezipienten muss vorgegriffen werden […]“ (Steinhoff 2007: 124). Es wird die Fähigkeit vorausgesetzt, mögliche Einwände bzw. Gegenargumente von LeserInnen zu antizipieren und vorwegzunehmen. Damit ist eine Überleitung auf den spezifischen Bereich des konzessiven Argumentierens möglich:

> Namensgeber ist hier die „concessio“ aus der rhetorischen Figurenlehre. Mit einer „concessio“ wird ein gegnerisches Argument aufgegriffen. Zugleich wird es aber entschärft, wodurch die eigene Argumentation letztlich gestärkt wird: „Wissenschaftliche Argumentation wird wirkungsvoller, wenn man auch die Gegenposition fair darstellt und mit Gegenargumenten widerlegt.“ (Kruse 2002: 173). (Steinhoff 2007: 124)

Beim konzessiven Argumentieren geht es also darum, mögliche Einwände der Leserschaft anzusprechen und zu entkräften. Indem Gegenargumente berücksichtigt und widerlegt werden, kann die eigene Position an Plausibilität und Überzeugung gewinnen und dadurch gestärkt werden (vgl. Steinhoff 2007: 329). Damit dies gelingt, ist eine starke Orientierung am Adressaten nötig, was eine große kognitive Leistung darstellt. Beim Schreiben bedeutet das konzessive Argumentieren eine besonders große Herausforderung:

> Im Gegensatz zum mündlichen Argumentieren wird die Perspektivenübernahme und damit das konzessive Argumentieren beim schriftlichen Argumentieren durch die Lösung vom Situationskontext (Dekontextualisierung) erschwert. Das Fingieren eines virtuellen Dialogs beim schriftlichen Argumentieren erfordert vom Schreiber Kontextualisierungskompetenz, die Antizipation und Entkräftung von Gegenargumenten sowie die Beherrschung der Mittel kontroversen und hypothetischen Argumentierens (vgl. Feilke 2010: 218; Leitao 2003: 276). (Rezat 2011: 50)

Voraussetzung für ein erfolgreiches konzessives Argumentieren ist eine flexible, textsortenangemessene Leseranpassung als Teil sozialkommunikativer Schreibfähigkeiten (vgl. Jechle 1992: 55, 67; Steinhoff 2007: 330). Untersuchungen zur Schreibentwicklung haben gezeigt (z. B. Augst/Faigel 1986; Schneuwly 1988; Jechle 1992; Coirier/Golder 1993), dass sich elaborierte Formen des Argumentierens oder auch eine bewusste, differenzierte Hinwendung zum Adressaten erst im Übergang ins Erwachsenenalter entfalten (siehe dazu auch Steinhoff 2007: 330; Rezat 2011: 50-51). (Prä-) argumentative, einschließlich (prä-)konzessiver argumentativer, Strukturen sind allerdings bereits im Grundschulalter feststellbar, etwa in Form einer den Leser involvierenden Schreibweise mit einer Integration von antizipierten Argumenten und Gegenargumenten (vgl. z. B. Augst/ Faigel 1986: 135ff.; Feilke 2003: 188; Augst *et al.* 2007: 204). In nachfolgenden Phasen der schulischen und akademischen Sozialisation werden solche Verfahren des leserorientierten Argumentierens weiter ausgebaut und verfestigt (vgl. z. B. Steinhoff 2007: 2, 330; Augst *et al.* 2007: 204). Den Fokus in diesem Beitrag bildet die Gruppe der (spät-)adoleszenten SchülerInnen.

Zum konzessiven Argumentieren sind verschiedene Verfahren möglich, die in der wissenschaftlichen Literatur ausführlich beschrieben werden (siehe für einen Überblick z. B. Rezat 2011, 2009; Steinhoff 2007). Grundsätzlich kann zwischen expliziten und impliziten Verfahren unterschieden werden, d. h. dass konzessive Strukturen an der sprachlichen Oberfläche mit verschiedenen Mitteln markiert werden können oder eben nicht (vgl. Steinhoff 2007: 331ff.; Rezat 2011: 53ff.). Häufig wird „zwischen den Zeilen" konzessiv argumentiert. Solche impliziten Mittel sollen aus den folgenden Ausführungen ausgeklammert bleiben, da der Schwerpunkt auf korpuslinguistisch erfassbaren Oberflächenphänomenen liegt.

Explizite konzessive Konstruktionen können der Größe *Ausdruck* zugeordnet werden, die in Anlehnung an Feilke (1996: 65ff.) als

> eine sprachliche Einheit [beschreibbar ist], die ein Wort, aber auch mehrere Wörter umfassen kann, z. B. Kollokationen oder grammatische Konstruktionen. Die Zeichenqualität des Ausdrucks besteht darin, dass er ein bestimmtes Gebrauchsschema indiziert. Die Kollokation ‚der Frage nachgehen indi-

> ziert ein typisch wissenschaftliches Gebrauchsschema: Sie ist in der Wissenschaftsdomäne gebräuchlich und wird meist im Rahmen von textkommentierenden Abschnitten eingesetzt, zum Zwecke der Zielexplizierung und thematischen Organisation. (Steinhoff 2007: 79).

Dieser Ansatz weist einen klaren Bezug zu Feilkes Theorien der „Common-Sense-Kompetenz“ (1994) und der „idiomatischen Prägung“ (1996) auf. „Der Common sense ist ein auf die ökologischen Bedingungen des menschlichen Handeln bezogenes und durch diese Bedingungen pragmatisch konstituiertes und stabilisiertes intuitives Wissen.“ (Feilke 1994: 66). Die entsprechende Kompetenz kann folglich als Fähigkeit beschrieben werden, typische, pragmatisch bewährte sprachliche Gebrauchsmuster zu verwenden (vgl. Feilke 1993; Steinhoff 2007: 88-91). Dazu zählt auch das Wissen über idiomatische Ausdrücke (auf syntaktischer, z. B. *von Haus zu Haus*, semantischer, z. B. *die letzte Ruhe*, und pragmatischer Ebene, z. B. *gut, dass ich Sie treffe*, vgl. Feilke 1996: 211ff.), die „offenbar als ‚Standardlösungen' für rekurrente Formulierungsprobleme“ (Steinhoff 2007: 104) verwendet werden. Solche Ausdrücke bzw. „idiomatische Prägungen“, auf die in der Kommunikation ständig zurückgegriffen wird, sind „selektive und sozial verbindliche Ausdrücke, deren (volles) Gebrauchspotential kompositionell nicht prädiktabel ist.“ (Steinhoff 2007: 108; vgl. Feilke 1998: 74; sowie Feilke 1993: 17f.). Bestimmte Ausdrücke, einschließlich solcher der „Allgemeinen Wissenschaftssprache“, sind Indikatoren für bestimmte Kommunikationsbereiche, z. B. für Wissenschaftstexte (vgl. Steinhoff 2007: 107).

Typische Ausdrücke des konzessiven Argumentierens in wissenschaftlichen Texten zeigen die folgenden Beispiele aus Expertentexten (aus Steinhoff 2007: 331):

(4) **[T]** In Duisburg wie auch in vielen anderen Städten beruhte die Urteilsfindung damals noch auf der Befragung rechtskundiger Männer, die als Schöffen großes Ansehen hatten. **[K] Zwar** gab es im 14. Jh. schon Bemühungen um ein schriftgeleitetes Verfahren der Urteilsfindung […], **[P] doch** wurden sie in der Praxis unterlaufen, indem man sich darauf berief, daß der konkrete Fall nicht im Buch geregelt sei […]

(5) **[K] Sicherlich** ist es nicht zwingend, bei der Lektüre eines Textes immer den Autor mitzubedenken. [FN] Strukturalistische oder andere Formen werkim-

> manenter Textanalyse belegen dies zur Genüge. **[T] Allerdings** ist es eine gängige Praxis, so die Ausgangsthese der folgenden Überlegungen, literarische Texte (nicht nur aber auch) im Hinblick auf ihren Autor zu lesen.

Sie weisen übliche Elemente des konzessiven Argumentierens auf: „eine *These* des Verfassers (T), eine *Konzession* bzw. ein *Kontraargument* (K) und ein *Proargument* (P), das das Kontraargument entkräftet und die These des Verfassers stützt." (Steinhoff 2007: 331). Die Versprachlichung konzessiven Argumentierens durch explizite Indikatoren, u. z. durch zweiteilige *zwar*-Konstruktionen sowie durch Konstruktionen mit Modalwörtern – wie in den Beispielen (4) und (5) –, die als typisch für die Wissenschaftsdomäne gelten, sind, wie Steinhoff (2007: 332) gezeigt hat, in Experten- wie auch Studententexten häufig vorzufinden. Auch Rezat (2011: 53, weitere empirische Untersuchungen 2007, 2009) hat vergleichbare syntaktische Kombinationsmuster als prototypisch für konzessives Argumentieren beschrieben und in Experten-, aber auch in Schülertexten (Elementar- und Sekundarbereich) untersucht.

3. Die Untersuchung

Die in diesem Beitrag vorgestellte empirische Pilotstudie verfolgt dreierlei Ziele: Es werden Texte von OberschülerInnen ein Jahr vor der Matura bzw. dem Abitur auf das Vorkommen konzessiver argumentativer Konstruktionen hin untersucht. Beschränkt wird sich dabei auf Konstruktionen der „Alltäglichen Wissenschaftssprache" mit der entsprechenden syntagmatischen Typik, die für argumentatives Schreiben charakteristisch ist. Weiterhin werden mögliche Zusammenhänge mit außersprachlichen Variablen, im konkreten Fall *Region*, *Schultyp* und *Geschlecht*, überprüft. Zuletzt geht es darum, Möglichkeiten und Grenzen beim Einsatz bewährter computerlinguistischer Methoden zur Unterstützung der Untersuchung zu testen bzw. ggf. nötige Überarbeitungsschritte in der Vor- und Nachbearbeitung zu (er)fassen.

3.1. Datengrundlage und Methode

Die Datengrundlage bildet das deutschsprachige Lernerkorpus „KoKo“, das im Rahmen des Projekts „Bildungssprache im Vergleich: korpusunterstützte Analyse der Sprachkompetenz bei Lernenden im deutschen Sprachraum (KoKo)“ entstanden ist. Das Projekt wird seit 2010 am Institut für Fachkommunikation und Mehrsprachigkeit der EURAC in Zusammenarbeit mit Partnern der Fakultät für Bildungswissenschaften der Freien Universität Bozen und des Instituts für Germanistik der Universität Innsbruck durchgeführt (vgl. Abel/Glaznieks im Druck). Es gliedert sich in die Initiative „Korpus Südtirol“[2] ein, innerhalb welcher Südtiroler Texte systematisch gesammelt, archiviert und korpuslinguistisch erschlossen werden, um sie der Allgemeinheit zugänglich zu machen und den Sprachgebrauch des Deutschen in Südtirol zu dokumentieren (siehe z. B. Anstein/Oberhammer/Petrakis 2011; Abel/Anstein 2011).

Das Lernerkorpus „KoKo“ enthält 1503 Texte und weist eine Gesamtgröße von 811.330 Tokens auf (ohne Satzzeichen, Stand: Dezember 2012); die durchschnittliche Länge der Texte beträgt folglich 540 Tokens. Die AutorInnen der Texte sind SchülerInnen allgemeinbildender und berufsqualifizierender Oberschulen mit deutscher Unterrichtssprache, ein Jahr vor der Matura bzw. dem Abitur, aus drei Regionen des deutschen Sprachraums, und zwar aus Südtirol (Italien), Nordtirol (Österreich) und Thüringen (Deutschland). Die TeilnehmerInnen der Erhebung wurden in jeder Region durch eine stratifizierte Zufallsstichprobe mit den Schichtungsmerkmalen *Schultyp* und *Gemeindegröße* ermittelt. Die regionale Verteilung der Daten lässt sich aus Tabelle 1 ablesen:

Subkorpus (nach Region)	Gesamtanzahl
Nordtirol (NTIR):	233.098 Tokens (457 Texte)
Südtirol (STIR):	222.209 Tokens (520 Texte)

2 S. <http://www.korpus-suedtirol.it/index_en> (letzter Abruf September 2013).

Thüringen (TH):	353.674 Tokens (521 Texte)
ohne Angabe	2.349 Tokens (5 Texte)
Gesamt	811.330 Tokens (1503 Texte)

Tab. 1. Regionale Zusammensetzung des Lernerkorpus „KoKo".

Alle Texte wurden im Rahmen des regulären Unterrichts mit denselben Zeitvorgaben und unter vergleichbaren Bedingungen zu ein und demselben Thema verfasst. Die SchülerInnen sollten einen Erörterungsaufsatz zu folgendem Zitat von Hans Magnus Enzensberger schreiben und zum Thema auch persönlich Stellung nehmen:

> Der deutsche Schriftsteller und Essayist Hans Magnus Enzensberger (*1929) hat in einem Interview vom 4. Mai 2001 mit der Wochenzeitung „Die Zeit" unter anderem Folgendes gesagt:
> „Aber wissen Sie, ich finde, die Jugend ist sowieso keine beneidenswerte Phase des Lebens. Ich verstehe gar nicht, warum die Leute so einen Kult damit treiben. Ein junger Mensch ist labil, unsicher, schwankend, hat keine Souveränität, macht jede Dummheit mit. Denken Sie nur an diese Klamottensucht, ein Leben in der Diskothek, schrecklich. Wenn der eine ein Motorrad hat, muss der andere auch eines haben. Das ist doch entsetzlich. Man muss froh sein, wenn man das überstanden hat."

Über eine Fragebogenerhebung wurden umfassende Informationen zu soziolinguistisch relevanten Variablen der SchülerInnen erfasst, wie etwa zu Gemeindegröße, Schultyp, Geschlecht, sozioökonomischem Hintergrund, zur Sprachbiographie (z. B. Dialektgebrauch, Medienkonsum, Schreibgewohnheiten). Diese Informationen sind als Metadaten mit den Korpusdaten verknüpft und ermöglichen ein Filtern der Daten nach unterschiedlichen Variablen.

Die handgeschriebenen Texte wurde zunächst nach vorgegebenen Richtlinien mit einem XML-Editor transkribiert. Anschließend wurden die Daten korpuslinguistisch aufbereitet und abfragbar gemacht. Dazu wurden eine strukturelle und linguistische Annotation, und zwar eine Tokenisierung, eine Lemmatisierung und ein POS-Tagging (Treetagger, Schmid 1994) sowie nötige manuelle Korrekturen durchgeführt. Die Daten wurden für ein Korpusabfragesystem

(CQP, Evert 2009) indiziert und aufbereitet (Corpus Workbench IMS Stuttgart, Christ 1994). Außerdem wurden einige Perl-Scripts erarbeitet, die Filteroptionen und somit die Suche über Metadaten ermöglichen.

Die hypothesentestende, oberflächenorientierte Untersuchung im Hinblick auf das Vorhandensein von Indikatoren konzessiven Argumentierens basiert auf dieser Grundlage. In Anlehnung an Steinhoff (2007: 332ff.) wurden folgende Phänomene in die Analysen einbezogen:

Zum einen wurden *zwar*-Konstruktionen betrachtet. Der Adverbkonnektor *zwar* kommt, mit kataphorischer Verweisrichtung, vorwiegend in zweiteiligen Konstruktionen mit Konnektoren wie *aber, jedoch, trotzdem* etc. vor (vgl. Pasch *et al*. 2003).[3] *Zwar* hat in solchen Konstruktionen eine einräumende Funktion, mit ihm wird zum Kontraargument hingeführt. Die Entkräftung der Konzession wird durch einen Konnektor der *aber*-Klasse eingeleitet (vgl. Steinhoff 2007: 333). Feilke (1996: 217ff.) zählt *zwar …, aber*-Konstruktionen zur Gruppe der „syntaktischen Prägungen". Sie dienen als formale Muster der Kommunikation mit Leerstellen für einzufügende Inhalte. Für das hypothesentestende Vorgehen wurden in der vorliegenden Pilotstudie die Korrelate *aber, jedoch, doch, dennoch, allerdings, ohne, trotzdem, hingegen, nichtsdestotrotz, dagegen, indessen, demgegenüber, gleichwohl* in die Untersuchung einbezogen. Solche Konstruktionen wurden in der Korpusstudie von Steinhoff (2007) in Experten- und Studententexten sowie in einer Pilotstudie von Rezat (2011)[4] teilweise auch in Schülertexten nachgewiesen.

---

3 Zu einem kritischen Überblick bezüglich der grammatischen Konzessivität und deren Beschreibung in unterschiedlichen Grammatiken siehe z. B. Rezat (2007: 93-258).

4 Rezat (2011: 53) verweist darauf, dass im Deutschen eine Reihe von Mitteln zur Markierung konzessiver Relationen zur Verfügung stehen. Sie selbst nimmt auf der Grundlage ihrer empirischen Erhebung konzessiver Konstruktionen (Rezat 2007, 2009) anhand von Expertentexten z. B. eine Dreiteilung syntaktischer Kombinationsmuster vor, die unterschiedliche lexikalische und grammatische Einheiten enthalten. Diese umfassen dieselben Strukturen, die auch Steinhoff 2007 näher untersucht, doch bezieht sie auch weitere lexikali-

Zum anderen wurden Konstruktionen mit Modalwort näher analysiert. Modalwörter dienen dazu, die Einstellung des Sprechers zum Geschehen auszudrücken. Sie beziehen sich nicht auf einzelne Wörter, sondern auf einen ganzen Satz (vgl. Helbig/Buscha 1986: 504-505). Mit Modalwörtern „wird die Geltung einer Proposition explizit thematisiert" (Steinhoff 2007: 347). Die vorliegende Untersuchung beschränkt sich in Anlehnung an Steinhoff (2007: 348) auf die Modalwörter *freilich, natürlich, sicher, sicherlich, zweifellos*. Gemeinsam ist ihnen, „dass sie als *Gewissheitsindikatoren* verwendet werden […]" und „dass sie im Hinblick auf eine Aussage Sicherheit signalisieren" (Steinhoff 2007: 348). Modalwörter können zur „Realisierung einer Konzession, anschließend zur Explizierung einer These verwendet" (Steinhoff 2007: 348) werden. Sie werden oft auch „nicht explizit konzessiv verwendet, d. h. nicht in einem Zusammenhang, wo Gegenargumente explizit angeführt und entkräftet werden, sondern für bloße Einräumungen" (Steinhoff 2007: 348). Häufig sind Modalwörter Teile von zweiteiligen Konstruktionen und werden ähnlich wie *zwar*-Konstruktionen mit Korrelaten der *aber*-Klasse verwendet, z. B. mit *aber, doch, jedoch* etc. (vgl. Steinhoff 2007: 349). Solche zweiteiligen Konstruktionen wurden in der vorliegenden Studie ebenfalls berücksichtigt.

Für die Untersuchung wurden die Vorkommen mittels quantitativer, (semi-)automatischer Korpusabfragen und (Ko-) Okkurrenz-Analysen der korpuslinguistisch aufbereiteten Texte ermittelt, die jeweiligen satzgrenzenüberschreitenden Verwendungskotexte manuell geprüft (Kontrolle hinsichtlich der konzessiven Verwendung der extrahierten Phänomene, jedoch nicht hinsichtlich der kontextuellen Passung) und Vorkommenswahrscheinlichkeiten mit statistischen Maßen berechnet (Log Likelihood, vgl. Dunning 1993; Kilgarriff/Tugwell 2002).

---

sche Mittel der Einräumung in ihre Studie ein, wie z. B. *ich weiß, es stimmt, dass*, sowie Fälle, die unmarkiert bleiben. In der vorliegenden Pilotstudie wird bewusst ausschließlich ein kleiner, klar begrenzter Ausschnitt explizit markierter konzessiver Konstruktionen ausgewählt.

Die folgende Korpusabfrage zur Ermittlung von *zwar*-Konstruktionen mithilfe der CQP-Abfragesyntax dient der Veranschaulichung des Vorgehens:

(6) [targetlemma="zwar"] []* ([targetlemma="aber| allerdings| dagegen| demgegenüber| dennoch| doch| gleichwohl| hingegen| indessen| jedoch| nichtsdestotrotz| ohne| trotzdem"]) within 2s

Die Abfragen bringen Ergebnislisten hervor, wie sie in (7)-(10) aufgeführt sind und die es weiter zu bearbeiten und untersuchen gilt:

(7) ... Auch ein Leben in der Diskothek ist völlig übertrieben. Zwar steigt die Anzahl von Diskogängern in unserer Region erheblich, jedoch gibt es auch genug Jugendliche, die, wie ich ... (ID 1002)

(8) ... nicht mehr an seinen Eltern, man wird selbstständig. Zwar nur in einem gewissen Maß, aber immerhin. Die argumente die Enzensberger gebracht hat, um ... (ID 1005)

(9) ... von Hans Magnus Enzensberger nicht überein. Die Jugend hat zwar auch ihre tiefen aber es ist trotzdem eine wichtige Entwicklungsphase im Leben eines jeden ... (ID 1018)

(10) ... Sachen zu entwickeln oder gar zu verbessern. Man ist zwar froh diese Zeit überstanden zu haben, mit ihren Höhen und Tiefen, aber wenn man erwachsen ist und selber dann Nachkommen hat, ... (ID 1019)

Außerdem wurden exemplarisch qualitative, manuelle Feinanalysen der Verwendungskontexte durchgeführt (u. a. Kontrolle hinsichtlich der Passung der extrahierten Phänomen im Verlauf der konzessiven Argumentation).

Die Ergebnisse werden im folgenden Abschnitt vorgestellt.

3.2. Ergebnisse

Insgesamt finden sich 389 *zwar*-Konstruktionen mit Korrelat in 320 von insgesamt 1503 Texten. Das bedeutet, dass 79% der Texte keine Vorkommen aufweisen. Wenn solche Konstruktionen vorkommen, dann werden sie durchschnittlich einmal, maximal jedoch viermal pro

Text verwendet. Die Untersuchung hat eine eindeutige Präferenz des Vorkommens von *zwar* in Kombination mit *aber* ergeben (222 von 389 *zwar*-Konstruktionen), gefolgt von *jedoch* (80 Vorkommen) und *doch* (61 Vorkommen) (vgl. Tabelle 2).

Anzahl Vorkommen	absolut	Prozent
Gesamt	389	100
Anzahl Texte	absolut	Prozent
mit zwar & Korrelat	320	21
ohne zwar & Korrelat	1183	79
Gesamt	1503	100
Anzahl Vorkommen/Text	absolut	Prozent
4	2	1
3	6	2
2	51	16
1	261	82
Gesamt	320	100
Zwar	Korrelat	Anzahl Vorkommen absolut
Zwar	Aber	222
Zwar	Jedoch	80
Zwar	Doch	61
Zwar	Dennoch	12
Zwar	Ohne	6
Zwar	Trotzdem	4
Zwar	Allerdings	4

Tab. 2. Vorkommen (quantitativ) von *zwar*-Konstruktionen im Lernerkorpus „KoKo".

In der vorliegenden Studie wurden einige Gruppenvergleiche durchgeführt. Dabei wurden die Vorkommenswahrscheinlichkeiten der *zwar*-Konstruktionen bezüglich ihrer regionalen Verteilung, nach

Schultyp und Geschlecht mithilfe von Log Likelihood-Maßen[5] überprüft. Wie aus Tabelle 3 hervorgeht, konnten in keinem Fall signifikante Unterschiede festgestellt werden; die größten Unterschiede gibt es zwischen allgemeinbildenden und berufsqualifizierenden Schulen in Nordtirol.

Anzahl Vorkommen	absolut	Prozent		log likelihood
STIR	107	28	STIR vs. TH	0,10
TH	177	46	STIR vs. NTIR	0,24
NTIR	105	27		
Gesamt	389	100		
Anzahl Vorkommen	absolut	Prozent		log likelihood
STIR AHS	60	56	STIR AHS-BHS	1,79
STIR BHS	47	44		
Gesamt	107	100		
Anzahl Vorkommen	absolut	Prozent		log likelihood
NTIR AHS	60	58	NTIR AHS-BHS	3,25
NTIR BHS	44	42		
Gesamt	104	100		
Anzahl Vorkommen	absolut	Prozent		log likelihood
TH AHS	136	77	TH AHS-BHS	0,25
TH BHS	41	23		
Gesamt	177	100		

5 * signifikanter Unterschied mit $p < 0.5$; kritischer Wert bei 3,84
** signifikanter Unterschied mit $p < 0.01$; kritischer Wert bei 6,63
*** signifikanter Unterschied mit $p < 0.001$; kritischer Wert bei 10,83
**** signifikanter Unterschied mit $p < 0.0001$; kritischer Wert bei 15,13. Zu den kritischen Werten siehe z.B. Rayson/Berridge/Francis 2004.

Anzahl Vorkommen	absolut	Prozent		log likelihood
STIR W	67	63	STIR W-M	0,07
STIR M	39	37		
Gesamt	106	100		
Anzahl Vorkommen	absolut	Prozent		log likelihood
NTIR W	62	62	NTIR W-M	0,23
NTIR M	38	38		
Gesamt	100	100		
Anzahl Vorkommen	absolut	Prozent		log likelihood
TH W	88	51	TH W-M	0,00
TH M	83	49		
Gesamt	171	100		

Tab. 3. Gruppenvergleiche (quantitativ) von *zwar*-Konstruktionen im Lernerkorpus „KoKo“ nach Region, Schultyp und Geschlecht[6].

Zusätzlich zu den *zwar*-Konstruktionen wurden konzessive Konstruktionen mit Modalwort untersucht. 45% der Texte weisen solche Phänomene auf (gezählt wurden Vorkommen mit und ohne Korrelat). Diese Konstruktionen scheinen somit häufiger auf als diejenigen mit *zwar*. Die Vorkommen mit und ohne Korrelat halten sich in der vorliegenden Studie in etwa die Waage (47% vs. 53% der Vorkommen). Die am häufigsten verwendeten Modalwörter sind *natürlich*, *sicher* und *sicherlich*. Die Verwendung der Korrelate schließlich ist jener der *zwar*-Konstruktionen vergleichbar.

Einen Überblick über die Vorkommen der konzessiven Modalkonstruktionen gibt Tabelle 4. Die Vorkommen von Modalwörtern mit den entsprechenden Korrelaten werden in Tabelle 5 dargestellt.

6 Legende der verwendeten Abkürzungen: STIR = Südtirol, NTIR = Nordtirol, TH = Thüringen; AHS = allgemeinbildende Schule, BHS = berufsqualifizierende Schule; W = weiblich, M = männlich.

Anzahl Vorkommen	absolute	Prozent
Modalwort mit Korrelat	984	47
Modalwort ohne Korrelat	1096	53
Gesamt	2080	100
Anzahl Texte	absolute	Prozent
mit Modalw. & Korrelat	675	45
ohne Modalw. & Korrelat	828	55
Gesamt	1503	100
Modalwort	Vorkommen mit Korrelat absolut	Vorkommen ohne Korrelat absolut
Natürlich	597	549
Sicher	198	353
Sicherlich	189	189
Zweifellos	0	3
Freilich	0	2
Anzahl Vorkommen/Text	absolute	Prozent
5	6	1
4	11	2
3	51	8
2	155	23
1	452	67
Gesamt	675	100

Tab. 4. Vorkommen (quantitativ) von Modalwort-Konstruktionen im Lernerkorpus „KoKo“.

Natürlich	Korrelat	Anzahl Vorkommen absolut
Natürlich	Aber	287
Natürlich	Doch	164
Natürlich	Jedoch	85

Natürlich	Ohne	21
Natürlich	Dennoch	14
Natürlich	Allerdings	13
Natürlich	Trotzdem	10
Natürlich	Hingegen	2
Natürlich	Dagegen	1
Gesamt		597
Sicher	Korrelat	Anzahl Vorkommen absolut
Sicher	Aber	94
Sicher	Doch	61
Sicher	Jedoch	24
Sicher	Ohne	8
Sicher	Dennoch	4
Sicher	Trotzdem	3
Sicher	Allerdings	2
Sicher	nichtsdestotrotz	1
Sicher	Hingegen	1
Gesamt		198
Sicherlich	Korrelat	Anzahl Vorkommen absolut
Sicherlich	Aber	94
Sicherlich	Doch	44
Sicherlich	Jedoch	29
Sicherlich	Dennoch	7
Sicherlich	Trotzdem	6
Sicherlich	Ohne	6
Sicherlich	Allerdings	3
Gesamt		189

Tab. 5. Vorkommen (quantitativ) von Modalwort-Konstruktionen und Korrelaten im Lernerkorpus „KoKo“.

Die Gruppenvergleiche (Modalwort-Konstruktionen mit Korrelat) lassen – ebenso wie bei den *zwar*-Konstruktionen – keine signifikanten Unterschiede erkennen (siehe Tabelle 6).

Anzahl Vorkommen	absolut	Prozent		log likelihood
STIR	292	29	STIR vs. TH	1,24
NTIR	287	29	STIR vs. NTIR	0,61
TH	427	42		
Gesamt	1006	100		
Anzahl Vorkommen	absolut	Prozent		log likelihood
STIR AHS	156	53	STIR AHS-BHS	1,69
STIR BHS	136	47		
Gesamt	292	100		
Anzahl Vorkommen	absolut	Prozent		log likelihood
NTIR AHS	146	51	NTIR AHS-BHS	0,10
NTIR BHS	141	49		
Gesamt	287	100		
Anzahl Vorkommen	absolut	Prozent		log likelihood
TH AHS	321	75	TH AHS-BHS	0,00
TH BHS	106	25		
Gesamt	427	100		
Anzahl Vorkommen	absolut	Prozent		log likelihood
STIR W	180	63	STIR W-M	0,03
STIR M	108	38		
Gesamt	288	100		
Anzahl Vorkommen	absolut	Prozent		log likelihood
NTIR W	182	65	NTIR W-M	0,10
NTIR M	97	35		
Gesamt	279	100		

Anzahl Vorkommen	absolut	Prozent		log likelihood
TH W	228	55	TH W-M	0,53
TH M	186	45		
Gesamt	414	100		

Tab. 6. Gruppenvergleiche (quantitativ) von Modalwort-Konstruktionen mit Korrelat im Lernerkorpus „KoKo“ nach Region, Schultyp und Geschlecht.

Zuletzt wurden einige exemplarische Analysen des Verlaufs konzessiver Argumentation, in denen *zwar*-Konstruktionen aufscheinen, manuell untersucht – als Vorarbeit für künftige, zu vertiefende Untersuchungen[7]. Zunächst konnte dabei festgestellt werden, dass konzessives Argumentieren unter der Verwendung der genannten Indikatoren auf zwei Ebenen geschieht: Zum einen sind Reaktionen auf konkrete Argumente erkennbar, im konkreten Fall auf den Inputtext von Hans Magnus Enzensberger, wie das folgende Beispiel zeigt:

(11) Auch ist die Pauschalisierung junge Menschen seien labil nicht richtig. Zwar sind viele Jugendliche anfälliger für Magersucht, Drogen und Sekten, jedoch ist dies nicht allein davon abhängig, wie alt ein Mensch ist, sondern [...]. (ID1156)

Zum anderen werden weitere, zusätzliche Argumente antizipiert, wie etwa in diesem Textausschnitt:

(12) Viele ältere Menschen sagen, dass früher alles anders war. Doch wenn man von Erzählungen hört, war früher doch nicht alles anders. Zwar waren die Lebensverhältnisse und manche Einstellungen unterschiedlich, aber auch zu der Zeit wollte man in die Diskothek gehen oder den Klamottentrend nachgehen [...] (ID1234)

Ausschnitte aus einem Schülertext (siehe Tabelle 7) sollen beispielhaft den Verlauf einer konzessiven Argumentation und einen Analyseansatz verdeutlichen helfen.

---

7 U. a. in Rezat (2009) wurden Verfahren zur textbezogenen Argumentationsanalyse entwickelt.

Textstelle (thematische Einheiten)	Eingeständnis an HME	Argumentation contra HME	Untermauerung/ Begründung zum vorhergehenden Satz
Zugegeben, in einigen Punkten kann man Herrn Enzensberger nichts vorwerfen.	X		
„Ein junger Mensch ist labil, unsicher, schwankend, hat keine Souveränität, macht jede Dummheit mit“			X
Diese Aussage ist zwar sehr verallgemeinernd und wirft alle Jugendlichen in einen Topf, …		X	
… mag aber trotzdem auf viele Jugendliche zutreffen.	X		
Aber erstens ist jeder Mensch ein Individuum und unterscheidet sich somit von allen anderen …			X
… und zweitens: Was ist schon so schlimm daran ?			X
Natürlich ist man unsicher wenn das erste Vorstellungsgespräch bevorsteht.	X		
Natürlich lässt man sich oftmals leicht verführen und beeinflussen, …	X		
… aber was soll man dagegen tun, wenn man bislang noch keine Erfahrungen sammeln konnte und sich nur auf den eigen Instinkt verlassen kann?		X	
Über welches Gesicht schleicht sich kein Lächeln , wenn man an die „guten alten Zeiten“ zurückdenkt und Erinnerungen vorbeiziehen, wie sorglos das Leben noch war und was für Dummheiten man manchmal machte .			X
Natürlich ist man als Jugendlicher bei jeder Dummheit dabei .	X		

Grenzen wollen ertestet werden, man will Respekt und Prestige. Auch wenn man meist erst im Nachhinein realisiert, was man eigentlich getan hat, bzw. was man riskiert hat.		X	
Nach dem Motto: „Schlau war's nicht, aber geil!“			X

Tab. 7. Beispiel – Verlauf eine konzessiven Argumentation im Lernerkorpus „KoKo“.

Am Beispiel in Tabelle 7 zeigt sich, dass zunächst jeweils Eingeständnisse an Hans Magnus Enzensberger (HME) gemacht, anschließend Proargumente für die eigene These vorgebracht werden. Auch Untermauerungen bzw. Begründungen zu vorgebrachten Argumenten sind feststellbar. Bei genauerer Analyse fällt auf, dass es im dritten Satz („Diese Aussage ist zwar sehr…“) genau umgekehrt ist: Im *zwar*-Teil befindet sich die These gegen HME bzw. die eigene Position, während der *aber*-Teil das Eingeständnis enthält. Auffälligerweise wird im daran anschließenden Satz, der an dieser Stelle fälschlicherweise mit dem Konnektor *aber* eingeleitet wird, eine Begründung ausgedrückt, die im Grunde genommen die Gegenthese, also die Argumentation contra HME, stützt (nämlich, dass man nicht alle Jugendlichen in einen Topf werfen kann bzw. dass jeder Mensch ein Individuum ist, das sich von anderen unterscheidet). Ein Anschluss mit kausalem *denn* o. Ä. hingegen würde mit der restlichen Argumentation übereinstimmen. Zukünftige Analysen müssen zeigen, ob weitere Schwierigkeiten beim konzessiven Argumentieren im „KoKo“-Korpus auftreten.

4. Diskussion und Ausblick

Die vorgestellte Pilotuntersuchung zur Verwendung konzessiver Argumentationsstrukturen hat Folgendes gezeigt: Die expliziten *zwar*-Konstruktionen mit Korrelat werden nur in etwa 20% der Schülertexte

und darin durchschnittlich einmal pro Text verwandt. Dabei werden *aber, jedoch* und *doch* am häufigsten als Korrelat eingesetzt. Ähnlich verteilte Präferenzen zeigen beispielsweise auch die Vorkommen in den Experten- und Studententexten in der Untersuchung von Steinhoff (2007). In Steinhoffs Untersuchung wurden ebenfalls argumentative Texte untersucht, in denen Ausdrucksmittel der „Alltäglichen Wissenschaftssprache" und Merkmale konzessiven Argumentierens zu beobachten sind. Allerdings ist zu betonen, dass ein direkter Vergleich mit seinen Ergebnissen keinesfalls zulässig ist, da sein Korpus aus Zeitschriftenartikeln professioneller Wissenschaftler und aus studentischen Hausarbeiten zusammengestellt ist, die folglich aus einer anderen Domäne stammen und außerdem durchschnittlich viel länger als die Schülertexte des „KoKo"-Korpus sind[8].

Zwar-Konstruktionen mit Korrelat zur konzessiven Argumentation sind somit kein sehr häufiges, aber auf jeden Fall ein nachweisbares Phänomen in Texten von OberschülerInnen[9]. Zumindest für die *zwar…, aber*-Konstruktion ist dies nicht überraschend. Es gibt Hinweise darauf, dass die „Verwendung des Konnektivs ‚aber' […] prototypisch für den Erwerb konzessiven Argumentierens zu sein [scheint]" (Rezat 2011: 59). (Prä)konzessive Argumentationsstrukturen sind in Ansätzen bereits im Grundschulalter feststellbar. Dabei treten Strukturen mit *aber* vergleichsweise häufig auf (vgl. Rezat 2011; auch Augst *et al.* 2007), eine eindeutig konzessive Verwendung scheint in argumentativen Grundschultexten allerdings selten vorzuliegen. Dennoch spricht nach Rezat (2011: 59) „vieles für die Annahme, dass der Erwerb konzessiver literaler Prozeduren über adversative

8 Steinhoffs Daten stammen aus verschiedenen geisteswissenschaftlichen Bereichen. Das „Studententexte-Korpus" umfasst 296 Texte (insgesamt 1.668.655 Wörter) mit jeweils durchschnittlich 5.637 Wörtern, das „Expertentexte-Korpus" hingegen 99 Texte (insgesamt 863.084 Wörter) mit jeweils durchschnittlich 8.718 Wörtern. Außerdem wurde auf ein Vergleichskorpus mit journalistischen Texten aus Cosmas 2 des IDS Mannheim zugegriffen (siehe dazu Steinhoff 2007: 151ff.).

9 Bei Steinhoff (2007) weisen insgesamt 13% der Experten- und 21% der Studententexte keine Vorkommen solcher Konstruktionen auf.

literale Prozeduren – hier: ‚aber' – erfolgt. ‚Aber' bildet sozusagen, die Schaltstelle im Erwerb, die dann konzessiv ausdifferenziert wird."

Weitere alternative Korrelate treten später hinzu, wenn auch häufiger konzessiv argumentiert wird. Rezat (2011: 60) beispielsweise weist in einer von ihr durchgeführten Pilotstudie solche Konstruktionen öfter in Sekundar- als in Grundschultexten nach; in den Texten der HauptschülerInnen stellt sie keine Vorkommen fest – im Unterschied zu denen der GymnasialschülerInnen[10]. Die Gruppenvergleiche der vorliegenden Studie haben hingegen weder bezüglich der unterschiedlichen Schultypen (AHS vs. BHS), der drei getesteten Regionen (Südtirol, Nordtirol, Thüringen) noch bezüglich des Geschlechts signifikante Unterschiede ergeben.

Was die Verwendung konzessiver Strukturen mit lexikalischen und grammatischen Einheiten betrifft, stellt Rezat in ihrer Untersuchung fest, dass in den Texten der GymnasialschülerInnen der „einräumende Teil der Konzession bis auf wenig Ausnahmen markiert [wird], und zwar durch lexikalische Mittel der Einräumung oder allgemeine Verstärkungsmittel des Gegensatzes." (Rezat 2011: 61). Am häufigsten treten solche Strukturen auf, in denen die Gegenbehauptung durch adversative Konnektive wie *aber, doch, jedoch, allerdings*, die Einräumung durch lexikalische Mittel wie *es ist mit Sicherheit so, dass …, in gewisser Weise teile ich ihre Ansicht, vielleicht* etc. oder durch ein allgemeines Verstärkungselement wie *zwar, sicherlich, natürlich* markiert wird (vgl. Rezat 2011: 62).

Häufiger als adversative Konnektive konnten im „KoKo"-Korpus konzessive Konstruktionen mit Modalwort nachgewiesen werden. In Steinhoffs Studie (2007: 334ff.) ist hingegen eine gegenläufige Tendenz bemerkbar. Die am häufigsten verwendeten Modalwörter im „KoKo"-Korpus sind *natürlich, sicher* und *sicherlich*, dieselben, die

10 Bei diesem Ergebnis (Rezat 2011) muss jedoch die kleine Pilotierungsstichprobe berücksichtigt werden: Die Grundlage dazu bilden Texte zu argumentativen Schreibaufträgen von SchülerInnen unterschiedlicher Jahrgangsstufen (3,4,8,9) und Schulformen (Grundschule, Hauptschule, Gymnasium). Das Korpus der GrundschülerInnen umfasst 47 Texte, das der Sekundarschülerinnen 62 Texte (davon 14 von HauptschülerInnen). (siehe dazu Rezat 2011: 54; über die Textlänge gibt es in dem Beitrag keine Informationen).

auch in Steinhoffs Studentenkorpus am häufigsten verwendet werden. In seinem Expertentexte-Korpus sind es *zweifellos, natürlich* und *freilich*. Er weist darauf hin, dass StudentInnen Modalwörter häufiger explizit konzessiv verwenden als ExpertInnen (Steinhoff 2007: 352). Dies lässt auf die Tendenz schließen, dass unerfahrene SchreiberInnen den/die LeserIn eher explizit durch einen Text führen, was teilweise an einer Übererfüllung der entsprechenden Konventionen erkennbar ist. Diese Annahme könnte auch das relativ häufige Vorkommen von Modalwörtern in konzessiven Argumentationen im „KoKo"-Korpus erklären.

Resümierend lässt sich festhalten, dass mit den angewandten Methoden in den untersuchten Schülertexten Vorkommen konzessiver argumentativer Strukturen nachgewiesen werden konnten. Die quantitativ ermittelten, auf die konzessive Verwendung hin überprüften Vorkommen lassen, wie erwähnt, keine signifikanten Unterschiede bezüglich Region (Südtirol, Nordtirol, Thüringen), Schultyp (allgemeinbildende vs. berufsqualifizierende Oberschulen) und Geschlecht erkennen. Es gibt hingegen unterschiedliche qualitative Ausprägungen der Vorkommen, die es näher zu untersuchen gilt. Während für die Abfrage lexikalischer und grammatischer Phänomene und dabei z. B. die Bewältigung großer Textmengen ein korpuslinguistischer Zugang Vorteile bringt, hat die Analyse längerer Textausschnitte (s. oben Tab. 7) gezeigt, dass für detailliertere Beobachtungen ein quantitativer, im Wesentlichen auf Kotexte beschränkter Ansatz nur bedingt geeignet ist.

Als nächste Schritte stehen daher umfassende Feinanalysen konzessiver argumentativer Strukturen an, die u. a. Einzeläußerungen sowie die kontextuelle Passung betreffen. Auch sollen Strukturen ohne explizite Markierung der Einräumung oder der Gegenargumente bzw. gänzlich ohne erkennbare lexikalische Indikatoren an der Textoberfläche in die Analysen einbezogen werden.

Einige Anregungen für die Didaktik sollen den Beitrag abrunden: Konzessive Konstruktionen, die sich des Ausdrucksinventars der „Alltäglichen Wissenschaftssprache" bedienen, stellen für das Verfassen argumentativer Texte eine wichtige sprachliche Handlungsressource bereit, deren Nutzung die Bildungsinstitutionen vermitteln helfen sollen (vgl. Ehlich/Valtin/Lütke 2012: 25-27). Bereits ab der

Grundschule kann an argumentatives Schreiben herangeführt werden, während elaborierte argumentative Kompetenzen im Wesentlichen erst im frühen Erwachsenenalter entfaltet werden können (vgl. z. B. Feilke/Augst 1989; Becker-Mrotzek 1997; Pohl 2007; Steinhoff 2007: 43, 131). Ein produktiver *und* rezeptiver Umgang mit argumentativen Texten bildet eine wichtige Voraussetzung für den Aufbau von Argumentationskompetenz (vgl. Garaté/Melero 2004). SchülerInnen der Oberstufe können bereits ansatzweise mit Handlungsmustern wissenschaftlicher Texte vertraut gemacht werden und erste Erfahrungen mit wissenschaftlichen Texten, z. B. durch das Schreiben einer Facharbeit, sammeln (vgl. Steets 1999, 2003). Dabei ist es möglich, die Aufmerksamkeit ganz bewusst auf bestimmte Teilbereiche des (wissenschaftlichen) Schreibens, u. a. auf konzessive argumentative Strukturen, zu lenken.

Literatur

Abel, Andrea / Anstein, Stefanie 2011. Korpus Südtirol – Varietätenlinguistische Untersuchungen. In Abel, Andrea / Zanin, Renata (Hgg.) *Korpora in Lehre und Forschung*. Bozen-Bolzano: University Press, 29-54.

Abel, Andrea / Glaznieks, Aivars im Druck. Wo Sprachkompetenzforschung auf Varietätenlinguistik trifft: Empirische Befunde aus dem Varietäten-Lernerkorpus „KoKo“. In Lenz, Alexandra / Glauninger, Manfred (Hgg.) *Variation und Varietäten des Deutschen in Österreich - Theoretische und empirische Perspektiven.* Frankfurt/M.: Peter Lang (i.D.).

Anstein, Stefanie / Oberhammer, Margit / Petrakis, Stefanos 2011. Korpus Südtirol – Aufbau und Abfrage. In Abel, Andrea / Zanin, Renata (Hgg.) *Korpora in Lehre und Forschung*. Bozen-Bolzano: University Press, 15-28.

Augst, Gerhard / Disselhoff, Katrin / Henrich, Alexandra / Pohl, Thorsten / Völzing, Paul-Ludwig 2007. *Text – Sorten – Kompe-*

tenz. Eine echte Longitudinalstudie zur Entwicklung der Textkompetenz im Grundschulalter. Frankfurt/M.: Peter Lang.

Augst, Gerhard / Faigel, Peter 1986. *Von der Reihung zur Gestaltung. Untersuchungen zur Ontogenese der schriftsprachlichen Fähigkeiten von 13-23 Jahren*. Unter Mitarbeit von Karin Müller und Helmuth Feilke. Frankfurt/M.: Peter Lang.

Becker-Mrotzek, Michael 1997. *Schreibentwicklung und Textproduktion. Der Erwerb der Schreibtätigkeit am Beispiel der Bedienungsanleitung.* Opladen: Westdeutscher Verlag.

Becker-Mrotzek, Michael / Böttcher, Ingrid 2006. *Schreibkompetenz entwickeln und beurteilen.* Berlin: Cornelsen Scriptor Verlag.

Christ, Oliver 1994. A Modular and Flexible Architecture for an Integrated Corpus Query System. In *Proceedings of COMPLEX 1994, 3rd Conference on Computational Lexicography and Text Research - Budapest, Hungary, July 7-10*, 23-32.

Coirier, Pierre / Golder, Caroline 1993. Writing Argumentative Text. A Developmental Study of the Acquisition of Supporting Structures. *European Journal of Psychology of Education* VIII/2, 169-181.

Dittmann, Jürgen / Geneuss, Katrin A. / Nennstiel, Christoph / Quast, Nora A. 2003. Schreibprobleme im Studium – Eine empirische Untersuchung. In Ehlich, Konrad / Steets, Angelika (Hgg.) *Wissenschaftlich schreiben – lehren und lernen*. Berlin – New York: de Gruyter, 155-185.

Dunning, Ted 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* 19/1, 61-74.

Ehlich, Konrad / Graefen, Gabriele 2001. Sprachliches Handeln als Medium diskursiven Denkens. Überlegungen zur sukkursiven Einübung in die deutsche Wissenschaftskommunikation. *Jahrbuch Deutsch als Fremdsprache* 27, 351-378.

Ehlich, Konrad / Valtin, Renate / Lütke, Beate 2012. *Expertise „Erfolgreiche Sprachförderung unter Berücksichtigung der besonderen Situation Berlins“*. <http://www.berlin.de/imperia/md/content/sen-bildung/schulqualitaet/expertise_sprachfoerderung.pdf?start&ts=1355928831&file=expertise_sprachfoerderung.pdf> (letzter Abruf September 2013).

Ehlich, Konrad 1993. Deutsch als fremde Wissenschaftssprache. *Jahrbuch Deutsch als Fremdsprache* 19, 13-42.

Ehlich, Konrad 1999. Alltägliche Wissenschaftssprache. *InfoDaF* 1/1999, 3-24.

Evert, Stefan 2009. *The CQP query language tutorial*. Technical report, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. <http://cwb.sourceforge.net/temp/CQPTutorial.pdf> (letzter Abruf September 2013).

Feilke, Helmuth 1993. Sprachlicher Common sense und Kommunikation. Über den gesunden Menschenverstand, die Prägung der Kompetenz und die idiomatische Ordnung des Verstehens. *Der Deutschunterricht* 6, 6-21.

Feilke, Helmuth 1994. *Common sense-Kompetenz. Überlegungen zu einer Theorie des ‚sympathischen' und ‚natürlichen' Meinens und Verstehens*. Frankfurt/M.: Suhrkamp.

Feilke, Helmuth 1996. *Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik*. Frankfurt/M.: Suhrkamp.

Feilke, Helmuth 1998. Idiomatische Prägung. In Barz, Irmhild / Öhlschläger, Günther (Hgg.) *Zwischen Grammatik und Lexikon*. Tübingen: Niemeyer, 69-80.

Feilke, Helmuth 2003. Entwicklung schriftlich-konzeptualer Fähigkeiten. In Bredel, Ursula / Günther, Hartmut / Klotz, Peter / Ossner, Jakob / Siebert-Ott, Gesa (Hgg.) *Didaktik der deutschen Sprache. Ein Handbuch*. Band 1. Paderborn – München – Wien – Zürich: UTB GmbH, 178-192.

Feilke, Helmuth 2010. Schriftliches Argumentieren zwischen Nähe und Distanz – am Beispiel wissenschaftlichen Schreibens. In Ágel, Vilmos / Henning, Mathilde (Hgg.) *Nähe und Distanz im Kontext variationslinguistischer Forschung*. Berlin – New York: de Gruyter, 209-231.

Feilke, Helmuth / Augst, Gerhard 1989. Zur Ontogenese der Schreibkompetenz. In Antos, Gerd / Krings, Hans P. (Hgg.) *Textproduktion. Ein interdisziplinärer Forschungsüberblick*. Tübingen: Niemeyer, 297-327.

Garaté, Milagros / Melero, Angeles 2004. Teaching How to Write Argumentative Texts at Primary Schools. In Rijlaarsdam, Gert *et*

al. (Hgg.) *Effective learning and teaching of writing*. 2nd edition. Amsterdam: Kluwer, 323-337.

Gogolin, Ingrid / Lange, Imke 2010. Bildungssprache und Durchgängige Sprachbildung. In Fürstenau, Sara / Gomolla, Mechthild (Hgg.) *Migration und schulischer Wandel: Mehrsprachigkeit*. Wiesbaden: VS-Verlag, 107-127.

Habermas, Jürgen 1981. Umgangssprache, Bildungssprache, Wissenschaftssprache. In Habermas, Jürgen (Hg.) *Kleine politische Schriften I–IV*. Frankfurt/M.: Suhrkamp, 340-363.

Helbig, Gerhard / Buscha, Joachim 1986. *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. 12., unveränderte Auflage. Leipzig: Enzyklopädie.

Jechle, Thomas 1992. *Kommunikatives Schreiben. Prozeß und Entwicklung aus der Sicht der kognitiven Schreibforschung*. Tübingen: Narr.

Kilgarriff, Adam / Tugwell, David 2002. Sketching Words. In Corréard, Marie-Hélène (Hg.) *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*, Euralex, 125-137.

Klieme, Eckhard *et al*. 2003. *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Bonn: BMBF.

Leitao, Selma 2003. Evaluating and Selecting Counterarguments. Studies of Children's Rhetorical Awareness. *Written Communication* 20/3, 269-306.

Nussbaumer, Markus / Sieber, Peter 1994. Texte analysieren mit dem Zürcher Textanalyseraster. In Sieber, Peter (Hg.) *Sprachfähigkeiten. Besser als ihr Ruf und nötiger denn je*. Aarau: Verlag Sauerländer, 141-186.

Paek, Solja 1993. *Die sprachliche Form hypothetischen Denkens in der Wissenschaftssprache*. München: Iudicium.

Pasch, Renate / Brauße, Ursula / Breindl, Eva / Waßner, Ulrich Hermann 2003. *Handbuch der deutschen Konnektoren. Linguistische Grundlagen der Beschreibung und syntaktische Merkmale der deutschen Satzverknüpfer (Konjunktionen, Satzadverbien und Partikeln)*. Berlin - New York: de Gruyter.

Pohl, Torsten 2007. *Studien zur Ontogenese wissenschaftlichen Schreibens*. Tübingen: Niemeyer.

Rayson, Paul / Berridge, Damon / Francis, Brian 2004. Extending the Cochran rule for the comparison of word frequencies between corpora. In Purnelle Gérard / Fairon Cédrick / Dister Anne (Hgg.) *Le poids des mots: Proceedings of the 7th International Conference on Statistical analysis of textual data* (JADT 2004), *Louvain-la-Neuve, Belgium, March 10-12, 2004 (vol. II)*. Louvain: Presses universitaires de Louvain, 926 – 936.

Redder, Angelika 2014. Wissenschaftssprache – Bildungssprache – Lehr-Lern-Diskurs. In Hornung Antonie / Carobbio, Gabriella / Sorrentino, Daniela (Hgg.) *Diskursive und textuelle Strukturen in der Hochschuldidaktik. Deutsch und Italienisch im Vergleich*. Münster: Waxmann.

Rezat, Sara 2007. *Die Konzession als strategisches Spiel*. Heidelberg: Universitätsverlag Winter.

Rezat, Sara 2009. Konzessive Konstruktionen. Ein Verfahren zur Rekonstruktion von Konzessionen. *Zeitschrift für Germanistische Linguistik* 3, 8-12.

Rezat, Sara 2011. Schriftliches Argumentieren: Zur Ontogenese konzessiver Argumentationskompetenz. *Didaktik Deutsch* 17/31, 50-67.

Schmid, Helmut 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing* <ftp://ftp.ims.uni-stuttgart.de/pub/corpora/tree-tagger1.pdf> (letzter Abruf September 2013).

Schmidlin, Regula / Feilke, Helmuth 2005. Forschung zur literalen Textkompetenz – Theorie und Methodenentwicklung. In Feilke, Helmuth / Schmidlin, Regula (Hgg.) *Literale Textentwicklung*. Forum Angewandte Linguistik. Bd. 45. Frankfurt/M.: Peter Lang, 7-18.

Schneuwly, Bernard 1988. *Le langage ecrit chez l'enfant. La production des textes informatifs et argumentatifs*. Lausanne: Delachaux et Niestlé.

Schründer-Lenzen, Agi / Henn, Dominik 2009. Entwicklung eines Computerprogramms zur Analyse der schriftlichen Erzählfähigkeit. In Lengyel, Drorit / Reich, Hans / Roth, Hans-

Joachim / Döll, Marion (Hgg.) *Von der Sprachdiagnose zur Sprachförderung*. Münster: Waxmann, 91-107.

Steets, Angelika 1999. Schreiben in der Oberstufe – Überlegungen zu einer wissenschaftspropädeutischen Schreibausbildung. *Mitteilungen des Deutschen Germanistenverbandes* 3, 398-421.

Steets, Angelika 2003. Wissenschaftspropädeutik in der Oberstufe: Die Facharbeit. *Der Deutschunterricht* 3, 58-70.

Steinhoff, Torsten 2007. *Wissenschaftliche Textkompetenz. Sprachgebrauch und Schreibentwicklung in wissenschaftlichen Texten von Studenten und Experten.* Tübingen: Niemeyer.

Tente, Christina / Furchner, Ingrid / Ruhmann, Gabriela 1999. Von der Schreibberatung für Studierende zur Lehrberatung für Dozenten. In Kruse, Otto / Jakobs, Eva-Maria / Ruhmann, Gabriela (Hgg.) *Schlüsselkompetenz Schreiben. Konzepte, Methoden, Projekte für Schreibberatung und Schreibdidaktik an der Hochschule*. Neuwied: Luchterhand, 61-72.

I corpora in contesti pedagogici

DENISE MILIZIA

5. Phrasal Verbs and Phrasal Units: Political Corpora Within the Walls of the Classroom

1. Introduction

This chapter has a twofold purpose. First, it focusses on a feature much dreaded by learners, that is phrasal verbs, those verbs followed by one, two or even more adverbial or prepositional particles, each of which has a semantically autonomous existence but together they make a meaning that cannot be predicted from the meanings of the parts. All the examples provided here are taken from an 18 million-word corpus of political speeches (henceforth ABC)[1] delivered by British and American politicians from November 2001 to November 2011. Second, the present research aims to prove the several advantages of teaching and learning through corpora, drawing only on real-life language data, i.e. on language that has in fact occurred (Römer 2008: 116), giving students a richer language-learning experience (Dodd 1997: 131) in a significantly enriched learning environment (Aston 1997: 51).

After more than three decades of research on electronic corpora, many course books are still lagging behind this development and seem to be rather unaffected by advances in Corpus Linguistics. Kennedy (1992: 366) cautions that "invented examples can present a distorted version of typicality or an over-tidy picture of the system"; Sinclair (1991: 5) calls it "an absurd notion that invented examples can actually represent the language better than real ones": we cannot trust

1 The corpus, originally called BBB from the initials of the three politicians under investigation, Blair, Bush and Berlusconi, has recently been baptized ABC, American British Corpus.

our ability to make up examples, and this practice would also be "just nonsense" from a semantic point of view (Firth 1957: 24).

As Tribble (2012) has shown, Corpus Linguistics in the classroom still remains a minority sport. In his last survey it emerged that the gap between corpora and the teaching reality is still far too wide, and the extent to which corpora and concordances are being used in the classroom is as yet fairly limited (Römer 2008: 123). After an informal survey carried out among colleagues, it seems to me that the reason why corpora are still not relied on as much as they should be lies in the fact that, being corpora genuine language, it is often believed that they are likely to include language which is too convoluted, too complicated to start with, hence they should become part of the syllabus only in higher levels of achievement. Thus, most educators prefer, for example, to teach one-word rather than multi-word verbs, starting with expressions like *Barack Obama took office in January 2009*, or *Barack Obama took the sacred oath in January 2009*, procrastinating structures like *Barack Obama was sworn in as the 44th President of the United States in January 2009* to the latest possible moment. Hence, not only do teachers tend to postpone the teaching of phrasal verbs, but also learners tend to avoid their use as long as they can although native speakers not only manage them with aplomb but seem to prefer them to single-word alternatives.

Undoubtedly, learning the Latinate verb is easier than learning the phrasal verb, especially when such a form has a correspondent in the students' native language (Side 1990: 145), as is the case of *tolerate* for Italian students, even though the instances of *put up with* in our spoken political corpus are higher in terms of frequency than those of *tolerate*. Contrary to common belief, substituting the phrasal verb with a single-word equivalent or synonym very often seems out of place and sounds rather awkward, given that in several cases phrasal verbs and their synonyms have different ranges of use, meaning and collocations (Sinclair/Moon 2011: v). Students are often taught phrasal verbs in combination with their Romance one-word 'equivalent' – *tolerate* equals *put up with*, *postpone* is the same as *put off*, *look into* means *investigate* – but, as Quirk *et al*. (1985: 1162) maintain, the substitution of a multi-word verb with a single-word verb may not be entirely reliable. In this respect, McArthur and Atkins (1974: 6) add that verbs of classical origin may have similar meanings

but may have often unsuitable overtones of formality, pomposity or difficulty.

It is often said (Sinclair/Moon 2011) that phrasal verbs tend to be rather colloquial or informal and more appropriate to spoken than written English: this was confirmed in our analysis of EU documents, where only five phrasal verbs were found: *lay down*, *provide for*, *set out*, *agree upon*, *carry out*. Such scant presence in this type of specialized language is hardly surprising, given also that legal language, and European legal language in particular, exhibits several differences with respect to general written English, being even more restricted and farther from spoken language. In this respect, Reppen (2010) shows that in academic writing almost overwhelmingly a single-word verb is preferred to a multi-word verb which, instead, is usually avoided. Conversely, spoken English abounds in multi-word verbs, be they phrasal, prepositional, or phrasal-prepositional, hence they do need to be taught if we want our students to sound native-like and if native English is still to be regarded as a target model. It is worth underlining that, even though the current debate on English as an international language is questioning the status of the idealized native speaker as a target model (Bernardini 2004: 30), we cannot deny that most of us rely on native speaker language performance and strive to achieve the native standard as closely as possible.

It is worth adding, at this point, that most political speeches are not *ad lib*, off-the-cuff, as it were: speeches proper and statements are often pre-prepared, written-to-be-spoken, whereas interviews are the most extemporaneous, thus exhibiting similar features to spoken English. It is in fact in interviews that we most frequently encounter phraseology, in the form of both phrasal verbs and phrases. We are not going to delve here into the complicated issue of the differences between spoken and written language, but it cannot be denied that spoken language is often believed to be more chaotic, incomplete, convoluted, and containing linguistic phenomena that defy traditional written norms, which would be regarded by pedagogical grammars as proscribed and unacceptable. Whether this is true or not, what students need the most is to be exposed to real language, and setting these features aside from the teaching material would mean to make them a disservice. We need spoken data for teaching spoken language, and we need authentic data to teach authentic language. Spoken data can

achieve high authenticity, serve as communication aids, and provide irreplaceable models of the target language (Mauranen 2004a).

The present chapter presents a project carried out with undergraduates at the Department of Political Science at the University of Bari. The main purpose of the research was to allow students to work exclusively with real and attested spoken language, in order to capture the regularities/irregularities that are reflected only in authentic data. Phrasal verbs, in the case in point, receive scant attention both in traditional books and dictionaries, and are usually relegated to the end of the book. It is well-known that phrasal verbs create endless problems for learners: they seem to be anarchic, unstable, one-off items that do not fit into a tidy description and relate in mysterious ways to word meaning (Sinclair 1996); yet they are very common in spoken language, in both political and general language, and corpora are an invaluable resource to enable students, and teachers alike, to learn (and teach) the words and the phrases that the learners will use later on in order to handle texts in their subject areas (Gavioli 2006).

2. Ambiguity and disambiguity of one-word and multi-word verbs

Students at Political Science are particularly motivated and interested in analysing political speeches, not only because being informed on topical issues is part of their syllabus, but also because they do not major in English nor are they likely to become language professionals, and one of the main skills they will need in their professional lives is giving presentations: what they mostly need is record language spoken in situations in which they are likely to find themselves (Aston *et al*. 2004).

As to their level of linguistic competence, the initial entry test submitted on the first day of the semester (March 2010) had shown that 70% of students had reached a B1 level (although not all of them had earned the PET certificate), 20% had reached a B2 level (with

only ten students holding a FCE). Only a very tiny percentage had a good mastery of the English language, with two students holding the TOEFL certificate and one the Proficiency Certificate of English released by Cambridge University. The remaining were students who had never studied English before (at least so they claimed), or had attended an English course many years ago, mainly adults and working students.

We very often watch and listen to the videos in class, despite the lack of ad-hoc labs and the huge number of attendees – between 400 and 500. Our courses, apart from a short module on the language of the Treaty of Lisbon and the failed European Constitution, focus mainly on spoken language, in particular on the language as uttered by British and American politicians in government and in opposition. At the end of the semester, students are invited to carry out their own project choosing one or more items to investigate: those who accept the challenge become enthusiastic about the insights they gain and come up with very interesting results that may be sometimes unknown to the teacher. Most students are particularly attracted by ambiguity and synonymy and, unsurprisingly, phrases and phrasal verbs. The verbs they usually choose as an item to analyse are *take*, *make*, *give* and *turn*: these are verbs they are all familiar with when taken individually, and what fascinates them most is how they change their meaning when in context and, first and foremost, how several multi-word verbs have no semantic relationship with the single-word lexical verb.

A phrasal verb is a difficult word to isolate semantically and it has to be put into context because in most of its usage it contributes to meaning in combination with other words. Frustrating enough for the student, most of the time also the one-word lexical verb needs its language environment in order to carry meaning, as is the case of the verb *turn*, which, as mentioned earlier, is well-known to students in their first year of university. What needs to be highlighted, though, is that the meaning they are familiar with is the one that the dictionaries, both monolingual and bilingual, rank first, as the figures below show:

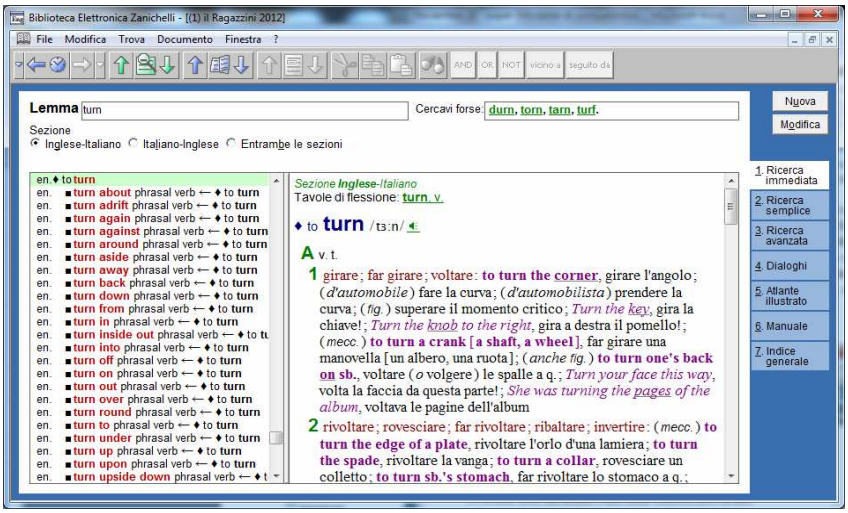


Figure 1a. The verb *turn* in the *Ragazzini* (2012).

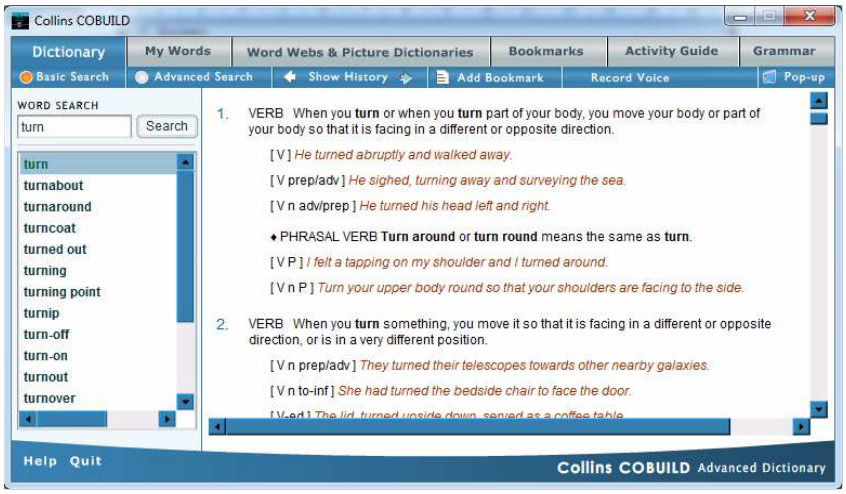


Figure 1b. The verb *turn* in the *Collins Cobuild* (2011).

It did not come as a surprise that the verb *turn* with the meaning of "move something / move your body or part of your body so that it is facing in a different or opposite direction" was hardly found in the

spoken political corpus, where the semantic preference of this verb is made of collocates belonging to the field of age:

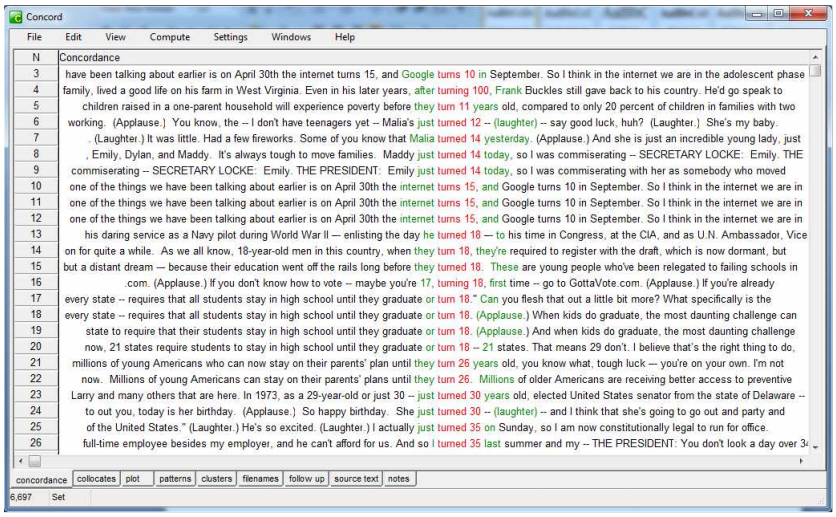


Figure 2. The verb *turn* in the spoken political corpus.

Even though the sense of *turn* followed by a number, as shown in Figure 2, was somewhat unfamiliar to most students, its meaning was eventually disambiguated in context. Yet, the semantic field of age was not immediately clear, given that the well-known expression *years old* was spelt out only in very few cases (lines 5, 21 and 23). Conversely, in the bilingual dictionary, only after the 20 transitive meanings and the 15 intransitive meanings of the verb *turn*, and only after all the idiomatic phrases that the verb lends itself to create, was one example provided with the meaning which, instead, is the most common in ABC (namely, "he has just turned twenty").

The verb *turn* is particularly rich in making combinations with grammatical words like *about, away, back, down, in, off, on, out*, and these words are themselves very common. The phrasal verbs with *turn* that the vast majority of students had already encountered in their high school studies were *turn on* and *turn off*, and to a lower degree, *turn down* and *turn up*, but for all of them the collocates were usually a heater, the radio, or other piece of equipment, thus with the meaning

of "increasing/lowering the amount of sound, heat or power being produced, by adjusting the controls" (*Collins Cobuild Advanced Dictionary* 2011: 1685).

Apart from the very frequent cluster *turn a blind eye*, in the investigation of the verb *turn* we found out that the most frequent particle collocating with *turn* was *out*, followed by *over* and *back*, as the two-word and the three-word clusters generated by *WordSmith Tools* (Scott 2007) illustrate below:

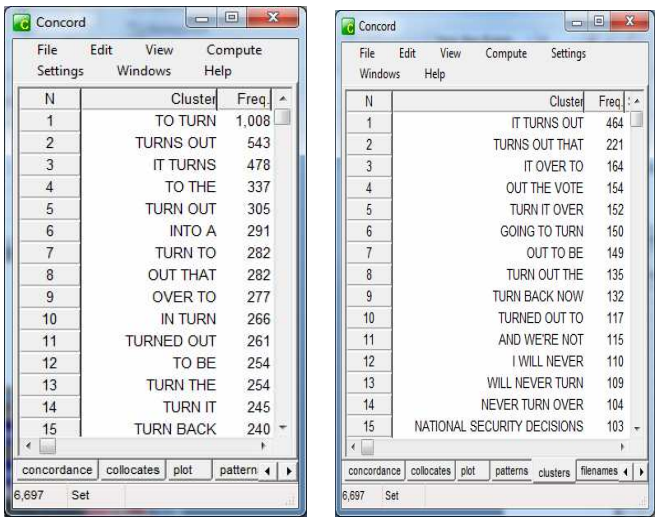


Figure 3. Two- and three-word clusters around the verb *turn* in ABC.

Interestingly, the verb *turn out* was new to most students. It was thus found out that the dictionary, which used to be an invaluable and crucially important resource in traditional courses, barely helps on some occasions: what is really needed is the context of a word, that only corpora can provide. It was soon evident that the verb *turn out* had no semantic relationship with *turn*, and that, just like single-word lexical verbs, phrasal verbs are prone to multiple ambiguity, with some having a meaning when used transitively and another meaning when used intransitively (*Collins Cobuild Phrasal Verb Dictionary* 2011).

Confusing enough for the students, it soon emerged that the two most frequent uses of *turn out* in ABC were both intransitive, as the concordance lines below indicate:

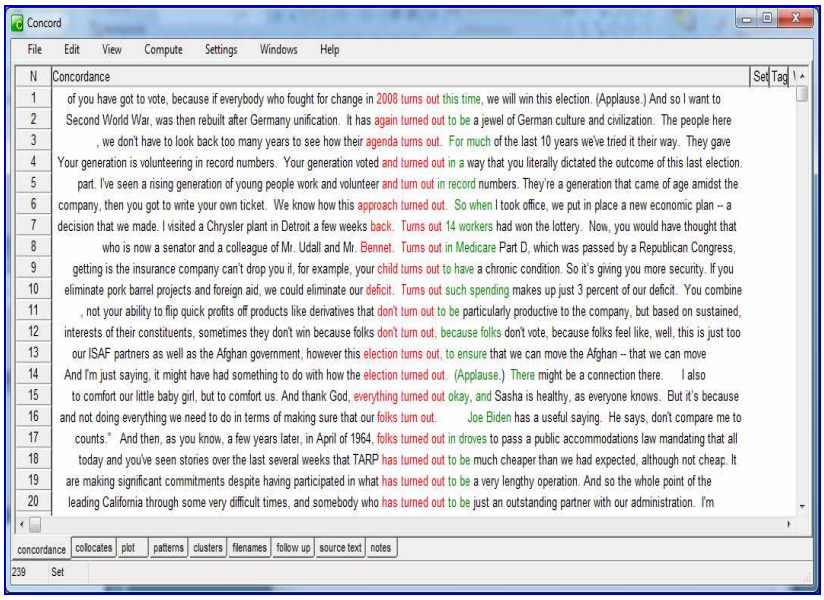


Figure 4. Concordance lines of the verb *turn out* in ABC.

Enlarging the context in lines 1 and 4 respectively, we read that "all of you have got to vote, because if everybody who fought for change in 2008 turns out this time, we will win the elections", and "Your generation voted and turned out in a way that you literally dictated the outcome of this last election". It is clear from the evidence of the data that ABC features less salient meanings than those provided in dictionaries.

Students immediately noticed that in lines 7 and 10 in Figure 4 the subject of the sentence is omitted: this omission would be regarded as an aberration by written grammarians; indeed aberrations do occur in spoken performances, even more "important" than the simple omission of the subject, and it would be a disservice to our learners to have them believe the contrary.

Relying on the two-, three- and four-word clusters facility provided by *WordSmith Tools*, of all the phrasal verbs we managed to unveil in ABC, *turn out* ranked 21st (Figure 5). Even though it was not positioned top of the list, students found it to be one of the most "political" verbs in the whole corpus, together with *look forward to* (rank 9th), *move forward* (rank 10th), *fire up* (rank 20th), *give up* (rank 24th), *give up on, give it up for, be sworn in* (rank 30th), and *opt out of* (rank 45th); the phrase *make sure* (ranking 1st in the phrases around the verb *make*) was also labelled the most "political" cluster, followed by *make clear*, uttered repeatedly by British and American politicians.

Below is a list of the most frequent two-word phrasal verbs in the spoken political corpus as generated by *WordSmith Tools* (Milizia 2012):

1. deal with	16. come back	31. set out
2. look at	17. keep on	32. take on
3. talk about	18. stand up	33. find out
4. think about	19. get up	34. come in
5. provide with	20. fire up	35. come out
6. set up	21. turn out	36. point out
7. go back	22. open up	37. go through
8. look for	23. come up	38. look like
9. look forward	24. give up	39. look back
10. move forward	25. get into	40. meet with
11. figure out	26. go up	41. come together
12. end up	27. call on	42. go ahead
13. go on	28. put forward	43. hand down
14. be back	29. call off	44. turn up
15. go ahead	30. swear in	45. opt out

Figure 5. Two-word verbs in ABC.

Needless to say, providing the students with long lists of phrasal verbs out of context is virtually pointless for the reasons mentioned above. The list was produced only to find out which verbs politicians are more likely to utter in their speeches and interviews; only some of these were investigated in depth.

In the paragraphs below first the most frequent one-word verbs and the most frequent as well as the most productive particles will be shown. In this chapter only two verbs will be examined in greater detail: *deal with*, because it was found to be the most recurrent in the two-, three-, and four-word clusters, and *figure out*, because, in spite of its high frequency in the spoken corpus, it seems indeed rather infrequent in general English and, correspondingly, receives scant attention in traditional reference texts. As a consequence, students are usually not familiar with it.

3. Most typical and most common phrasal verbs in spoken politics

After generating a list of two-word phrasal verbs, we became eager to check whether the verbs that were occupying a top position in the list of two-word phrasal verbs were ranking so high also in the list of single-word lexical verbs. Thus a wordlist was produced, and the most frequent verbs were unveiled manually. For the sake of convenience, Figure 6 exhibits only the first 45 verbs, where, of course, auxiliaries and modals were left out[2]:

2 The verb *have*, in fact, was ranking 1st, followed by *are*, *be*, *was*, *can* and *has*. The first lexical verb was *go*, which in ABC was first found in the form of *going*. Yet, it may be argued that this form of the verb *go* is also found in expressions such as *We are not going to turn a blind eye*, or *Our problems are going to be discussed in several meetings* where the verb *go* is not to be regarded as a lexical verb proper, but rather as an auxiliary to express future or already decided events (Murphy 2011: 40-46), or for plans and intentions, predictions and commands (Carter *et al*. 2011: 224). This manual analysis was not carried out.

1. go	16. come	31. deal
2. want	17. give	32. talk
3. make	18. put	33. build
4. know	19. look	34. meet
5. think	20. keep	35. find
6. get	21. believe	36. face
7. work	22. tell	37. ask
8. thank	23. hope	38. bring
9. like	24. create	39. try
10. need	25. set	40. stand
11. say	26. mean	41. use
12. help	27. provide	42. remember
13. take	28. move	43. lead
14. see	29. clean	44. afford
15. let	30. cut	45. ensure

Figure 6. Most frequent lexical verbs in ABC.

It was interesting to notice that the two lists (Figure 5 and Figure 6) shared only very few rankings: with the exception of the verb *think*, ranking 5th in the wordlist and 4th in the phrasal verbs lists (i.e. *think about*), the others displayed a very different order. For instance, the first verb of the wordlist, *go*, ranks only 7th in the multi-word verbs list, i.e. *go back*, and only further down in the list do we find *go on*, *go up*, *go through* and *go ahead* (Figure 5). Furthermore, it came as a big surprise that the verb *deal*, ranking down in the list in position 31, was instead the most frequent two-word phrasal verb uttered by politicians. Students were expecting to find *make* top of the list of the phrasal verbs as well, given that it ranked 3rd in the list of single-word verbs, but it was later found that the verb *make* ranks so high because it lends itself to create several phrasal units rather than phrasal verbs, *make sure* being the most recurrent in politics (Milizia 2013). It was apparent, at this point, that many words are frequent because of their strong constructional tendency: they occur in many phrases (Summers 1996: 262; Sinclair 1992: 162), and *make*, in fact, did not rank so high

carrying the meaning of "create or produce something by working" (Milizia 2013).

Once the single-word lexical verbs were generated, we started to wonder also whether the most frequent adverbial and prepositional particles in the corpus were also the most productive, namely those more willing to combine with a verb to form a phrasal verb. With this purpose in mind, two lists were produced, one displaying the most common and the other the most productive particles in ABC:

1. to	1. up
2. in	2. out
3. for	3. in
4. on	4. on
5. with	5. down
6. at	6. for
7. by	7. with
8. up	8. back
9. out	9. forward
10. over	10. into
11. back	11. through
12. down	12. like
13. forward	13. together
14. after	14. off
15. across	15. away
16. under	16. aside
17. off	17. over
18. ahead	18. ahead

Figure 7. Most frequent particles (on the left) and most productive particles (on the right) in ABC.

It was very interesting to find that the order of the particles appears somewhat different when it comes to displaying the most productive ones: *up* and *out* are by far the most productive particles combining with a large number of different verbs in ABC (these findings are also corroborated by Biber *et al.* 1999: 413). Phrasal verbs with *up* usually convey perfectivity in the sense of thoroughness and are often referred to as 'completive phrasal verbs', even though, confusing and frustrating enough for the learner, also *out*, in combination with some verbs, describes activities carried out thoroughly or completely

(*Collins Cobuild Phrasal Verbs Dictionary* 2011: 39; McCarthy/O'Dell 2009: 30-32).

It has been noted (cf. Sinclair/Moon in *Collins Cobuild Dictionary of Phrasal Verbs* 1989), in this respect, that it would be more proper if, in dictionaries, verbs were grouped by particles (*give over, get over, tide over, take over*) rather than their verbal component (*give up, give out, give over*), as is still common practice. By doing so, we would notice that the semantics of phrasal verbs is not as arbitrary as is often held to be, and this categorization might be helpful also for students who would have a less abstract and more systematic and logical picture of these problematic and opaque verbs.

Dictionaries are, in the case in point, somewhat misleading because, as is in fact shown in Figure 1a, all the phrasal verbs are grouped under the verb *turn*, even though they are not semantically related[3].

The approach of citing the phrasal verb under the preposition or the adverb would be more sensible, in that the main communicative function is carried out by the particle (Side 1990: 146), and it is the particle which indeed receives the semantic focus of the sentence (Declerck 1977: 306).

This said, it is easily understandable why phrasal verbs are a language feature that are known to cause perpetual problems to learners: the two (or three, or even four) words composing the verb create a new meaning that does not seem to be systematically related to either or both of the original words. Dictionaries usually make special provision for them, and grammars for learners seem to make apologies for their very existence (Sinclair 1991).

3 Also the visual layout is, in a way, confusing: the arrow next to the verb would seem to suggest that the preposition or the adverb co-occurring with the verb simply extends the meaning of the verb alone. Indeed, the meanings of the phrasal verbs around *turn* are more often than not far from being transparent or simply an extension of the verb.

3.1. Deal with

As mentioned earlier, the verb *deal with* was chosen as an item to investigate in some of the students' final project because it emerged as the most prominent in terms of frequency in the two-word and three-word clusters, and it ranked 3rd in the four-word clusters list (i.e. *to deal with the*, after *I look forward to* and *to figure out how*), as can be seen in Figure 8:

1. deal with	1. to deal with	1. I look forward to
2. look at	2. look forward to	2. to figure out how
3. talk about	3. look at the	3. to deal with the
4. think about	4. to move forward	4. it turns out that
5. provide with	5. to figure out	5. going to keep on
6. set up	6. it turns out	6. I was sworn in
7. go back	7. get out of	7. if you look at
8. look for	8. set up a	8. to get out of
9. look forward	9. come up with	9. put people back to
10. move forward	10. was sworn in	10. I ran for President
11. figure out	11. go ahead and	11. what's going on
12. end up	12. live up to	12. to pay for it
13. go on	13. I ran for	13. going to call on
14. be back	14. to stand up	14. to live up to
15. go ahead	15. to give up	15. you put it in

Figure 8. Multi-word verbs included in the two-, three- and four-word clusters in ABC.

Most students were familiar with *deal with* and could use it proactively, but mainly with the meaning of "be about something" (*Collins Cobuild Phrasal Verbs Dictionary* 2011: 52). What they were not aware of was the negative semantic prosody that is almost always associated with it, which emerged very clearly from the concordance lines, where *crisis* is omnipresent in the vicinity of *deal with*, together with *a H1N1 virus, a broken health care system, a child's sickness, a common problem, a common threat, a budget's deficit, criminality,*

corruption, among others, as can be observed in the concordance lines below:

N	Concordance
34	can do what we've always done — whether it was on a farm or dealing with a crisis overseas. We shape our own destiny as a
35	. So we've had a real challenge over the last couple of years of dealing with a crisis but not taking our eye off the ball in terms of
36	was not just to rescue the economy. Our job, as I said, was to deal with a decade of putting off tough decisions. I'm not just
37	that African leaderships and governments have to take. They are dealing with a different scale of problem, and therefore this is the
38	. In the long run, our economy is going to be fine. Right now we're dealing with a difficult situation and, Mr. Secretary, I want to
39	tomorrow, it's much easier for a man of peace, like Uribe, to deal with a difficult problem that he inherited. Q President Bush,
40	possible for business to locate. That's why in Britain, in spite of dealing with a difficult deficit, we are cutting our rate of
41	Turkish and the Armenian people. If they can move forward and deal with a difficult and tragic history, then I think the entire world
42	reform? Lisbon At the Lisbon summit next month, Europe will be dealing with a dilemma confronting us all: economic reform. How
43	we can do. We are dealing with a trebling of oil prices, we are dealing with a doubling of food prices and we are dealing with a
44	once in a while you'd go to a specialist, but basically you were dealing with a family doctor. Increasingly, the economics of
45	. That's just not fair. You shouldn't be punished for getting sick or dealing with a family crisis. That's why I'll require employers to
46	a state that was able to make critical investments even as it was dealing with a fiscal crisis; a state that became better positioned
47	that was able to make investments in education even as it was dealing with a fiscal crisis; a state that has now been able to
48	Minister: My hon. Friend is right to say that as well as having to deal with a fiscal crisis-which has been handed over to us by the
49	wrongs of that dispute, but more about the ways and means one deals with a given crisis, like this one for instance. And I think it
50	whole point of the G20 is that the whole world must take action to deal with a global problem, which requires global solutions. The
51	market is affected, home building has come to a halt. So we are dealing with a global financial situation. I believe that we have
52	of taxpayers. I think some of your questions forget we are dealing with a global financial turbulence that is affecting every
53	today. Why are we in London? We are in London because we are dealing with a global economic recession, and we are in London
54	had a shower. (Laughter.) So I came off the plane. I knew I was dealing with a good man when I looked in his eyes and he was

Figure 9. Concordance lines of *deal*with* in ABC.

No clear and detailed guidance is given in traditional reference texts as to the typical behavior of this verb, that is its preferred collocates, the "best friends" it likes to occur with, even though two of the three items it is associated with in the *Ragazzini* (2012) are *problems* and *complaints*, together with *enquiry*.

The *Macmillan Phrasal Verbs Plus Dictionary* (2009) provides seven meanings of *deal with*: 1) do sth; 2) control an emotional situation; 3) buy and sell something; 4) take necessary action; 5) be about sth; 6) talk about sth; 7) punish s.b. A wider context of some of the lines in Figure 9 shows that the most frequent use is no. 4, "take necessary action". A few instances are displayed below:

(1) No decisions can be quite as difficult as the decisions that African leaderships and governments have to take. They are dealing with a different scale of problem, and therefore this is the moment I think to help that African leadership by coming up to the mark with the things that we need to do. (Prime Minister Tony Blair, May 2005)

(2) I'll also stand up for paid leave. Today, 78 percent of workers covered by FMLA don't take leave because it isn't paid. That's just not fair. You shouldn't

be punished for getting sick or dealing with a family crisis. (President Barack Obama, June 2008)

(3) My hon. Friend is right to say that as well as having to deal with a fiscal crisis − which has been handed over to us by the Labour party − there is also a social crisis: social mobility has gone down; inequality has gone up; the gap between rich and poor has increased; and child poverty has increased by 100,000 since 2004 alone. (The Deputy Prime Minister, Nick Clegg, July 2010)

The overall unfavourable prosody that this verb carries within itself has already been analysed (Partington 2003: 18), and it is confirmed in ABC. From a careful look at the many instances in the spoken political corpus, some collocates carry a prima-facie good prosody, like the extract below:

(4) I will tell you that part of the warmth I feel toward Germany is because I like Chancellor Merkel a lot. I've now dealt with a lot of world leaders, and I think that Chancellor Merkel is smart, practical, and I trust her when she says something. (President Barack Obama, June 2009)

Starting from the assumption that a word is learnt through encounters with it in speech and writing, and that a word is loaded with the cumulative effects of those encounters such that it is part of our knowledge of it (Hoey 2005), we might speculate that in the expression "dealing with a lot of world leaders" such an action is represented as a difficult and complicated business.

In the investigation of this phrasal verb, Partington (2003) found instances of *deal with* occurring with positive items, such as *success*, in the collocation *deal with success*. The occurrence of *deal* with a positive concept such as that of *success* is certainly referring to all the success-related novelties and possible problems that are usually inextricably linked to success.

It very often happens that the words in the vicinity of the node carry an apparently neutral semantic aura, as is the case in the concordance lines below, where *members of Congress, a population, a world which is changing so fast* do not clearly carry a positive or negative pragmatic load:

(5) That said, we do have a population that we have to deal with, a population that has a role in our economy, a population which I think all Americans want to see treated humanely. And so the President had proposed some time ago a guest worker program, a temporary worker program, not an amnesty, because our laws have to be respected. (Secretary of State, Condoleezza Rice, April 2006)

(6) Many of them have confronted circumstances more perilous than those which face Britain today. But perhaps few will have dealt with a world which is changing so fast. From Beijing to Seoul, from Washington to San Paolo, leaders must work out what it all means for their countries and where their national interest lies. (Prime Minister David Cameron, November 2010)

Endorsing Sinclair's (2003) claim that rarely are words neutral, a careful reading of the extracts above shows that even though the words co-occurring with the node verb are neutral, the aura of difficulty is strongly perceived in the less immediate cotext, where items like *ugly process, something wrong, perilous, dark times, difficulties* corroborate the general tendency that *deal with* likes to co-occur mainly in negative situations and unpleasant circumstances.

Semantic prosody, as well as semantic preference, are extremely important, therefore students should be made aware of these features of language. The prosody is not reported in dictionaries and may not even be accessible to an individual's introspection: its trace may only really be evident by concordancing numerous examples.

Needless to say, students are usually not made aware of these restrictions, which are often swept under the carpet in teachers' and course books' explanations and definitions of phrasal verbs (Waibel 2008).

3.2. Figure out

As Figure 8 shows, the verb *figure out* ranks 11th in the two-word list, occupying an even higher position in the three- and four-word list, i.e. respectively position 5 and 2. Hardly any student was familiar with this verb, which is barely encountered in textbooks and rarely used by high school teachers. As the data in ABC shows, instead, the verb *figure out* was found to be very common in spoken political discourse, corroborating the assumption that phrasal verbs tend to be rather

colloquial or informal and more appropriate to spoken than written English.

This discrepancy provides evidence that different phraseology occurs with different frequency in different text-types, and, conversely, that different text-types have significantly different phraseology (Stubbs 2007).

The verb *figure out* is one of those phrasal verbs which rarely have an independent existence from the preposition: no instance of *figure* was found in ABC and there is brief mention of *figure on* in the *Collins Cobuild Advanced Dictionary* (2011) and in the *Collins Desktop* (2005), but not even a star was provided to indicate the frequency of the verb *figure out*. Unsurprisingly, not one occurrence of *figure out* was found in EUC, the written European Union Corpus that students of Political Science have been assembling in the last eight years and have been using as a study corpus as well as a reference corpus. In the *Macmillan Phrasal Verbs Plus Dictionary* (2009), *figure out* is marked by two stars, clearly indicating that it is not a very frequent item. The *MPVPD* provides the two following meanings: 1) to be able to understand something or solve a problem = work out; 2) to understand what someone is like and why they behave the way they do = work out, make out. Figure 10 displays some typical uses of *figure out* as uttered by British and American politicians:

N	Concordance
121	a series of meetings taking place all over the Capitol trying to figure out how to get the thing together -- that was a messy
122	for it. If you want to cut taxes, great, cut taxes. But you've got to figure out how to fill the revenue that results when you lose that
123	They're barely keeping their heads above water. They're trying to figure out how to retire. They're seeing more and more of their
124	what it was like for a worker who was struggling to get by, to figure out how to take advantage of opportunities at community
125	people in Elkhart can't afford it. The single mom who's trying to figure out how to keep her house can't afford it. And whether
126	and our long-term economic success. But we've got to figure out how to pay for that. I think it's really important to
127	particularly when unemployment is 10 percent. What's hard is to figure out how to solve these problems. Our long-term deficit, if
128	tough days over the next several days. People are still trying to figure out how to organize themselves. There's going to be fear,
129	neighbor had gotten the same car for $6,000 less, you'd want to figure out how to get that deal. And that's what reform is all
130	as a society to not leave people out, then we're going to have to figure out how to pay for it. If we don't, then we should
131	. It's a big entitlement. Medicare -- it exists. We've got to figure out how to keep it from bankrupting the country without
132	issue is something that I have dealt with my whole life, trying to figure out how to juggle work-family balance in the process of
133	years. Everything we spend on our veterans -- we would have to figure out how to come up with that much money. That's the
134	millions of people out there who are struggling, who are trying to figure out how to pay the mortgage; they're seeing their homes
135	to be made tooling around with these computers, trying to figure out how to communicate with each other more efficiently.
136	you, Mr. President. Congress, as you alluded to, is trying to figure out how to pay for all of this reform. Have you told House
137	now, whose homes are underwater right now, who are trying to figure out how to pay the bills or send their kids to college right
138	fact that we had some breakout sessions, we're going to have to figure out how to break out of these silos and integrate these
139	markets, community gardens are really critical. But we have to figure out how to make this more affordable. And Bancroft
140	willing to cooperate with you. It's just not that difficult to figure out how to deal with the disease of which thousands of
141	banks, the vast majority of banks are healthy. We've got to figure out how to raise their capital -- the point that you were

Figure 10. Concordance lines of *figure*out* in ABC.

It is evident from the concordance lines that there is a great consistency of patterning around the verb both to the left and to the right: the first position to the left of *figure out* (N-1) is occupied by the word *to* in 70% of the examples, suggesting a preponderance of infinitive usage. In position N-2, the verb *try/trying* is frequently found, together with *have to*, *hard*, *difficult*: it seems that *figure out* has a semantic feature of 'difficulty' and is found in a slightly negative prosodic environment, occurring with items like *problems*, *bankrupt*, *diseases*, *taxes* in its proximity.

Turning to position N+1, two words dominate the pattern: *how* and *ways* (which are identical in terms of semantics), followed in a much lower percentage by *why*. A close look at the concordances shows that both *how* and *ways* are most of the time followed by *to*, and the two bundles, *how to* and *ways to*, are so common at position N+1 and N+2 that they almost seem an inherent component of the phrase *figure out how to* and *figure out ways to*. Figure 11 shows three- and four-word clusters around the verb *figure out*:

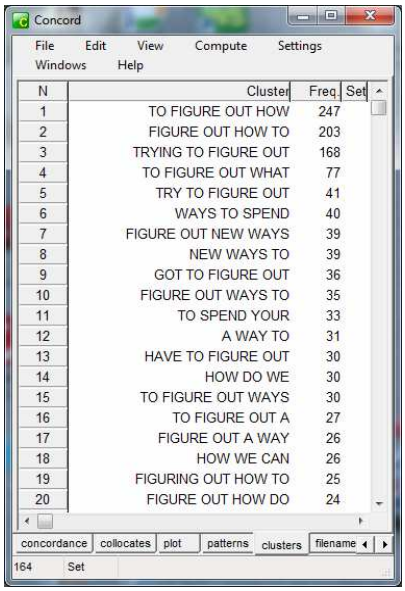


Figure 11. Clusters around *figure out* in ABC.

The table shows that *figure out a way to* is also very recurrent, even more so when an adjective premodifies the noun, as in expressions such as *figure out a better way*, *figure out a coy way*, *figure out a different way*, etc.

It was interesting to find out that, when providing a synonym for *figure out*, the *Macmillan Phrasal Verbs Plus Dictionary* (2009) gives two more phrasal verbs rather than a single word lexical verb, that is *work out* and *make out*. Yet, reviewing the evidence of usage, we noticed that while *figure out* seems to share quite a few collocates with *work out* − like the verb *try to* the left and *how to* and *a way to* to the right (no instance of *work out ways to* was found) − hardly any instances where *figure out* can be interchanged with *make out* were found.

Let us now look at some extended concordances extracted from ABC:

(8) I'm very glad to see the terminology "war on terror" has been banned because the bottom line is that the people who attacked us on 9/11 or in London or in Mumbai are murderers, and if we talk about a war on terror, we enlarge them in their own societies to be warriors. They are not warriors, they want to be known as warriors, but they are murderers, and we have to figure out a different way to deal with it. (Madeleine Albright, March 2009)

(9) Let's start thinking of each other as Americans first, figuring out how we can help one another, figuring out how we can move this country forward. I'm confident we'll do great. Thank you, everybody. God bless you. Thank you. (President Barack Obama, January 2010)

(10) And folks who are out of work right now, whose homes are underwater right now, who are trying to figure out how to pay the bills or send their kids to college right now − they don't have five or 10 years to wait. And so we've got to still work very hard in the short term to dig ourselves out of this enormous hole. (President Barack Obama, September 2010)

The several examples that were randomly chosen to illustrate the behaviour of *figure out* seemed all to belong to American speakers. With this purpose in mind, students carried out a manual analysis, confirming that all the instances of the node verb were found in speeches by American politicians. McCarthy and O'Dell (2009: 142) include *figure out* in the American list of phrasal verbs, adding that the British equivalents are *suss out* and *work out*, even though not one instance of *suss out* was found in ABC as uttered by British politicians.

In the distinction that Sinclair and Moon (2011: vi) make among the four different types of phrasal verbs, *figure out* belongs to the first category, namely those combinations "where the meaning of the whole cannot be understood by knowing the meanings of the individual verbs and particles"[4]. These are certainly the most problematic for foreign learners, even more so for those who have

4 The other types of combination are the following: combinations where the verb is always used with a particular preposition or adverb, and it is not normally found without it; combinations where the particle does not change the meaning of the verb, but it is used to suggest that the action is performed thoroughly, completely, or continuously; combinations where the verb and particle both have meanings which may be found in other combinations and uses, but there is overwhelming evidence in data that they occur together.

several years of English behind them, for whom learning was based on the memorization of a large number of small items and on a literal rote learning of long lists of items, thus on an analytic rather than a holistic learning style. What I am arguing here is that if students are made aware from the start that *figure out* is a two-word verb with its own meaning which has no semantic relationship with the individual verb nor with the preposition *out*, learners will store in their mind the two-word verb as a holistic unit, as a single "big word" (Ellis 1996: 111), just like a native speaker who has acquired the larger units as single wholes in the first place (Mauranen 2004a: 205). What is needed is certainly an enormous amount of exposure to real data, so that students can acquire not only the phrasal verbs but also the strong collocational associations between phrasal verbs and other words.

As mentioned earlier, *figure out* is not usually covered in traditional reference texts. In this respect, I agree with Mauranen (2004b: 97) that spoken language is a domain where learners need to work out many linguistic features on their own, in that they cannot expect enough help from textbooks, teachers or reference materials.

This is exactly one of the many advantages of corpora: they allow us to observe *what* is typically said in given circumstances, in political speeches in the case in question, and *how* it is typically said, and to relate the two.

4. Conclusion

After a very intensive course on phrasal verbs and phraseology in general, students said, in an anonymous form they were given at the end of the semester, that they very much enjoyed working on real and authentic language, despite a certain initial skepticism: even if modern students are today Internet savvy and *au fait* with technology in general, this change in approaching a foreign language initially scared more than fascinated them, with some of them who would have rather attended a traditional course based on books and dictionaries only, given that any kind of *change* causes anxiety to the human psyche

(Hunston 2007). Needless to say, it was up to the teacher to motivate the students and start with easy tasks, as difficult tasks would certainly turn any novice against corpora and they would very likely lose interest soon (Mauranen 2004b: 100).

After several years of teaching through corpora, both to undergraduate and graduate students, it is possible for me to claim that they respond extremely well to being offered corpus data, even though it is not an easy task to start reading "vertically" – as we do with concordance lines – after having read "horizontally" all our life, nor to free our mind of existing assumptions. Experience has shown that students like to get their hands on corpora and concordances themselves and find out about language patterning and the behavior of words in an "autonomous" way (Bernardini 2002: 165), and that they like to be confronted as directly as possible with the data, thus becoming linguistic researchers, or "language detectives", to borrow John's (1997: 101) words.

When we read concordances, we feel like we are in a new territory even if we are re-examining patterns which have been described before. Even though most students are eager to learn, it is common belief that the English taught at university must be somehow a "repetition" of what was already taught in school. Students were surprised to learn that the course was not a "repetition" but was altogether "new stuff", not only because the syllabus included the language of politics but mainly because it included attested spoken language.

It is well known, in fact, that the language that is generally taught is written language, and most traditional grammar texts, despite ideological lip-service to the priority of speech, do not describe the spoken register adequately. The differences between speech and writing come to the fore when it comes to investigating phrasal verbs, and despite the centrality of spoken language in most language teaching syllabi, pedagogic descriptions tend to be far from adequate in L2 teaching materials (Mauranen 2004b: 96). In conventional grammars phrasal verbs proper, prepositional verbs and phrasal-prepositional verbs are tucked away, well off-centre, considered as marginal phenomena, almost aberrations (Sinclair 1996).

The examples illustrated in this chapter only aimed to provide a short outlook of the complex area of phrasal verbs: many features

have been left out, as well as the many restrictions on their use which add to their complexity, constituting a further problem for learners who, as a result, tend to avoid phrasal verbs as long as they can, preferring to rely on larger, rarer and clumsier words which can make their language sound stilted and awkward (Sinclair 1991: 79). The whole trend in the historical development of English has been towards the replacement of words by phrases; yet, if learners have to choose between *carry out* and *perform* or *undertake*, or between *put up with* and *tolerate*, evidence shows they opt for the latter (Milizia 2012: 156), strongly avoiding the former.

As to the phenomenon of avoidance, it has been argued (Liao / Fukuya 2004: 211) that the tendency to avoid phrasal verbs decreases with increasing proficiency, but it is also true that structural dissimilarity is impedimental to the acquisition of phrasal verbs (Dagut / Laufer 1985): Swedish and German learners, in fact, seem to have fewer problems in actively using them than Italian and Chinese learners who do not have an equivalent to English phrasal verbs, even though Waibel (2008) in her research speaks of Italian phrasal verbs in instances like *Metti giù la valigia*, *Metti la valigia giù*. In her paper where she looks at the factors that make the learning of phrasal verbs so complex, McPartland (1989) argues that semantic complexity plays a major role and that idiomatic, non-transparent and non-compositional phrasal verbs are in general more avoided than literal ones: the more idiomatic, the less familiar students are with them. This is corroborated by Italian learners who seem to handle verbs such as *find out* or *look for*[5] fairly easily, hardly managing verbs such as *live up to* or *stand up for*.

It seems that Italian students passively know more phrasal verbs than they actively use, thus performing better in the receptive than in the productive tasks, so they would all understand *It's my pleasure to introduce the newly sworn-in Director of the Peace Corps to the*

5 The verbs *find out* [*scoprire*] and *look for* [*cercare*] are among the most widely used by Italian students, and are usually covered at the beginning of grammar books. Conversely, *live up to* [*essere all'altezza di*] and *stand up for* [*prendere le parti di, essere favorevole a*] are usually avoided by Italian learners, most likely because of their opaqueness.

American people but would certainly say *It's my pleasure to introduce the new Director of the Peace Corps to the American people*, avoiding the adjectival phrasal verb altogether.

What our students need is a huge amount of exposure to genuine data: this way they will appreciate that each phrase operates as a whole, as though a single big word, and will not expect *look, forward, to* or *turn, a, blind, eye, on* to have a meaning on their own, because the words have been stored in the mind like a holistic unit. The evidence of the data shows that the set of English phrasal verbs is constantly growing and changing, and new combinations are appearing and spreading in both general and specialized language, thus this is not a feature of language that can be overlooked.

I conclude quoting Searle (1975: 68), when he says that there is a conversational maxim that reads "Speak idiomatically unless there is some special reason not to", and agreeing with Henry and Roseberry (2001: 121) when they claim that the ability to master a specialized language, political language in the case in question, can "bring the learner more success in job hunting".

References

Aston, Guy 1997. Enriching the Language Environment: Corpora in ELT. In Wichmann, Anne / Fligelstone, Steven / McEnery, Tony / Knowles, Gerry (eds) *Teaching and Language Corpora*. London: Longman, 51-64.

Aston, Guy / Bernardini, Silvia / Stewart, Dominique 2004. Introduction: Ten Years of TaLC. In Aston, Guy / Bernardini, Silvia / Stewart, Dominique (eds) *Corpora and Language Learner*. Amsterdam: Benjamins, 1-18.

Bernardini, Silvia 2002. Exploring New Directions for Discovery Learning. In Kettermann, Bernhard / Marko, Georg (eds) *Teaching and Learning by Doing Corpus Analysis. Proceedings of the Fourth International Conference on Teaching and*

Language Corpora, Graz 19-24 July 2000. Amsterdam: Rodopi.

Bernardini, Silvia 2004. Corpora in the Classroom: An Overview and Some Reflections on Future Developments. In Sinclair, John (ed.) *How to Use Corpora in Language Teaching*. Amsterdam: Benjamins, 15-36.

Biber, Douglas / Johansson, Stig / Conrad, Susan / Finegan, Edward / Leech, Geoffrey 1999. *The Longman Grammar of Spoken and Written English*. London: Longman.

Carter, Ronald / McCarthy, Michael / Mark, Geraldine / O'Keeffe, Anne 2011. *English Grammar Today, An A-Z of Spoken and Written Grammar*. Cambridge: Cambridge University Press.

Collins Cobuild Phrasal Verbs Dictionary 2011. London: Collins.

Collins Cobuild Advanced Dictionary 2011. London: Collins.

Collins Desktop inglese. Dizionario Inglese-Italiano, Italiano-Inglese 2005. Milano: Boroli.

Dagut, Menachem / Laufer, Batia 1985. Avoidance of Phrasal Verbs – A Case for Contrastive Analysis. *Studies in Second Language Acquisition* 7, 73-80.

Declerck, Renaat 1997. Some Arguments in Favor of a Generative Semantic Analysis of Sentences with an Adverbial Particle or A Prepositional Phrase of Goal. *Orbis* 26, 297-340.

Dodd, Bill 1997. Exploiting a Corpus of Written German for Advanced Language Learning. In Wichmann, Anne / Fligelstone, Steven / McEnery, Tony / Knowles, Gerry (eds), *Teaching and Language Corpora*. London: Longman, 131-145.

Ellis, Nick 1996. Sequencing in SLA: Phonological Memory, Chunking, and Points of Order. *Studies in Second Language Acquisition* 18: 91-126.

Firth, John Rupert 1957. A Synopsis of Linguistic Theory, 1930-1955. (Reprinted in Palmer, Franck (ed) *Studies in Linguistic Analysis*, London: Longman, 1-32.)

Gavioli, Laura 2006. *Exploring corpora for ESP learning*. Amsterdam: Benjamins.

Hunston, Susan 2007. Semantic Prosody Revisited. *International Journal of Corpus Linguistics* 12/2, 249–268.

John, Tim 1997. Contexts: The Background, Development and Trialling of a Concordance-based CALL Program. In

Wichmann, Anne / Fligelstone, Steven / McEnery, Tony / Knowles, Gerry (eds) *Teaching and Language Corpora*. London: Longman, 100-115.

Henry, Alex / Roseberry, Robert 2001. Using a Small Corpus to Obtain Data for Teaching a Genre. In Ghadessy, Mohsen / Henry, Alex / Roseberry, Robert (eds) *Small Corpus Studies and ELT. Theory and Practice*. Amsterdam: Benjamins.

Kennedy, Graeme 1992. Preferred Ways of Putting Things with Implications for Language Teaching. In Svartvik, Jan (ed.) *Directions in Corpus Linguistics: Proceedings of the Nobel Symposium 82 Stockholm, 4-8 August 1991*. Berlin: Mouton, 335-373.

Liao, Yan / Fukuya, Joshinori 2004. Avoidance of Phrasal Verbs: The Case of Chinese Learners of English, *Language Learning* 54.

Macmillan Phrasal Verbs Plus Dictionary 2009. Oxford: Macmillan.

Mauranen, Anna 2004a. Speech Corpora in the Classroom. In Aston, Guy / Bernardini, Silvia / Stewart, Dominique (eds) *Corpora and Language Learner*. Amsterdam: Benjamins, 195-211.

Mauranen, Anna 2004b. Spoken Corpora for an Ordinary Learner. In Sinclair, John (ed.) *How to Use Corpora in Language Teaching*. Amsterdam: Benjamins, 89-105.

McArthur, Tom / Atkins, Beryl 1974. *Dictionary of English Phrasal Verbs and their Idioms*. London: Collins.

McCarthy, Michael / O'Dell, Felicity 2009. *English Phrasal Verbs in Use*. Cambridge: Cambridge University Press.

McPartland, Pamela 1989. The Acquisition of Phrasal Verbs by Non-Native Speakers of English. *CUNYForum* 14, 150-156.

Milizia, Denise 2012. *Phraseology in Political Discourse. A Corpus-Linguistics Approach in the Classroom*. Milano: LED.

Milizia, Denise 2013. Spoken Political Corpora in the Classroom. Proceedings of the XXV AIA Conference: *Regenerating Community, Territory, Voices. Memory and Vision*, University of L'Aquila, September 15-17.

Murphy, Raymond [3]2011. *English Grammar in Use. A Self-Study Reference and Practice Book for Intermediate Students of English.* Cambridge: Cambridge University Press.

Partington, Alan 2003. *The Linguistics of Political Argument*. London: Routledge.

Quirk, Randolph / Greenbaum, Sidney / Leech, Geoffrey / Svartvik, Jan 1985. *A Comprehensive Grammar of the English language*. London: Longman.

Ragazzini 2012. Dizionario Inglese-Italiano, Italian-English dictionary DVD, Zanichelli.

Reppen, Randy 2010. *Using Corpora in the Language Classroom*. Cambridge: Cambridge University Press.

Römer, Ute 2008. Corpora and Language Teaching. In Lüdeling, Anke / Kytö, Merja (eds) *Corpus Linguistics. An International handbook*, Volume 1. Berlin: Mouton.

Side, Richard 1990. Phrasal Verbs: Sorting them out. *ELT Journal* 44/2, 144-152.

Sinclair, John 1991. *Corpus, Concordance, Collocations*. Oxford: Oxford University Press.

Sinclair, John 1992. Trust the Text. In Davis, Martin / Ravelli, Louise (eds) *Advances in Systemic Linguistics*. London: Pinter, 5-19.

Sinclair, John 1996. The Search for Units of Meaning. *TEXTUS* 9/1, 75-106.

Sinclair, John 2003. *Reading Concordances*. London: Longman.

Sinclair, John / Moon, Rosamund 2011. Introduction. *Collins Cobuild Phrasal Verbs Dictionary*. London: Collins.

Scott, Mike 2007. *WordSmith Tools* 5.0. Lexical Analysis Software Limited.

Searle, John 1975. Indirect Speech Acts. In Cole, Peter / Morgan, Jerry (eds) *Speech Acts, Syntax and Semantics* 3. New York: Academic Press, 59-82.

Stubbs, Michael 2007. An Example of Frequent English Phraseology: Distributions, Structures and Functions. In Facchinetti, Roberta (ed.) *Corpus Linguistics 25 Years on*. Amsterdam / New York: Rodopi, 89-105.

Summers, Della 1996. Computer Lexicography: The Importance of Representativeness in Relation to Frequency. In Thomas, Jenny / Short, Michael (eds) *Using Corpora for Language Research*. London: Longman, 260-266.

Tribble, Christopher 2012. Paper presented at the 10th Teaching and Language Corpora Conference (TaLC), July 11-14, 2012, Warsaw, Poland.

Waibel, Birgit 2008. *Phrasal Verbs. German and Italian learners of English compared*. Saarbrücken: Vdm Verlag Dr Muller Aktiengesellschaft & Co. Kg.

CARMEN ARGONDIZZO / ASSUNTA CARUSO / IDA RUFFOLO[1]

6. The Use of Specialised Corpora: From Research to Pedagogy

1. Introduction

Corpora are invaluable resources "in order to uncover the linguistic patterns which can enable us to make sense of the ways that language is used in the construction of discourses" (Baker 2006). The linguistic identification of a given type of discourse aims to offer insights into teaching a particular genre in the field of English for Specific Purposes. In this specific context, our research interests have focused on the analysis of specialised corpora involving travel promotion texts, on the one hand, and political speeches, on the other. The first aspect of the research project aimed at analyzing travel promotion texts (TPT) using software tools to reveal the meaning of nature as it is represented in these texts and the function that it serves within the texts. The TPT corpus was created to investigate the linguistic devices used by the producers of the texts to present the destination as both producers and readers want it to be. The second aspect of the research focused on political speeches concerning the Middle East peace process. A specialized corpus (Roadmap corpus) was created in order to analyse and compare the metaphors used by the politicians involved in the peace process. The investigation aimed at showing how the same peace process may be differently conceptualised at different points in time and by different parties on account of different socio-

1 Although the authors have contributed to the present chapter as a team, they have individually devoted more attention to the following sections: Carmen Argondizzo 1, 3, 6; Assunta (Susie) Caruso 4.2, 4.2.1, 5.2, 5.2.1; Ida Ruffolo 2, 4, 4.1, 4.1.1, 5, 5.1.1, 5.1.2.

political factors or ideologies. Several pedagogical applications arose out of these two studies for tertiary students of English as a Foreign Language, majoring in Tourism, Communication, Political Science, Economic and Social Sciences. This chapter briefly reports on research findings and offers extensive reflections on the use of the above mentioned corpora in university classroom settings where the aim was to raise students' awareness on possible mismatches between saying and meaning (TPT corpus) and on the variables adopted by politicians of different ideologies and cultures when using metaphors (Roadmap corpus). The overall aim was to allow students to bring a critical mind to readings, while aiding both comprehension and production skills and provide them with strategies, supplemented by concordancing and corpus linguistic techniques, which enable them to explain how language is actually used. If successfully applied, this didactic approach should encourage researchers to keep on identifying pedagogical techniques which can transform pure technological linguistic tools into meaningful teaching and learning strategies.

This chapter is organized in three interrelated sections. Firstly, we focus on the use of two specialised corpora in two English language courses[2] held at the University of Calabria (Italy). Secondly, we illustrate the course objectives, the corpora used in class, students' analyses and interpretation. Finally, we express our conclusive remarks.

2. Theoretical background: Use of specialised corpora in ESP settings

The role of language corpora and computer tools for language education is a well-established issue. Indeed, over the past three decades, there has been an increasing interest into how corpora and

2 Classes were organised at the Faculties of Economics and Political Science in the academic year 2011-2012.

corpus evidence can be applied to the teaching and learning of languages. This is exemplified by the flourishing literature on the issue (e.g. Wichmann *et al.* 1997; Partington 1998; Burnard/McEnery 2000; Aston 2001; Hunston 2002; Aston *et al.* 2004; Sinclair 2004; Römer 2005; Gavioli 2005; Scott/Tribble 2006). However, there are some criticisms regarding the pedagogical applications of corpora. Widdowson (2000), for instance, highlights that data retrieved from a corpus can be considered partially real since corpora provide decontexualized language. Therefore, it must be recontextualized in the classroom. Swales (2002) is skeptical about the fact that corpora require a bottom-up approach, while Cook (1998) claims that using frequency lists in a classroom may impoverish language teaching because these lists focus only on what is frequent rather than pointing out rarer yet more effective expressions. Yet, some scholars (Gavioli 2005; Römer 2004, 2006) discuss the potential use of corpus data as course materials and highlight how corpus-derived exercises have not fully enriched the teaching and learning environment and that they need to be empowered within a classroom. Indeed, there is "very little request from teachers and learners, who either ignore the possibility of using corpora inside classroom activities or do not see their relevance for teaching/learning" (Gavioli 2005:1). Moreover, Römer (2006) claims that general and specialized corpora offer valuable resources for both teachers and learners, although they have not been fully applied.

It is interesting to highlight that "there has been much interest in the creation and exploitation of specialised corpora in academic and professional settings" (McEnery *et al*. 2006: 60). In fact, in recent years, studies into professional discourse have been fruitfully informed by corpora, providing a degree of objectivity and representativeness (McEnery *et al*. 2006). Indeed, corpus-based analysis is widely used to examine spoken and written discourse in professional contexts. As suggested by Gavioli (2005), there are essentially two main reasons why EFL/ESP teachers should use corpus tools in the classroom: a) corpus work provides students with a useful source of information about ESP language aspects, and b) the process of 'search-and-discovery' implied in the method of corpus analysis may facilitate language learning and, therefore, promote autonomy in learning language use. Moreover, using corpus tools in the classroom puts the students at the centre of the teaching-learning process (Johns 2002). These are tools that make the student become a

linguistic researcher who plays an active role in discovering meaningful patterns of selected lexical items, related to the learner's field of study.

Implementing specialised corpora in ESP classrooms may be of great value to ESP teachers and learners thanks to the methodological advantages that their specificity can lead to; indeed these types of corpora are "(1) carefully targeted (2) specialized structures are likely to occur with more regular patterning and distribution (3) the pedagogical goals in terms of how they are used and applied are likely to be easier to define and delimit" (O' Keeffe *et al.* 2007: 198).

3. Teaching and learning context

The study was carried out within an ESP course for students majoring in Economic and Social Sciences and Political Science. Moreover, one third of the students in the Economic and Social Sciences course were Tourism majors who had chosen the course as an elective. The students, all native speakers of Italian, were enrolled in their third year and had already taken two modules (i.e. 30 hours per module) of English language in the faculty.

The courses, taught by the authors of the present chapter, were an integration of English for Specific Purposes (ESP), with emphasis on the language used in politics, advertising and publicity, marketing and tourism with a focus on corpus linguistics and discourse analysis. As for the course organization, the teachers met each group twice a week for five weeks resulting in 30 hours of course work. The first sessions involved an introduction to corpus linguistics and concordancing software which focused on illustrating concordance lines and collocates, while giving examples on how these tools and devices could be used to analyse the language. Hands-on sessions followed the theoretical introduction in order to give students the opportunity to familiarize with the software and its various utilities (Johns 2002). Subsequently, an in-depth investigation of the two corpora was carried out, moving into the application of the language strategies and corpus tools presented. The following section will illustrate the main features of the two corpora used in class.

4. The corpora

The two corpora chosen for the study are the TPT Corpus and the Roadmap Corpus which respectively consist in travel promotion texts and political speeches. These two specific corpora were chosen for a double purpose: to raise students' awareness of the linguistic devices in both tourism and political discourse, and to encourage students to adopt a critical approach when analysing a text. Moreover, both corpora were investigated with students enrolled in different degree courses in order to compare the results and feedback from students with diverse study backgrounds.

4.1. TPT Corpus

The *TPT Corpus* was compiled with the aim of reflecting on the relationship between three areas of research − the natural environment, tourism and discourse − and how this relationship is affected by and affects society as a whole. The underlying idea of the research carried out using the TPT Corpus was that social and economic actors, i.e. advertisers and tourism entrepreneurs, are *exploiting* the idea of nature by constructing it according to the contemporary ideology and culture.

Moreover, the choice to focus on the concepts of ecotourism as a form of responsible and sustainable tourism was also drawn on a more personal interest. This interest derives from the discussions and debates on the definition of ecotourism and eco-friendly destinations aroused in class with students majoring in Tourism[3]. Although students are aware of the principles that ecotourism sites should follow in order to achieve certain goals and benefits, there is an aspect on which they fail to respond: how can a tourist (who is not an expert in this field of research) understand whether the destination chosen is truly eco-friendly? Is it a matter of trusting the travel agent? Or do tourists have to be informed on specific guidelines? Based on these premises, the discourse used in travel advertising was analysed with

3 The students were 1st year students enrolled in the second level degree course in Tourism at the University of Calabria (Italy).

the aim of understanding whether advertisers, travel agencies, local communities are promoting real eco-friendly destinations or if they are simply following the trend of the cultural and social stances regarding the environment.

4.1.1. TPT corpus data

The *TPT Corpus* includes one main genre type, namely articles in specialized magazines collected over seven years, precisely from January 2003 to March 2010. The articles included in the corpus were taken from *Travel Weekly (TW),* a British periodical, and *Travel Agent (TA),* an American journal.

The corpus was collected with the aim of investigating how travel promotion texts use the terms *nature* and *natural,* specifically to explore whether these terms are used in tourism advertising with a deceptive meaning. In particular, the study intends to investigate how advertisers describe nature and how the search terms *nature* and *natural* are employed within these texts in order to attract potential 'green tourists'.

In order to build the *TPT Corpus,* the articles were accessed from the journals' websites and retrieved through the online library catalogue. The accessibility and the opportunities provided by this database enabled the creation of a corpus containing all the articles from 1st January 2003 to 31st March 2010 which included the words *nature* and *natural* in the headline and/or lead and/or in the body of the text. Once downloaded as PDF files, the articles were saved in .txt format in order to be processed by the software *WordSmith Tools 5* (Scott 2008). The full size of the corpus consists of 311,520 running words. Table 1 shows information regarding the *TPT Corpus.*

	Tokens	Corpus files	Concordance hits for *nature*	Concordance hits for *natural*
TPT Corpus	311,520	466	419	606

Table 1. Corpus Data.

As illustrated in the table, two separate queries were carried out within the corpus, one for *nature* and the other for *natural*; results showed 419 occurrences of *nature* and 606 of *natural*. The analysis of the terms was carried out by focusing on collocation, i.e. "the statistical

tendency of words to co-occur" (Hunston 2002: 12). The collocates of the two node words, *nature* and *natural*, were calculated with a 3:3 span, which is the calculation that "counts the instances of all words occurring within a particular span" (Hunston 2002: 69), for example, three words to the left of the search term and three words to the right. The decision fell upon this particular span since the collocates obtained seem to be the true collocates of the two search terms (Baker 2006). Specifically, this span seems to include words which are included in the noun phrases containing the words *nature* and *natural*. Not all lexical items were taken into account for the research: in this study lexical items specifically refer to nouns and qualifying adjectives "which seem to be the most suitable categories to provide a thorough description of the natural environment" (Ruffolo 2011: 44). In order to fully analyze the significance of collocates and, therefore, reveal interesting patterns, a statistical approach was adopted to identify the lexical attraction between the two analyzed words (Ruffolo 2011); the formula adopted for this study was T-score[4].

4.2. Roadmap Corpus

Since Lakoff's influential paper on the metaphors used to justify the war in the Persian Gulf (1991) there have been a number of studies of the metaphors used to talk about war or terrorism (e.g. Voss *et al*. 1992; Pancake 1993; Rohrer 1995; Sandikcioglu 2000; Lakoff 2001; Hellìn Garcia 2008; Steuter/Willis 2008), but very little has been published on the metaphors used to describe peace. As Bridgeman (2000: 2) argues, despite the value which is given to peace in political rhetoric, "it has not proven to be a particularly well-studied subject among researchers". Therefore, the Roadmap corpus was compiled to analyse the use of metaphors for peace in American, Israeli and Palestinian political discourse on the Middle East peace process. The analysis has been focussed on this particular peace process since a metaphor itself has been used to name it. As Semino (2008: 117) argues:

---

4 The cut-off point for the analysis of the collocates follows Hunston's (2002) suggestion that "a t-score of 2 or higher is normally taken to be significant" (Hunston 2002: 72).

> First, particular metaphors can be used both to 'name' and 'frame' particular initiatives, with which they become inextricably associated. Second, once a particular metaphor occupies a prominent position in the public domain, it can be alluded to and exploited in different ways by different participants in political debates.

Research shows that the use of metaphor in framing political discussion has become a major linguistic device, and a tool used by politicians to frame political arguments (Hellìn Garcia 2008).

4.2.1. Corpus data

The corpus under investigation here is a specialized corpus of political speeches dealing with the Middle East Roadmap peace process. In order to construct this corpus, speeches and statements given by American President George W. Bush, Israeli Prime Minister Ariel Sharon and Palestinian Authority Prime Minister Mahmoud Abbas during the first four years of the Middle East Roadmap peace process (2002-2005) were collected. This includes a total of approximately 30,000 words comprising 20 speeches related to peace/the peace process. A detailed description of the corpus as a whole can be found in Table 2, which shows the date range, the number of speeches, and the number of words per politician.

Political speeches	Date range	Number of speeches/ % of corpus	Number of words/ % of total words
George W. Bush	2002-2005	7 (35%)	9,669 (32%)
Ariel Sharon	2002-2005	8 (40%)	9,873 (33%)
Mahmoud Abbas	2003-2005	5 (25%)	10,535 (35%)
Total		**20 (100%)**	**30,077 (100%)**

Table 2. Data on the corpus.

The corpus has been divided into 3 sub-corpora, so as to separate the speeches of each politician, with the speeches chronologically ordered by date.

The following sections will focus on the students' experience working with the corpora in the classroom.

5. Taking the corpora into the classroom

The investigation of the ways students analyse and interpret the corpora is the core of this chapter. The following sections will describe the methodological approach employed in class and the reflections provided by students throughout the course.

5.1. TPT corpus: Methodology employed in class

The initial part of the course was devoted to a brief introduction to the text genre which students would be analysing, that is travel promotion texts. Students were given selected extracts from the corpus and then asked to analyse the collocates of *nature* and *natural*[5]. A second phase involved a more qualitative investigation with the examination of the selected concordance lines and the surrounding co-text.

As for the analysis of the collocates, students were invited to focus on the most frequent ones. They immediately noticed the predominance of nouns compared to adjectives, an evidence that was quite unexpected for them due to the type of texts they were analysing (promotional and descriptive). In particular, the collocates that drew their attention the most were *reserve/reserves*, *trails*, *park* and *habitat*, all closely related to the idea of preserving and protecting the environment, while they focused very little attention on the adjectives identified.

Moving further into the investigation, they were encouraged to express their personal interpretation of the collocates and explain the reason why the text producers had chosen to employ more nouns than adjectives. After looking more carefully at the concordance lines (see Tables 3 and 4 for samples of concordances given in class) containing some of the most recurring collocates and guided by the instructors in the analysis, students were able to investigate the texts in-depth.

---

5 The list of collocates were given to students; due to limit of time we were not able to have many hands-on sessions.

Concordance		
as home to the 203-acre Booby Pond	**Nature**	reserve. Encourage clients to visit all
bicycles, tennis and a tour of the Valriche	**nature**	reserve. Sample price: Thomson
center and networks of publicly accessible	**nature**	trails and interpretive stations that will
diving, walking, visits to Owen Island,	**nature**	trails to explore, biking. Attractions
the Zambezi River. Early risers will enjoy a	**nature**	walk as the islands are home to

Table 3. Sample of concordances of *nature*.

Concordance		
Falls Entices Three countries share breath-taking	**natural**	attractions. For years, the world-class
Vegas. While more of an engineering feat than a	**natural**	attraction, a visit to the dam - one of
Thailand does, and the superlatives don't stop at	**natural**	attractions. The destination has some
synonymous with the country's exotic appeal and	**natural**	beauty. The hurricane season, which
initiatives. Costa Rica's primary appeal is its	**natural**	beauty and attractions, although this

Table 4. Sample of concordances of *natural*.

First, they pointed out that the text writers were using nouns to present what nature offers rather than describing it with attractive adjectives. Moreover, the specific nouns used, such as *habitat*, *reserve*, *trail* were chosen to attract those tourists who want to be 'environmentally responsible'.

5.1.1. Moving to the qualitative analysis

The second step of the analysis regarded the analysis of the concordance lines along with the surrounded co-text. These were used to identify nature by classifying the collocates according to four different nature categories[6]:

6 The classification follows Thelander (2002) and Hansen (2002).

- *Artificial nature*: Few natural elements are present in the descriptions of these sites but they are mainly controlled and arranged by men. The image of human mastery is clearly depicted, while the role of nature is limited.
- *Tamed nature*: The natural elements are always presented in the texts with continuous reference to human artefacts. Moreover, nature is considered as a recreational resource, a playground for tourists. There is little attention for nature.
- *Untamed nature*: Human artefacts are present but they do not dominate the environment. Nature is to be observed, it is a spectacle thanks to which people experience tranquillity, and a refuge/escape from the alienating effects of city life.
- *Accessible wild nature*: there is little or no human impact. Nature is exotic and pure. It recalls the idea of untouched nature that should only be observed.

The following examples illustrate students' classification which shows how their interpretation differed to a certain extent from the researchers' assumptions[7].

The extracts[8] below, both classified by the researcher as *artificial nature*, were given to the students who were asked to highlight the words or expressions that would guide them in the understanding of the depiction of nature. In both cases, the students fully agreed with the original classification. Indeed, the expressions pointed out by the students highlight how nature is presented as a contour of human activities since people's happiness does not come from nature but from the act of being involved in activities with others. The selected items highlight those elements which suggest how the natural world is anything but natural.

(1) The administrative capital of South Africa, Pretoria, is a *city of government buildings and embassies.* […] City highlights National Botanical Gardens: a

7 The researchers mentioned above are the authors of the present chapter. The categories used in class for the analysis were identified by the authors, however the students were not given a model classification against which their answers would be measured. The students' answers were compared with the authors' for the sole purpose of this chapter.

8 Due to space limit, only some examples have been included in the present chapter. For a wider range of examples see Ruffolo (2011).

mass of subtropical and temperate plants with ***paved*** nature trails and a ***tea garden***. (TW36_06.txt)

(2) St Lucia to play on its **natural beauty**. [...]
The initiative will cover 74 *properties* divided into five categories: *guest houses, self-catering apartments, limited service hotels, full service hotels and villas*. The draft plan includes a website, brochure, CD-ROM, fliers targeted at niche market segments, point-of-sale material for agents, advertising and *fam trips*. (TW253_08.txt)

Interestingly enough, there were instances in which the students' analysis and classification differed from the researchers' interpretation. The following extracts were originally classified by the researchers as *accessible wild nature*, since they express the need and the willingness to protect and preserve the natural world in its authentic state. The students, on the contrary, classified them as *artificial nature*, focusing on human mastery and dominance rather than on the idea of preservation. Indeed, the underlined words in the following extracts were highlighted by the students in order to emphasize the idea of human presence in the natural world. Although students' categorization was well supported by the text, they were lacking specific notions on the protection and preservation of natural areas. As a consequence, their interpretation was partial.

(3) The GB penchant for a tipple had travelled as far as a *remote* nature reserve in Jordan. Relatively new to the concept of ecotourism, Dana was the Jordanian Royal Society for the *Conservation* of Nature's first *ecotourism* ***project***. It ***became*** a *protected* reserve in the '90s, thanks to financial support from the UN. (TW91_03.txt)

(4) The Algarve hills invite us to take *long walks*. [A] *great diversity of flora and fauna that can be observed* at locations such as the Ria Formosa Natural Park […] and Vila Real de Santo Antonio Nature Reserve. ***Come and recharge your batteries*** in the peaceful countryside of the Algarve. (TA56_06.txt)

(5) Shaumari Nature Reserve: Shaumari was ***created*** as a breeding centre for *endangered wildlife*. Following breeding programmes with some of the world's leading *wildlife* parks and zoos, the reserve is now a thriving *protected* environment for some of the Middle East's *rarest* species. (TW91_03.txt)

The second step of the qualitative analysis aimed at understanding how these travel promotion texts elaborate the idea of nature and what functions this idea serves within these texts. The functions used for the classification were the following[9]:

- Spiritual function, which recalls a type of nature that can provide us with a kind of spiritual therapy helping us to ground and rebalance our lives;
- Nourishing/nurturing function, which refers to tourists who experience nature to renew themselves in order to face the alienating effects of city life[10];
- Cultural function, which regards historical-cultural information as well as non-environmental (biological) information;
- Global Economy function, concerned with economic profits that the tourism companies gain from advertising the naturalness of a location;
- Local Economy function, which regards the economic benefits of local people and communities in the host areas;
- Promotional function, which shows how the concept of nature is simply used as a contour, since the natural world seems to be constructed according to society expectations;
- Educational function, which refers to all those initiatives promoting environmental ethic;
- Recreational function, which highlights how the natural aspect of a holiday may hide the pure entertaining feature of the tourist location;
- Preservative function, which points to the relationship between humans and nature as a sort of 'guardianship' of nature;
- Aesthetic function, which consists in the constructing of the physical environment as a 'landscape' not for production yet embellished for aesthetic appropriation.

---

9 The classification follows Hansen and Machin (2008).

10 Although very similar, there is a substantial difference between the two: the spiritual function refers to an experience of wilderness which leads to the pleasure and serenity that a place can evoke, providing tourists with long-term effects, while the nourishing/nurturing function has more short-term benefits, referring to a type of nature which may offer a sense of tranquillity through the simple activities provided by the natural world (Ruffolo 2011).

The following extract was classified by students as examples of preservative and recreational functions. Generally speaking, these texts were written with the aim of attracting tourists with a low level of environmental commitment (Holden 2008). As we can see in the examples below, nature is mentioned but it is not represented as something requiring special attention or consideration on behalf of the tourist, as in other instances already highlighted in Holden's work (2008).

(6) The island is criss-crossed by trails, all detailed in a Cyprus Nature Trails map from the CTO, outlining 48 *walks* taking in the Island's beauty spots. The beauty of Cyprus is its compact size − in winter, many operators offer *free car hire*, so there's no excuse for not getting out and about. (TW45_05.txt)

The following extract is an example of how locations are promoted for eco-aware tourists who are tourists interested in nature and foreign cultures and/or in conservation programmes. This text, as well, was categorized as preservative and recreational by the students:

(7) The Free State's natural open spaces have created an extensive system of parks and reserves. In addition to the nearly 30,000-acre Golden Gate Highlands National Park, there are more than 80 provincial, municipal, and privately owned nature parks, nature reserves, *game reserves*, and *game farms*. Within these parks activities include *game viewing, birdwatching, hiking, horseback riding and 4X4 off-road dri*ving. Qwaqwa National Park borders Golden Gate. (TA66_06.txt)

While students reached an agreement on the previous examples, the extracts below have triggered interesting debates in class since students had various ideas and differing interpretations.

Extract (8) has been classified as promotional by those students who focused on human activities with no references to the natural world, while other students considered the same activities as references to economic efforts for local sustainment. They classified them as examples of the local economy function which regards the economic benefits of local people and communities in the host areas. Remarkably enough, the debate focused on how the economic aspects seem to override the environmentalist ones, although in an attempt to enhance local income.

(8) Visit the Domaine Les Pailles, a 1,200-hectare nature park with lots of facilities including a *Chinese restaurant*, an *Indian restaurant*, *pizzeria* and a *casino*. […] Suitcase essential: *swim suit* and *sun cream*. (TW27_06)

Another instance, which aroused an interesting and lively debate, was extract (9): the preservative function versus the recreational one. Attracting international tourists, who are pure vacationers, may provide a positive economic contribution to environmental protection, but may also cause damage to the protected areas (e.g. over-popularity, displacement of indigenous peoples) (Ruffolo 2011).

(9) Portugal's diverse geography offers countless options for nature-lovers. The terrain ranges from rugged mountains and verdant forests to grass-covered plains and pristine, sandy beaches. Portugal's *commitment to maintaining ecological balance* has led the country to *preserve* its natural beauty by *designating a number of national and natural parks*. (TA233_03.txt)

5.1.2. Students' reflections

The students were able to identify a mismatching between saying and meaning. They suggested that, although the authors focus on environmental conservation, there is a major concern for the economic issues and benefits not considered in a sustainable way. Moreover, the students emphasized that the media, in this case travel promotion texts, do not always provide a complete picture on *ecotourism sites*, since essential information for the evaluation of the locations is sometimes missing. Yet, they often point out how a linguistic analysis can help to scratch the surface since the thorough reading and analysis of texts can aid in noticing possible mismatching and inconsistencies and in developing critical thinking.

5.2. The Roadmap corpus: methodology employed in class

As for the Roadmap corpus which was aimed at analysing metaphors, the first step was to give students a selection of concordance lines (76) of the target domain *peace*, taken from the three sub-corpora. The target domain is the concept that becomes understood (the more abstract or unfamiliar concept). It is then related to the source domain which is the other concept which somehow facilitates understanding or discussion of the target. We explained to the students that they

would be carrying out a linguistic analysis as well as a conceptual metaphor analysis.

mourn the damage done to the hope of	**Peace**	the hope of Israel's and the Israeli
living side by side in	**Peace**	and security
opposed the	**Peace**	process and seek the destruction
as progress is made toward	**Peace**	settlement activity in the
implementing our road map toward	**peace.**	Reaching that destination
it is possible to bring	**Peace**	to the Middle East
be true contributions to building	**Peace**	in the region.
It accelerates the	**Peace**	process.
not – and never will be a partner to	**peace.**	He does not want peace.
	Peace	and terror cannot coexist.
joined with Arafat and sabotaged the	**Peace**	process with a series of the most
are the primary obstacle to any	**Peace**	process.
nflict and destroying and chance for	**peace.**	This is another attempt to
towards an Israeli partner for making	**peace.**	But partnership is not

Table 5. Sample of concordances for *peace*.

The MIPVU procedure (Steen *et al*. 2010) was selected for the first part of the analysis, i.e. metaphor identification, since it is made up of a precise set of guidelines which allows for the systematic identification of linguistic metaphors. After reading the text for general understanding, the first step is to identify the lexical units. Then, the main line of research is to find the contextual meaning of the lexical unit. The contextual meaning is then compared to a more basic meaning of the lexical unit. This is done with the help of dictionaries. MIPVU generally works with the Macmillan English Dictionary when making decisions about contextual meanings and basic meanings since this dictionary is contemporary and corpus-based. After this, it should be examined if the two meanings can be contrasted and compared. When deciding about sufficient distinctness, the MIPVU guidelines state that when a lexical unit has more than one separate numbered sense description within its grammatical category, these senses are regarded as sufficiently distinct; when a lexical unit has only one numbered sense description within its grammatical category, this counts as the basic sense and any difference with the contextual sense of the item under investigation will count as sufficient distinctness. When this is the case, then the lexical unit can

be marked as a metaphor related word (MRW). Our research involved analysing every single lexical unit in the political speeches in order to decide whether it could be coded as a metaphor-related word or not. MRWs are seen as potential expressions of underlying metaphorical mappings, those which indicate the source domain of a metaphor. It would have been extremely time consuming and not relevant for this project to do the same with the students. Therefore, in class the students were asked to work only with the words highlighted by the instructors. Let us take this sentence from one of Ariel Sharon's speeches as an example:

(10) It is true that this is not a shining *path* which will *lead* us to instant, magical solutions, but I am certain that only by *going forward* in this *direction*, *step* by *step*, will we be able to achieve security for the Israeli people, and *reach* the peace we all yearn for (Sharon_Herzliya_12.02)

The students analysed all the words in italics, deciding together both on the contextual meanings and the basic meanings. In this sentence, for instance, the contextual meaning of *path* is 'the way that someone takes to achieve something', and the basic, more concrete meaning is 'a way from one place to another that people can walk along'.

All the words highlighted by the instructors are MRWs. Our objective here was for the students to understand which source domain the metaphor keyword could belong to, rather than decide on metaphoricity or not.

Once all the metaphor keywords had been analysed, the metaphorical expressions were grouped together according to source domain. The students decided together on the name for each category. The identified source domains are the following: MOVEMENT/ JOURNEY, PERSONIFICATION, BUSINESS, BUILDING, DESTRUCTION, CONFLICT, LIGHT, PLANT, GAME, RELIGION, and DEATH.

The next step was to come up with a mapping structure between the source and the target domain. Here we have the correspondences between the target domain PEACE and the source domain MOVEMENT/ JOURNEY. Based on their analysis of the concordance lines, students saw the following correspondences:

- The peace process is a path / road / track
- Implementation of the peace process is movement forward / direction

- Rejection of the peace process is lack of movement / opposite direction
- Participants in the peace process are travellers
- Peace is a destination

These tasks were carried out to teach the students the steps involved in identifying metaphor keywords and source domains. Yet, the students did not analyse all the speeches in detail.

In the final sessions, we asked the students to carry out the same steps we had previously followed when using the sample concordance lines. The students analysed one speech per politician. Here we will be looking at an example of just one speech. Together with the students we chose a statement given by Prime Minister Sharon at a meeting with Palestinian Prime Minister Abbas in Jerusalem on July 1, 2003. This speech was selected because we calculated that the target word *peace* was used approximately once every 44 words, making this speech one of his most important 'peace speeches'. Once again, the students concordanced the target domain *peace/process*. They carried out a MIPVU analysis and identified the following keywords: *promote, lead, seek, extend (your) hand, pursue, price, compromises, collapse, derail, move, forward, step, take, advance, live, side by side, neighbour*.

Working in groups, students tried to explain the reason why certain keywords were used within the speech as a whole. A non-critical reading of metaphors, as in the example 'Israel extends a hand in peace', exposes its positive side: good will, courtesy, reconciliation and a sincere readiness to make peace (Gavriely-Nuri 2010). A more critical reading however, reveals the opposite. Indeed, the students reported that it actually shows Sharon's intent to highlight 'us' vs. 'them', the Israelis as the good neighbours extending their hand, and the Palestinians as the opponents who refuse to take it. Sharon depicts the Palestinians as the evil opponents who are trying to make the peace process collapse and trying to derail any possibility of peace. Sharon, on the other hand, is the good partner who is willing to pay a painful price and make painful compromises for peace. Near the end of the speech, he does however, bring the two parties together showing them taking bilateral steps towards the destination peace.

5.2.1. Students' reflections

The students' overall interpretation of the speech underlined that, within a peace framework, Sharon was presumably legitimating his policy towards terror and actually participating in a Just War discourse. The metaphors in Sharon's peace discourse are used to construct a positive self-image of the nation as well as of himself by demonstrating sensitivity to the positive social value, yearning for peace. Metaphors like 'we extend our hand in peace' emphasize the perceived asymmetry between the opponents. Indeed, the identification of underlying conceptual metaphors and the way they are represented in a text linguistically can provide the students with an organizational framework and shed light on the choice of lexical items that at first glance may seem almost random. An awareness of the use of metaphor and the manner in which it may be differently employed by politicians of different ideologies and/or cultures encourage deeper insights into a text, allowing students to bring a critical mind to readings and aiding both comprehension and, eventually, production skills.

6. Conclusions

This chapter has outlined how the use of specialised corpora in professional and academic settings is highly beneficial for both teachers and students. Corpus data can provide academic classes with invaluable information regarding the language and discourse employed in specific areas, the description of the text and the general socio-cultural contexts in which they are produced. The chapter described some activities that could help students approach corpus tools and strategies with the aim of investigating specialized discourse. The core aspect of these tasks is to raise students' awareness of this type of analysis and of its importance for their own language learning in LSP settings (Gavioli 2001). This project, carried out at the University of Calabria in the degree courses of Political Science and Economic and Social Sciences, has indeed highlighted how learners can become aware that corpora can be meaningful resources to understand underlying discourse which enhances critical

reading. Although the degree of motivation and the attitude varied when the students from the two degree courses approached the text for the first time, no significant difference was found in comparing the feedback obtained at the end of the course from students with a different academic field background. Students were somehow, and maybe appropriately, reluctant when presented with a topic less meaningful for their study interest. Yet, when exposed to corpora related to topics linked with their study field, the effort put forward and the interest shown in carrying out the activities were the same among all students. This shows how relevant topic selection is in the didactic process. Moreover, the students involved in this experience highly appreciated the final course assignment which required them to accomplish the following interrelated tasks: performing a linguistic analysis using the concordancing software and strategies presented in class, compile a mini-corpus, and present the mini-corpus they had created at the oral exam. Thus, their overall positive feedback on the different actions carried out in class was precious input for the research team and an encouragement for implementing other didactic activities which could give further emphasis to the productive application of corpora in the academic classroom.

References

Aston, Guy / Burnard, Lou 1998. *The BNC Handbook. Exploring British National Corpus with SARA.* Edinburgh: Edinburgh University Press.

Aston, Guy (ed.) 2001. *Learning with Corpora.* Bologna: CLUEB and Houston, TX: Athelstan.

Aston, Guy / Bernardini, Silvia / Stewart, Dominic (eds) 2004. *Corpora and Language Learners.* Amsterdam: John Benjamins.

Baker, Paul 2006. *Using Corpora in Discourse Analysis*. London: Continuum .

Bridgeman, Cathleen 2000. *What Kind of Peace is this? Metaphor in the US Press Coverage of the Israeli-Palestinian Negotiations*. Unpublished PhD thesis. University of South Carolina.

Burnard, Lou / McEnery, Tony (eds.) 2000. *Rethinking Language Pedagogy from a Corpus Perspective*. Bern. Peter Lang.

Conrad, Susan 2002. Corpus Linguistic Approaches for Discourse Analysis. *Annual Review of Applied Linguistics* 22, 75-95.

Cook, Guy 1998. The Uses of Reality. A Reply to Ronald Carter.. *ELT Journal*, 52/1, 57-63.

Gavioli, Laura 2001. The Learner as Researcher: Introducing Corpus Concordancing in the Classroom. In Aston, Guy (ed.) *Learning with Corpora*. Houston, TX: Athelstan, 108-137.

Gavioli, Laura 2005. *Exploring Corpora for ESP Learning*. Amsterdam: John Benjamins.

Gavriely-Nuri, Dalia 2010. The Idiosyncratic Language of Israeli 'Peace': A Cultural Approach to Critical Discourse Analysis (CCDA). *Discourse & Society* 21/5, 565-585.

Hansen, Anders 2002. Discourses of Nature in Advertising. *Communications: European Journal of Communication Research* 27/4, 499-511.

Hansen, Anders / Machin, David 2008. Visually Branding the Environment. *Discourse Studies* 10, 777-794.

Hellin Garcia, Maria Jose 2008. *Framing Terrorism via Metaphors: Jose Luis Rodriguez Zapatero (2004-2007)*. Unpublished Dissertation: University of Minnesota.

Holden, Andrew 2008. *Environment and Tourism*. Oxon: Routledge.

Huston, Susan 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Johns, Tim 2002. 'Data-driven Learning: The Perpetual Challenge'. In Kettemann, Bernhard / Marko, Georg. (eds) *Teaching and Learning by Doing Corpus Analysis*. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000, Amsterdam: Rodopi, 107-117.

Lakoff, George 1991. Metaphor and War: The Metaphor System Used to Justify War in the Gulf. *Peace Research* 23, 25-32.

Lakoff, George 2001. Metaphors of Terror. *In These Times.com* Available: <http://www.inthesetimes.com/issue/25/24/lakoff2524. html>.

McEnery, Tony / Xiao, Richard / Tono, Yukio 2006. *Corpus-based Language Studies. An Advanced Resource Book*. London: Routledge.

O'Keefe, Anne / McCarthy, Michael / Carter, Ronald 2007. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.

Pancake, Ann S. 1993. Taken by Storm: The Exploitation of Metaphor in the Persian Gulf War. *Metaphor and Symbolic Activity* 8, 281-95.

Partington, Alan 1998. *Patterns and Meanings. Using Corpora for English Language Research and Teaching.* Amsterdam: John Benjamins.

Rohrer, Tim 1995. The Metaphorical Logic of (Political) Rape: The New Wor(l)d Order. *Metaphor and Symbolic Activity* 10, 115-37.

Römer, Ute 2004. Using General and Specified Corpora in English Language Teaching: Past, Present and Future. In Campoy-Cubillo, Mari Carmen / Belles-Fortuño, Begoňa / Gea-Valor, M. Lusia (eds) *Corpus-based Approaches to English Language Teaching.* London: Continuum, 18-35.

Römer, Ute 2005. *Progressives, Patterns, Pedagogy. A Corpus-driven Approach to English Progressive Forms, Functions, Contexts and Didactics.* Amsterdam: John Benjamins.

Römer, Ute 2006. Pedagogical Applications of Corpora: Some Reflections on the Current Scope and a Wish List for Future Developments. *ZAA* 54/2, 121-134.

Ruffolo, Ida 2011. *Perceiving Nature Through Travel Promotion Texts: A Corpus-Based Discourse Study*. Unpublished PhD Dissertation. Arcavacata di Rende: Università della Calabria.

Sandikcioglu, Esra 2000. More Metaphorical Warfare in the Gulf: Orientalist Frames in News Coverage. In Barcelona, Antonio (ed.) *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective*. Berlin: Mouton de Gruyer, 299-320.

Scott, Mike 2007. *Wordsmith Tools.* Software. Oxford: Oxford University Press.

Scott, Mike / Tribble, Christopher 2006. *Textual Patterns. Key Words and Corpus Analysis in Language Education.* Amsterdam: John Benjamins.

Semino, Elena 2008. *Metaphor in Discourse*. Cambridge: Cambridge University Press.

Sinclair, John 1991. *Corpus Concordance Collocation.* Oxford: Oxford University Press.

Sinclair, John (ed.) 2004. *How to Use Corpora in Language Teaching.* Amsterdam: John Benjamins.

Steen, Gerard / Dorst, Aletta / Herrmann, Berenike / Kaal, Anna / Krennmayr, Tina / Pasma, Trijintje 2010. *A Method for*

Linguistic Metaphor Identification. Amsterdam: John Benjamins.

Steuter, Erin / Willis, Deborah 2008. *At War with Metaphor: Media, Propaganda and Racism in the War on Terror*. Lanham MD: Lexington Books.

Stubbs, Michael 1996. *Text and Corpus Analysis: Computer-assisted Studies of Language and Culture*. Oxford, UK / Cambridge, Mass: Blackwell Publishers.

Swales, John 2002. Integrated and Fragmented Worlds: EAP Materials and Corpus Linguistics. In Flowerdew, John (ed.) *Academic Discourse*. Harlow: Longman, 150-164.

Thelander, Âsa 2002. *The Image of Nature in Travel Advertisements*. Retrieved 4 May 2010, Available: www.nordicom.gu.se/mr/iceland/papers/fourteen/AThelander.rtf

Tognini-Bonelli, Elena 2001. *Corpus Linguistics at Work.*. Amsterdam: John Benjamins.

Voss, James F /Kennet, Joel / Wiley, Jennifer / Schooler, Tonya Y. 1992. Experts at Debate: The Use of Metaphor in the US Senate Debate on the Gulf Crisis. *Metaphor and Symbolic Activity* 7, 197-214.

Wichmann, Anne / Fligelstone, Steven / McEnery, Tony / Knowles, Gerry (eds.) 1997. *Teaching and Language Corpora.* London: Longman.

Widdowson, Henry 2000. The Limitations of Applied Linguistics. *Applied Linguistics* 21/1, 3-25.

ALESSANDRA LOMBARDI / SILVIA MOLETTA

7. Von der Hochschule in die Berufswelt und wieder zurück. Berufsbezogene Korpusarbeit im Unterricht *Deutsch als Fachsprache*

Vorwort

In diesem Beitrag geht es um den Einsatz von ad hoc erstellten Korpora zum Erwerb fachspezifischer Fremdsprachenkenntnisse im Unterricht Deutsch als Fremdsprache.

Nach einer kurzen Einführung in die Beweggründe für die Entwicklung des hier vorgeschlagenen experimentellen Unterrichtsmoduls (Abschnitt 1 und 2) wird im 3. Abschnitt ein knapper Überblick über das Pilotprojekt *Fachsprachen lernen mit Korpora* gegeben, das die Grundlage für den hier vorgestellten Vorschlag darstellt.

Ziele und Ergebnisse des vor einigen Jahren konzipierten und an der Katholischen Universität Brescia didaktisch umgesetzten Unterrichtsprojekts werden im 4. Abschnitt konkret veranschaulicht auf der Basis der persönlichen Erfahrung einer berufstätigen Kursteilnehmerin im Umgang mit einem selbst erstellten Fachtextkorpus aus deutschen und italienischen Einkaufsbedingungen.

Im Ausblick soll ein neues berufsorientiertes Unterrichtsmodul skizziert werden, das in Bearbeitung ist und in Zusammenarbeit mit Unternehmen der Provinz Brescia entwickelt werden soll[1].

---

1 Abschnitt 1-3 und 5 wurden von Alessandra Lombardi verfasst, Abschnitt 4 von Silvia Moletta.

1. Kontextualisierung

Wirft man einen Blick auf die aktuellen Stellenangebote von Unternehmen der Provinz Brescia, so findet man eindeutige Hinweise auf Besonderheiten des Umfeldes der lokalen Industrie, das sich durch ein Produktionsnetz aus kleinen und mittleren Unternehmen (KMU) auszeichnet, die vorwiegend im metallverarbeitenden Bereich tätig sind.

COMMERCIALE ESTERO AREA GERMANIA

Posizione:

A seguito dell'implementazione dell'ufficio commerciale di **un'azienda cliente operante in ambito metalmeccanico** situata alla prima periferia di Brescia, siamo alla ricerca di un giovane e brillante commerciale che segua la vendita di una linea di prodotti ad elevato contenuto tecnologico **per i paesi di lingua tedesca.**

Verranno prese in considerazione anche candidature junior alla prima esperienza.

Si offre inserimento in ambiente giovane e dinamico, con ottime possibilità di crescita professionale.

Requisiti:

Formazione tecnica

Ottima conoscenza tedesco, buona conoscenza inglese

Esperienza anche minima nella mansione

Disponibilità viaggi e trasferte

Sede di lavoro: prima periferia di Brescia

[Quelle: http://www.talentmanager.com/offerte-di-lavoro; 16.05.2012]

Abb. 1. Beispiel für Stellenangebot eines Brescianer Unternehmens.

Deutschland, Österreich und die Schweiz sind traditionell die wichtigsten Geschäftspartner der Brescianer Unternehmen. Das Handelsvolu-

men mit diesen Ländern beträgt im Jahre 2012 26% des gesamten Import-Exports der Provinz Brescia und bei den Exporten in die deutschsprachigen Länder steht Brescia nach Mailand an zweiter Stelle[2].

In den letzten zehn Jahren - auch angesichts der anhaltenden Wirtschaftskrise - haben sich die lokalen Bestrebungen nach mehr Internationalisierung eindeutig intensiviert. So hat die Nachfrage nach qualifiziertem Personal (insbesondere nach Fachkräften mit Fremdsprachenkompetenz, vor allem mit Deutschkenntnissen) wesentlich zugenommen.

Die Anforderungen, die an interessierte Bewerber gestellt werden, sind allerdings sehr breitgefächert, denn KMU suchen in der Regel „Multitasking-Personal!", also vielseitig einsetzbare, qualifizierte Mitarbeiter, welche neben den notwendigen Fremdsprachenkenntnissen auch über gewerblich-technische und organisatorische Kompetenzen verfügen. Dieses Profil bringen die Bewerber freilich nur selten mit.

2. Studentenpraktika in lokalen Unternehmen

Um unsere Fremdsprachen-Studierende auf die Anforderungen des lokalen Arbeitsmarkts vorzubereiten, wurde 1999 zwischen der Katholischen Universität Brescia und der Industriellenvereinigung Brescia (AIB) ein Abkommen zur Organisation von Studentenpraktika in lokalen international tätigen Unternehmen abgeschlossen. Die Kooperation bietet beiden Seiten viele Vorteile. In den letzten vier Jahren haben Studierende der Fakultät *Scienze linguistiche e letterature straniere* rund 300 Unternehmenspraktika erfolgreich absolviert.

Um einen Einblick in die Sprachtätigkeiten zu gewinnen, die Unternehmen von den PraktikantInnen forderten, wurde 2007 – im Rahmen einer Diplomarbeit im BA-Studiengang *Scienze linguistiche*

2 Quelle: Rilevazioni Export Settore Economia e Centro Studi AIB 2012.

(Fachrichtung: *Esperto linguistico d'impresa*) eine Untersuchung durchgeführt, ausgehend von den Praktikumsberichten, die von den Deutsch-Studierenden verfasst worden waren.

Die systematische Auswertung der Berichte lieferte eine breite Palette von geforderten Aufgabenstellungen (s. Tabelle unten), welche offensichtlich nur mit einem äußerst breitem Spektrum an Kompetenzen bewältigt werden können.

firmenbezogene Texte übersetzen (L2⇔L1)
firmenbezogene Texte herstellen (meist in der Fremdsprache)
Beziehungen mit ausländischen Kunden, Lieferanten, Geschäftspartnern usw. managen und gestalten (Telefon, Mail, Arbeitstreffen online u. *live*)
Firmenpräsentationen erarbeiten und vorstellen (meist in der Fremdsprache)
Marketing-Aufgaben erledigen
Gestaltung, Verwaltung und Pflege firmeninterner Datenbanken und Website
Logistik und Buchhaltung
Kundenbeanstandungen und Reklamationen bzw. Fehlermeldungen betreuen
Marktumfragen erstellen / *Kundenakquise*
Customer satisfaction / Kundenzufriedenheit ermessen
Besuche bei Kunden und potentiellen Kunden
Vorbereitung von Fachmessen, Teilnahme und *follow-up*
[Quelle: Diplomarbeit v. Rachele Cavagna 2007]

Abb. 2. Geforderte fachliche Kompetenzen und fachkommunikative Aufgaben am Praktikumsplatz.

Besonders problematisch ist − nach Angabe der PraktikantInnen − die Bewältigung von berufsbezogenen Kommunikationssituationen, insbesondere der qualifizierte, professionelle Umgang mit fachspezifischen Texten in der Fremdsprache.

Aus der Analyse der Praktikumsberichte ergab sich eine tiefe Kluft zwischen den von den Unternehmen geforderten Leistungen und den tatsächlichen Kenntnissen und Kompetenzen der PraktikantInnen, die sich von der Komplexität der Aufgaben sehr oft überfordert

fühlten. Das hing zweifellos auch zumindest teilweise mit dem Arbeitskontext bzw. den besonderen Arbeitsbedingungen in den meisten Unternehmen zusammen, d.h.

- mit einer meist eingeschränkten Auffassung der strategischen Rolle der fremdsprachlichen bzw. interkulturellen Kompetenz für eine effektive und effiziente Kommunikation mit dem Ausland (die verbreitete Einstellung sei: Wer eine Sprache studiert, der muss auch einen Firmenkatalog mühelos und zügig in die Fremdsprache übersetzen können!);
- mit der mangelnden Verfügbarkeit von finanziellen Mitteln (der Verzicht auf spezialisiertes Fachpersonal führt in den meisten Fällen dazu, dass in der Exportabteilung in der Regel nur ein Angestellter tätig ist, der für alle Zielmärkte zuständig ist und eine Vielzahl unterschiedlicher Aufgaben erledigen muss);
- mit den zur Verfügung gestellten Hilfsmittel, die oft unzureichend und zur Bewältigung qualifizierter fachkommunikativer Aufgaben ungeeignet sind (wie etwa veraltete bzw. nicht fachbezogene Nachschlagewerke oder eine inadäquate technische Ausstattung).

Trotz der erwähnten Kritikpunkte (reduzierte Vorstellung von der Komplexität firmeninterner/-externer Kommunikation, begrenzte finanzielle Ressourcen, hektische bzw. beruflich bedingte Arbeitszeiten, unzureichende Hilfsmittel) ist das oben geschilderte Arbeitsumfeld für AbsolventInnen in Fremdsprachen an der Katholischen Universität Brescia selbstverständlich von besonderem Interesse.

3. Das Pilotprojekt Deutsch als Fachsprache lernen mit Korpora

Fremdsprachenexperten, die bei kleinen oder mittelständischen italienischen Unternehmen tätig sind, müssen tagtäglich eine breite Vielfalt von Texten unterschiedlichen Fachlichkeitsgrads übersetzen, die

nicht nur sprachliche Herausforderungen mit sich bringen, sondern auch eine nicht zu unterschätzende inhaltlich-begriffliche Komplexität aufweisen und oft durch kulturspezifische Merkmale mitgeprägt werden (Beispiele dafür sind Verträge, Einkaufsbedingungen, Liefervorschriften, Beanstandungen, Bedienungsanleitungen oder technische Handbücher usw.). Die kommunikative Bewältigung dieser Übersetzungsaufgaben stellt für sprachkompetente Fachkräfte ohne übersetzerische Ausbildung eine besondere Herausforderung dar.

> (…) Dabei ist natürlich die Tatsache zu berücksichtigen, dass nach wie vor ein großer Teil der Übersetzer in Italien selbständig oder innerhalb von Übersetzerbüros arbeiten. Ein fast ebenso großer Teil tritt in ein Angestelltenverhältnis, das neben der eigentlichen Übersetzungsarbeit auch andere Aufgaben innerhalb des Betriebs (Betreuung, ausländischer Kunden, Einkauf, Verkauf, Verwaltung, usw.) verlangt. Beides ergibt sich unmittelbar aus der Tatsache, dass in Italien die Übersetzungsaufträge größtenteils von kleinen oder mittleren Firmen oder von sonstigen Auftraggebern stammen, die sich aus Gründen der Größenordnung keine eigenen Übersetzer leisten können.[3]

Um den Anforderungen des lokalen Arbeitsmarkts gerecht zu werden und den Bedarf nach mehr Praxisbezug in der universitären Ausbildung abzudecken, wurde 2007 das experimentelle Lernmodul *Deutsch als Fachsprache lernen mit Korpora* erarbeitet und bis 2011 im Bereich der Masterlehrveranstaltung *Testi specialistici della lingua tedesca* an Studierenden des 2. Jahres durchgehend angeboten.

In diesem Projektunterricht wurden selbst erstellte Mini-Fachtextkorpora als Hilfsmittel zum Erwerb von fachlichem und fachsprachlichem Wissen eingesetzt. Nach kurzer Einführung in die Fachtextanalyse und in die korpusbasierte Methode wurden die KursteilnehmerInnen aufgefordert − ausgehend von präzisen Fragestellungen − die eigenen einsprachigen (deutsch) oder zweisprachigen (italienisch-deutsch) Korpora auszuwerten, um Einsicht in strukturelle Merkmale bzw. in rekurrente Sprachgebrauchsmuster einzelner Fachtextsorten zu gewinnen.

---

3 Vgl. Soffritti (2004: 113-114).

Angestrebt wurde allerdings nicht die Vermittlung konkreter rezeptiver und produktiver Fachsprachenkenntnisse sondern der Erwerb eines *methodischen Wissens* für den selbständigen und effektiven Umgang mit deutschen Fachtexten.

Die Ergebnisse des Pilotprojekts waren sehr ermutigend[4], denn die Studierenden hatten zudem die Möglichkeit, neben den fremdfachsprachlichen Kenntnissen auch berufsbezogene Anwendungskompetenzen auszubauen, wie der nachfolgende Erfahrungsbericht einer berufstätigen Kursteilnehmerin veranschaulichen soll.

Im Rahmen des experimentellen Unterrichtsmoduls *Deutsch als Fachsprache lernen mit Korpora* hatte die Kursteilnehmerin die Möglichkeit, ihre am Arbeitsplatz erstellten Übersetzungen unter die Lupe zu nehmen und die im Kurs trainierten Arbeitsverfahren für ihren Berufsalltag nutzbar zu machen.

4. Das Pilotprojekt *Deutsch als Fachsprache lernen mit Korpora* aus Sicht einer berufstätigen Kursteilnehmerin

Der folgende Erfahrungsbericht wurde im Rückblick von einer der Kursteilnehmerin (des MA-Studiengangs *Scienze linguistiche* – Fachrichtung *Management internazionale*) verfasst und soll einen Einblick in Hintergründe des Projekts und Vorgehensweise der Studierenden vermitteln:

Ich bin zuständig für die Exportabteilung eines kleinen Familienunternehmens mit Sitz in der Provinz Brescia, das als Aluminiumdruckguss-Zulieferant für verschiedene Sektoren zuständig ist. Nach mehr als dreißig Jahren Tätigkeit auf dem italienischen Markt hat sich diese Firma kürzlich die Herausforderung der Internationalisierung gestellt, mit Deutschland und der Schweiz als erstrangige Zielmärkte.

4 Ausführlich dazu Lombardi (2011).

Nach drei Monaten Praktikumserfahrung im dritten Universitätsjahr wurde ich von der Firmainhaberin fest eingestellt. Nach meinem Bachelor-Abschluss habe ich als Teilzeitkraft gearbeitet und parallel dazu die zweijährige Ausbildungsstufe *Laurea Specialistica* abgeschlossen. Unter den kommunikativen Aufgaben, mit denen ich am Arbeitsplatz beauftragt wurde, war die Übersetzung von Einkaufsbedingungen aus dem Deutschen ins Italienische. Allgemeine Einkaufsbedingungen werden von einer Vertragspartei, dem Käufer, einseitig gestellt und sind ein primäres Instrument der Geschäftsabwicklung. Sie schaffen die sämtlichen Voraussetzungen und Gegebenheiten, unter denen ein Kauf erfolgt. Neben ihrer wirtschaftlichen Rolle haben sie gleichzeitig einen gesetzlichen Wert, weil sie den rechtlichen Rahmen eines Vertrags schaffen. Aufgrund ihrer komplexen fachspezifischen Natur bereitete mir die Übersetzung dieser Dokumente von Anfang an einige Schwierigkeiten.

Besonders problematisch war die Rekonstruktion der jeweils kontextgebundenen Bedeutungen von Präfixen in textsortentypischen Wortbildungen (s. Beispiele unten zum Fachwort *Lieferung*) sowie die Auffindung angemessener italienischer Entsprechungen angesichts der lückenhaften Aufnahme in den herkömmlichen von mir nachgeschlagenen zweisprachigen Wörterbüchern.

Durch den Aufbau und Auswertung eines im Unterricht *ad-hoc*-erstellten Vergleichskorpus aus deutschen und italienischen Einkaufsbedingungen konnte ich meine fachsprachliche Kompetenz erweitern und die idiomatische Qualität meiner Übersetzungsleistungen erhöhen. Nachfolgend die Kriterien zum Aufbau meines textsortenspezifischen Vergleichskorpus[5]:

Deutsches Teilkorpus		**Italienisches Teilkorpus**	
Umfang	50.410 Wörter	**Umfang**	79.878 Wörter
Texte	25 ungekürzte Einkaufsbedingungen (von	**Texte**	25 ungekürzte Einkaufsbedingungen (von verschiedenen

5 Die Kriterien basieren auf dem Wunschzettel nach Bowker/Pearson (2002: 193-209).

	verschiedenen Unternehmen)		Unternehmen)
Medium	Geschrieben	**Medium**	Geschrieben
Fachbereich	Wirtschaft, Recht	**Fachbereich**	Wirtschaft, Recht
Autoren	Zuständige der Einkaufsabteilung	**Autoren**	Zuständige der Einkaufsabteilung
Sprache	Deutsch (Muttersprachler)	**Sprache**	Italienisch (Muttersprachler)
Erscheinungs-datum	Zwischen 2004 und 2007	**Erscheinungs-datum**	Zwischen 2004 und 2007

Abb. 3. Textsortenspezifisches Vergleichskorpus aus deutschen und italienischen Einkaufsbedingungen.

Nachdem ich meine Übersetzungsschwierigkeiten erkannt und eingeschätzt hatte, habe ich das Korpus mit *WordSmith Tools* (Version 4.) durchsucht. Ausgehend von den Konkordanzen der im deutschen Teilkorpus aufgedeckten Schlüsselwörter habe ich eine kontextgebundene Analyse durchgeführt und angemessene Entsprechungen aus dem italienischen Teilkorpus herausgefiltert. Ein häufig auftretendes Übersetzungsproblem war für mich z.B. die fachlich korrekte Übersetzung deutscher Komposita, die in den von mir betrachteten Einkaufsbedingungen reichhaltig vorkommen und zentrale Begriffe vermitteln und denen im Italienischen oft syntagmatische Wortverbindungen entsprechen. Durch die Funktion *Cluster* konnte ich z.B. für das deutsche Kompositum *Produkthaftung* im italienischen Teilkorpus die entsprechende Mehrwortbenennung *responsabilità da prodotto* relativ schnell auffinden. Es sei hier angemerkt, dass die erste von mir gelieferte Übersetzung von *Produkthaftung responsabilità sul prodotto* war, ein zwar akzeptables und verständliches Pendant zum deutschen Kompositum das aber im italienischen Teilkorpus kaum belegt war.

Analogerweise waren beim ersten Übersetzungsversuch die Fachbezeichnungen *Produkthaftpflicht- und Rückrufversicherung* von mir eher intuitiv und teilweise inkorrekt übersetzt worden; die entsprechende vertragliche Bezeichnung *(assicurazione per responsabilità da prodotto e per ritiro dei prodotti)* trat allerdings im italienischen Teilkorpus mehrmals auf und konnte durch die Korpusabfrage

leicht ermittelt werden. Weitere im Vergleichskorpus aufgespürte fachspezifische Wörter und textsortenangemessene Formulierungen konnte ich direkt übernehmen und zur Steigerung der fachsprachlichen Qualität meiner Übersetzungsleistungen gezielt einsetzen.

Ein weiterer Vorteil der korpusbasierten Analyse war die Möglichkeit, die Produktivität bzw. das textsortenbedingte Wortbildungspotenzial rekurrenter Affixe systematisch auszuloten und potentielle Entsprechungen im Italienischen aufzufinden[6].

Das soll am Beispiel des Wortes *Lieferung* veranschaulicht werden, das im deutschen Teilkorpus eine Vielzahl von Derivaten aufweist und als Grundwort zahlreicher Determinativkomposita vorkommt.

	Deutsches Teilkorpus	**Italienisches Teilkorpus**
1.	Die Lieferung Jeder Lieferung ist ein Lieferschein beizufügen.	Consegna (287) In tal caso la consegna s'intenderà effettuata solo quando l'intera quantità ordinata sarà pervenuta al luogo di destinazione.
2.	Die Lieferungen (Pl.) Der Lieferant garantiert bei seinen Lieferungen die anerkannten Regeln der Technik, die vereinbarten technischen Daten und Spezifikationen sowie die Eignung zum Vertragszweck.	Consegne (47) Eventuali ritardi nelle consegne non daranno, in ogni caso, adito al Committente ad ottenere riduzioni di prezzo
3.	Die Nachlieferung Hält der Auftragnehmer den Liefertermin nicht ein, so ist der Auftraggeber ohne weitere Nachfristsetzung nach eigener Wahl berechtigt, Nachlieferung, Schadenersatz statt der Leistung wegen nicht oder nicht wie	Consegna sostitutiva Fornitura dei pezzi di ricambio La Committente godrà inoltre di tutti i diritti alla garanzia previsti dalla legge avendo la facoltà di pretendere dal Fornitore o l'eliminazione dei difetti o mancanza di qualità, oppure una fornitura sosti-

6 Vgl. Krausse (2008: 236).

	geschuldet erbrachter Leistung zu verlangen oder vom Auftrag zurückzutreten. Bei Lieferung mangelhafter Ware wird dem Lieferanten Gelegenheit zur Nacherfüllung (Nachbesserung / Nachlieferung) gegeben. Der Lieferant garantiert die Nachlieferung von Ersatzteilen bzw. Baugruppen 10 Jahre ab Lieferdatum.	tutiva con tutte le relative spese a carico dello stesso Fornitore senza con questo pregiudicare in alcun modo l'eventuale rivendicazione da parte della Committente dei diritti al risarcimento.
4.	Die Auslieferung Der Auftraggeber hat das Recht, nach eigenem Ermessen die Auslieferung der Werkzeuge zu verlangen oder die Werkzeuge durch den Lieferanten, für den Auftraggeber kostenfrei, verschrotten zu lassen. Bei Waren, die von uns an Dritte weiterverkauft werden, endet die Gewährleistungsfrist frühestens zwei Monate nach dem Zeitpunkt, in dem wir die Ansprüche des Dritten erfüllt haben, spätestens jedoch fünf Jahre nach Auslieferung der Ware an uns.	Restituzione consegna Mezzi di produzione / disegni / ecc. 9.1 I mezzi di produzione quali stampi, matrici, calibri, modelli, campioni, strumenti, disegni ecc. che la Committente abbia messo a disposizione del Fornitore dovranno essere restituiti alla stessa Committente a semplice richiesta di quest'ultima.
5.	Die Ersatzlieferung Die gesetzlichen Mängelhaftungsansprüche stehen uns ungekürzt zu. Wir sind berechtigt, vom Lieferanten nach unserer Wahl Mängelbeseitigung oder Ersatzlieferung zu verlangen.	Merce riparata Merce fornita in sostituzione Consegna sostitutiva Lotto di sostituzione In caso di sostituzione di forniture difettose il termine di garanzia decorre ex novo dalla data di consegna della merce/prestazione riparata o fornita in sostituzione. La Committente godrà inoltre di tutti i diritti alla garanzia previsti dalla legge avendo la facoltà di pretendere dal Fornitore o l'eliminazione dei difetti o mancanza di

		qualità, oppure una fornitura sostitutiva con tutte le relative spese a carico dello stesso Fornitore senza con questo pregiudicare in alcun modo l'eventuale rivendicazione da parte della Committente dei diritti al risarcimento. ACTEA avrà la facoltà di sospendere, dal totale dei pagamenti in quel momento dovuti al Fornitore, il pagamento di un importo pari al prezzo del lotto interessato, fino a quando il Fornitore non abbia consegnato i particolari o il lotto di sostituzione.
6.	Die Teillieferung Sind Teilmengen vereinbart, so ist der Lieferant verpflichtet, TRUMPF bei jeder Teillieferung die jeweils verbleibende Restmenge mitzuteilen.	Consegna parziale Non si accetteranno consegne parziali.
7.	Die Anlieferung Bei früherer Anlieferung als vereinbart behalten wir uns das Recht vor, die Rücksendung auf Kosten des Lieferanten vorzunehmen. Erfolgt die Anlieferung vor dem vereinbarten Liefertermin, behält sich die Stadtwerke Düsseldorf AG vor, die Lieferung nicht anzunehmen und diese auf Kosten und Gefahr des Lieferanten zurückzusenden.	Consegna anticipata Le forniture consegnate prima della data concordata si intendono consegnate soltanto alla data di consegna concordata.

Abb. 4. Ableitungen und Komposita von *Lieferung* im textsortenspezifischen Vergleichskorpus (Teil 1).

Wie aus der Tabelle ersichtlich, sind im italienischen Teilkorpus oft für ein und dasselbe deutsche Kompositum mehrere Entsprechungen zu finden wie z. B. *Ersatzlieferung* (5): *merce riparata, merce fornita in sostituzione, consegna sostitutiva lotto di sostituzione*. Das sind Va-

rianten, die je nach Kontext in den italienischen Einkaufsbedingungen praktisch synonymisch verwendet werden.

Interessant aber schwer zu dekodieren ist die fachspezifische funktionale Leistung einiger Präfixe wie „*An*“ in *Anlieferung (7)*, dem meist eine temporale Bedeutung zugeschrieben wird, wie aus der Analyse der Verwendungskontexte im deutschen Teilkorpus ermittelt werden konnte. Das Wort *Anlieferung* wird nämlich hier fast ausschließlich mit Bezug auf *frühere Lieferungen* verwendet.

8.	Die Lieferungsbedingungen (Pl.) Liefertermine und Lieferungsbedingungen.	Condizioni di consegna In particolare, qualora le condizioni di consegna concordate fossero DDP..
9.	Die Erstlieferungen (Pl.) – (1) Erstlieferungen ist ein Erstmusterprüfbericht beizufügen.	Consegna della prima campionatura Il Fornitore ha l'obbligo di consegnare beni identici agli eventuali campioni depositati in sede di gara e oggetto dell'aggiudicazione e/o dell'affidamento, mentre ha l'obbligo di ritirare, a proprio carico e spesa, quei beni non giudicati identici alla campionatura,
10.	Über- / Unterlieferungen (Pl.) Über- oder Unterlieferungen sind nur nach zuvor mit TRUMPF getroffenen Absprachen zulässig.	Quantità diversa Se non diversamente indicato nel PO il materiale non sarà considerato accettato dall' Acquirente fino alla data di consegna richiesta oppure se fornito in quantità diversa da quanto specificato nel PO.
11.	Die Ablieferung (11) Mängelansprüche sind rechtzeitig geltend gemacht, wenn sie innerhalb von 14 Tagen nach Ablieferung der Ware an den Verkäufer abgesandt werden.	Consegna Data di consegna Eventuali ammanchi dei prodotti spediti, devono essere notificati dall'acquirente entro e non oltre cinque giorni dalla data della consegna.

	Alle zu liefernden Maschinen, Apparate, Anlagen, Fahrzeuge und dergleichen müssen den zur Zeit der Ablieferung geltenden Bestimmungen des Geräte- und Produktsicherheitsgesetzes und seinen Verordnungen sowie der Maschinenrichtlinie (98/37/EG).	
12.	Die Warenlieferungen (Pl.) Zur Vermeidung verzögerter Bearbeitung bei TRUMPF sind Rechnungen nicht den Warenlieferungen beizufügen, sondern mit gesonderter Post zu übermitteln. Weiter hat der Lieferant Sorge dafür zu tragen, dass sämtliche erforderlichen Transport- und Frachtpapiere der Warenlieferung beigefügt sind.	Beni Merce Materiale Se non diversamente indicato nell'Ordine, il Venditore emetterà una fattura per ogni Ordine e, a meno che sia specificato un indirizzo di fatturazione diverso, le fatture dovranno accompagnare i Beni.

Abb. 5. Ableitungen und Komposita von *Lieferung* im textsortenspezifischen Vergleichskorpus (Teil 2).

Ein weiteres Beispiel ist das Wort *Ablieferung* (11), das in den meisten Kontexten mit Bezug auf das Lieferungsdatum benutzt wird. In diesem Fall könnte *data di consegna* eine textsortenangemessene Übersetzung sein.

Das Kompositum *Warenlieferungen* (12) wird benutzt, wenn es um die Unterlagen geht, die der Ware beizufügen sind. Auf Italienisch sind daher in diesem Sinne die Entsprechungen *merce*, *lotto*, *materiale*, *beni* statt *consegna* zu bevorzugen.

13.	Die Weiterlieferung Der Lieferant haftet dafür, dass weder die von ihm gelieferte Ware noch deren Weiterlieferung, -verarbeitung oder Benutzung durch TRUMPF Schutzrechte Dritter, insbesondere	Vendita Per vizi, difetti e non conformità rilevati nel periodo fra l'uscita dallo stabilimento di produzione dei sistemi di trattamento dell'aria e/o ai sistemi di trattamento delle

	Gebrauchsmuster, Patente oder Lizenzen verletzt.	acque, dei fanghi, o del gruppo sul quale il particolare viene montato (o, ove del caso, dopo la vendita o come ricambio o come parte sciolta) e la scadenza della garanzia data da ACTEA al cliente, avrà le seguenti facoltà.
14.	Mehr- / Minderlieferungen (1) Die Anerkennung von Mehr-oder Minderlieferungen behält Rhenus sich vor.	Quantità diversa (siehe Punkt 10)
15.	Die Sammellieferung Soweit wir mit den Transportkosten gesondert belastet werden, müssen den Rechnungen ferner die Originale und Kopien der Frachtbriefe mit voller Angabe der Fahrtstrecke, Wagennummer usw. und die Transportrechnungen beigefügt werden; im Falle einer Sammellieferung müssen diese Rechnungen das Gewicht und den Teilbetrag der gelieferten Waren angeben.	Consegna cumulativa Le fatture devono riguardare un ordine solo, ovvero non devono essere cumulative, anche se per merce uguale di più ordini, anche se telefonici.
16.	Die Zulieferungen (Pl.) Der Lieferant hat in jedem Fall auch ohne Verschulden für die von ihm etwa von Dritten beschafften Zulieferungen und Leistungen wie für eigene Lieferungen oder Leistungen einzustehen.	Consegne da parte di subfornitori In nessun caso sono comprese fra le cause di forza maggiore i ritardi di lavorazione e/o ritardate consegne da parte di subfornitori.
17.	Die Ersatzteillieferungen (Pl.) Für nachgebesserte Teile oder für Ersatzlieferungen beginnt die Verjährungsfrist für Sachmängelansprüche nach erfolgter Nacherfüllung neu zu laufen.	Fornitura sostitutiva La Committente godrà inoltre di tutti i diritti alla garanzia previsti dalla legge avendo la facoltà di pretendere dal Fornitore o l'eliminazione dei difetti o mancanza di qualità, oppure una fornitura sostitutiva.

18.	Die Belieferung (1) Soweit unsere Lieferer uns wiederholt oder laufend beliefern, verpflichten sie sich, nach Auslauf der Fertigung eines an uns gelieferten Produkts noch auf die Dauer von weiteren 8 Jahren unsere Belieferung mit Ersatzteilen zu gewährleisten, oder uns von der Produktionseinstellung frühzeitig zu benachrichtigen, so dass wir die Möglichkeit haben, eine Umstellung vorzunehmen oder eine Rest- Order zu erteilen!	Consegna Consegna della fornitura Fornitura La garanzia cessa con il decorrere di 24 mesi dalla consegna della fornitura.
19.	Die Hauptlieferung Die Lieferung der Technischen Dokumentation und aller geforderten Protokolle muss, wenn nicht anders vereinbart, Bestandteil der Hauptlieferung sein.	*

Abb. 6. Ableitungen und Komposita von *Lieferung* im textsortenspezifischen Vergleichskorpus (Teil 3).

Durch die Konkordanzanalyse konnte ich weitere sprachspezifische und textsortentypische Ausdrucksformen ermitteln, die für eine sprachlich und fachlich korrekte Übersetzung relevant sind wie z.B. Unterschiede im Numerus: *Ersatzlieferung* (17) ist im deutschen Teilkorpus sowohl im Singular als auch im Plural zu finden, während *fornitura sostitutiva* (18) im italienischen Teilkorpus ausschließlich im Singular vorkommt. Das Präfix Be- in *Belieferung* (18) modifiziert die Bedeutung vom Stammwort *Lieferung* anscheinend nicht.

Da ein Korpus − wenn auch reflektiert und nach überprüfbaren Kriterien aufgebaut − immer nur ein Ausschnitt aus der Gesamtheit aller produzierten Exemplare einer Textsorte darstellt, kann die Suche im Vergleichskorpus keineswegs als Allheilmittel betrachtet werden. Das ist klar ersichtlich aus der Tatsache, dass nicht für alle deutschen Komposita Entsprechungen im italienischen Teilkorpus ermittelt wer-

den konnten. Für das Wort *Hauptlieferung* (19) z.B. war kein einziger Beleg zu finden, da es sich hier höchstwahrscheinlich um eine kontextgebundene Wortbildung handelt.

Die Bilanz der Arbeit an und mit einem Korpus hat sich als sehr positiv erwiesen: durch die Korpusabfrage hatte ich Zugang zu einer großen Menge von Verwendungsbeispielen, die in Wörterbüchern nicht oder selten zu finden sind (s. *Produkthaftpflicht- und Rückrufversicherung*). Die Analyse von Fachwörtern im Kontext und der vergleichende Ansatz haben sich bei der Ermittlung textsortenangemessener Übersetzungslösungen als äußerst nützlich erwiesen (s. *Produkthaftung* und *Über- / Unterlieferung*). Durch die Konkordanzanalyse hatte ich ferner die Möglichkeit, die pragmatischen und funktionalen Leistungen vieler Affixe aus dem Kontext zu erschließen und unbekanntes Vokabular deduktiv zu lernen, was für Nichtmuttersprachler, die als Fremdsprachexperten in kleinen Unternehmen tätig sind, von großem Vorteil ist, wie am Beispiel *Anlieferung* gezeigt wurde.

Die im Kurs durchgeführte Korpusauswertung war darüber hinaus der Ausgangspunkt zur Erstellung eines firmeninternen elektronischen Fachglossars, das sowohl wirtschaftliche als auch juristische Termini umfasst und laufend aktualisierbar ist. Das ist ein sehr nützliches Hilfsmittel für den Fremdsprachenexperten, der auf diese Weise die Möglichkeit hat, gespeichertes terminologisches Wissen jederzeit schnell abzurufen und für neue kommunikative Aufgaben nutzbar zu machen, so wie Feyrer (2004: 32) hervorhebt:

> Durch die Internationalisierung von Wissen und fachlicher Zusammenarbeit macht Informations- und Wissensmanagement einen integrativen Bestandteil der Tätigkeit des Translators aus. Als Experte für intra- und interkulturelle bzw. -sprachliche Kommunikation hat er einem vielschichtigen und heterogenen Anforderungsprofil gerecht zu werden, das über die Funktion als reiner Sprach- und Kulturmittler weit hinausgeht. Er muss nicht nur Fachkommunikation transkulturell und intersprachlich realisieren, sondern auch auf die spezifischen Bedingungen der intendierten Rezeption des Zieltextes und des Zielpublikums im Sinne einer skoposorientierten Adaptation und Neugestaltung des Ausgangsmaterial eingehen können. Dazu bedarf es eines kom-

> plexen, die Besonderheiten von Ausgangs- und Zielsprache und -kultur wie auch die spezifischen Rahmenbedingungen der konkreten Arbeitssituation berücksichtigenden Informationsmanagement.[7]

Gelungenes, strategisches Informations- und Wissensmanagement gehört meiner Meinung nach zu den notwendigen Kernkompetenzen von Fremdsprachenexperten in KMU: die Speicherung und Verwaltung häufig vorkommender Fachwörter und der entsprechenden Übersetzungen ermöglicht es dem Fremdsprachenexperten, seine Textproduktionsaufgaben viel schneller und effektiver zu erledigen. Ein derartiges produktionsorientiertes Firmenglossar sollte vom auftragsgebenden Unternehmen als wertvolle nachhaltige Ressource angesehen werden.

5. Ausblick

Die europaweite ELAN-Studie über die Effekte von Fremdsprachendefiziten in kleinen und mittleren Unternehmen (*Effects on the European Economy of Shortages of Foreign Language Skills in Enterprise*)[8] ergab 2006, dass die meisten befragten KMU aufgrund mangelnder Fremdsprachenkenntnisse Exportaufträge nicht erhalten haben. Es fehlt offenbar an Schlüsselkompetenzen, die auch für die KMU von Brescia und Provinz äußerst strategisch sein können.

Vor diesem Hintergrund, und ausgehend von der positiven Erfahrung von Silvia Moletta (Punkt 4) sowie anderer Kursteilnehmerinnen wird das hier vorgestellte Unterrichtsmodul nun in Zusammenarbeit mit lokalen Unternehmen, die Interesse am Projekt gezeigt haben, weiterentwickelt. Es soll in Form eines didaktischen Zusatzangebotes für ausgewählte Gruppen von interessierten Studie-

---

7 Vgl. Feyrer (2004: 32).

8 Vgl. <http://ec.europa.eu/languages/documents/doc421_en.pdf> (letzter Abruf September 2013).

renden des 1. und 2. Jahres des Masterstudiengangs in *Scienze linguistiche* aufgenommen werden. Im Vergleich zum früheren Modul besteht die Neuheit darin, dass die Dozentin von qualifizierten Absolventinnen, die in lokalen Unternehmen tätig sind und korpuslinguistisch wie terminologisch ausgebildet sind, didaktisch unterstützt werden soll.

Im Kurs sollen im Auftrag der am Projekt beteiligten Unternehmen authentische fachkommunikative Aufgaben ausgeführt werden (wie z.B. die Erarbeitung einer Firmenpräsentation in deutscher Sprache oder die Lokalisierung der Firma-Website für deutschsprachige Adressaten oder die Übersetzung ins Deutsche von Unternehmensbroschüren bzw. des Firmenkatalogs u.a.). Dadurch sollen die Studierenden − zumindest ansatzweise – auf konkrete berufsrelevante Sprachtätigkeiten vorbereitet werden.

Methodisch soll das durch den Aufbau und die Abfrage von speziell für den jeweiligen Verwendungszweck erstellten Fachtextkorpora aus der jeweils ausgewählten firmenbezogenen Dokumenten erfolgen. Der zweite Schritt besteht in der Aufnahme und Aufbereitung der aus den Korpora gewonnenen Informationen in eine terminologische Datenbank.

Das Modul setzt sich also primär zum Ziel, den Wissens- und Kompetenzenaustausch zwischen universitärer Ausbildung und lokaler Berufswelt zu fördern – mit Vorteilen für beide Akteure. Zum einen werden die Kursteilnehmerinnen bei der Aufbereitung des firmenbezogenen Materials von der Fachberatung des Unternehmenspersonals profitieren können, zum anderen – und darüber hinaus – wird den Unternehmen ein qualitativ hochwertiges Kommunikationsprodukt und eine vielseitig einsetzbare terminologische Ressource zur firmeninternen Benutzung bereitgestellt.

Vorgesehen ist auch eine Vorstellung der jeweiligen Projekte vor den beteiligten Unternehmen. Das soll die Motivation der Studierenden stärken und ihr Bewusstsein bezüglich der besonderen Anforderungen des Arbeitens als Fremdsprachenexperten in lokalen Unternehmen fördern. Die Unternehmen sollen ihrerseits für die Vorteile einer nachhaltigen Qualitätssicherung ihrer Kommunikationsprozesse mit dem Ausland sensibilisiert werden. Denn eine effektive und effiziente internationale Unternehmenskommunikation fordert nicht nur

solide fach- und berufsbezogene Fremdsprachenkenntnisse, sondern – wie oben angedeutet – auch methodisches Wissen, ein adäquates Zeitpensum und aufgabengerechte Ressourcen.

Zitierte Literatur

Bowker, Lynne / Pearson, Jennifer 2002. *Working with specialized language. A practical guide to using corpora*. New York: Routledge.

Feyrer, Cornelia 2004. Informationsmanagement im Fokus „kundenorientierter" Textrezeption und -produktion. Ein Auftrag an die Translationsdidaktik. In Göpferich, Susanne / Engberg, Jan (Hgg.) *Qualität fachsprachlicher Kommunikation*. Tübingen: Gunter Narr Verlag, 31-50.

Krausse, Sylvana 2008. On the role of student research in the ESP classroom. A call for sustainable language skills. In Fortanet-Gomez, Inmaculada / Raisanen, Christine A. (Hgg.) *ESP in European Higher Education. Integrating Language and Content*. Amsterdam u. a.: Benjamins, 233-243.

Lombardi, Alessandra 2011. Fachsprachen lernen mit Korpora. Die Integration korpuslinguistischer Verfahren im Unterricht Deutsch als Fachsprache. In Knorr, Dagmar / Nardi, Antonella (Hgg.) *Fremdsprachliche Textkompetenz entwickeln*. Frankfurt/M. u.a.: Peter Lang, 217-236.

Soffritti, Marcello 2004. Fachübersetzung, Berufsbilder und Studienordnung – Grundperspektiven im Rahmen des aktuellen italienischen Hochschulgesetzes. In Rega, Lorenza / Magris, Marella (Hgg.) *Übersetzen in der Fachkommunikation – Comunicazione specialistica e traduzione*. Tübingen: Gunter Narr Verlag, 105-115.

NATACHA S.A. NIEMANTS

L'utilisation de corpus d'entretiens cliniques (français / italien) dans la didactique de l'interprétation en milieu médical

1. Introduction

Quand on parle de corpus dans la recherche et la didactique des langues, on pense généralement à des analyses quantitatives, à des investigations sur des éléments qui reviennent dans de grandes collections de textes écrits qu'un seul chercheur peut assez aisément construire. Lorsqu'on a à faire avec l'oralité de la parole (Galazzi 2002), on se heurte à la nécessité de transformer un discours oral dans un format analysable, ce qui requiert beaucoup de temps et d'énergie et ce qui limite la quantité de choses qu'un seul chercheur peut à la fois collecter et analyser.

Pour avoir des quantités de données orales qui se prêtent à des analyses quantitatives, les chercheurs s'engagent parfois dans des projets collectifs, où c'est l'ensemble de nombreux "sub-corpus" (Straniero Sergio / Falbo 2012) qui constitue une grande banque de données[1]. A cette fin, chacun de ces sous-corpus doit être construit de manière à satisfaire aussi bien les exigences locales d'un projet individuel que les exigences globales d'une entreprise partagée.

Dans ces pages nous nous attarderons sur deux exemples de sous-corpus bilingues français / italien. Le premier se compose d'entretiens simulés où deux professeurs et un étudiant jouent

1 C'est le cas, par exemple, de la banque de données outillée CLAPI (Corpus de Langue Parlée en Interaction), réalisée par le groupe ICOR (Université Lumière Lyon 2) et accessible à la page <http://corpus.univ-lyon2.fr>

respectivement le rôle du médecin italophone, du patient francophone et de l'interprète qui les aide à se comprendre (Tableau 1). Le second se compose d'entretiens cliniques réels, où de vrais médecins et patients communiquent avec l'aide d'un interprète dans des hôpitaux italiens et belges (Tableau 2).

Cet article s'articulera en trois volets : nous commencerons par préciser comment nous avons transcrit en conciliant des exigences locales et globales ; nous présenterons les quelques résultats de l'analyse des deux sous-corpus, ainsi que la manière dont ces résultats peuvent informer la formation des interprètes de dialogue ; nous finirons en avançant une proposition d'utilisation pédagogique du sous-corpus réel.

2. La transcription des entretiens cliniques

La transcription est le résultat de choix interprétatifs (*quoi* transcrire ?) et représentatifs (*comment* transcrire ?) qui se font entre deux ambitions en quelque sorte contradictoires : celle de transcrire tout ce que l'on entend et celle de faire en sorte que la transcription soit lisible. La transcription est donc une affaire de compromis qui dépendent du cadre théorique et des objectifs du chercheur.

La méthode de transcription choisie ne fait pas de grande différence quand on a à faire avec un nombre restreint d'interactions. Le chercheur connaît si bien ce qu'il transcrit, qu'il peut aisément localiser les faits linguistiques qui reviennent et / ou qui l'intéressent dans le cadre de son projet. Mais si l'on envisage de réutiliser le sous-corpus pour d'autres fins, voire dans le cadre d'un projet collectif, il faut alors prévoir la possibilité de modifier aussi bien le *quoi* que le *comment*.

2.1. Le logiciel EXMARaLDA

Le logiciel EXMARaLDA (Schmidt 2004) a été notre choix de compromis, dans la mesure où il permet d'ajouter et de cacher des détails – intervenant donc sur le *quoi* transcrire – ainsi que de visualiser et d'exporter les interactions en de multiples formats – ce qui relève plutôt du *comment*.

Ce logiciel fonctionne comme dans une notation musicale où à chaque interlocuteur correspond une ligne et où d'autres lignes peuvent être ajoutées pour de plus amples détails et / ou cachées lorsque ceux-ci n'intéressent pas (Figure 1). Prenons l'exemple de nos propres transcriptions. Une exigence locale de notre projet était la nécessité de transcrire les énoncés des participants italophones et francophones ainsi que les silences dans leurs conversations. Si un autre chercheur disposait de nos transcriptions, il pourrait bien vouloir ajouter des informations sur la façon dont ces énoncés sont prononcés ou, encore, des traductions de ces énoncés pour un public qui ne parle ni l'italien ni le français. Il lui suffirait d'ajouter deux lignes. Ce même chercheur pourrait tout aussi simplement cacher la ligne des silences [s] ou celle que nous avons consacrée à la description d'évènements concurrents [e] au cas où ces informations ne seraient pas pertinentes à son projet.

	9 [08:20.]	10 [08:20.8]	11 [08:21.5]	12 [08:23.1]	13 [08:25.2]
[s]	<pausa>				
[e]					
OPE [v]		buongiorno			
INT [v]			<pausa> bonjour madame		buongiorno dottore
PAZ [v]				<dur=1> bonjour docteur	

Figure 1. La notation musicale.

De plus, EXMARaLDA relie automatiquement la transcription à l'audio qui lui correspond : en double cliquant sur un segment, on peut visualiser et écouter le morceau de ligne sonore concerné (Figure 2). Cela facilite le travail du premier chercheur, qui peut revenir sur les transcriptions et les affiner, et également celui des chercheurs

successifs, qui pourront ajouter de plus amples détails sur un segment en l'écoutant à nouveau.

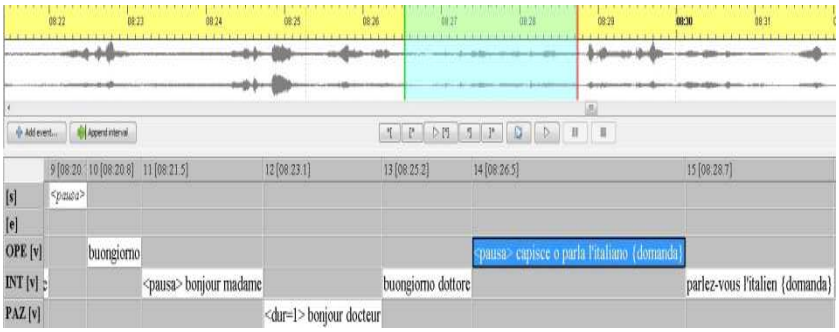


Figure 2. Le lien avec l'audio.

Par ailleurs, ce logiciel dispose d'un outil d'interrogation intégré, EXAKT, qui permet à tout instant de faire une concordance et de vérifier la fréquence d'un fait linguistique (cf. l'hésitation *ehm* dans la Figure 3).

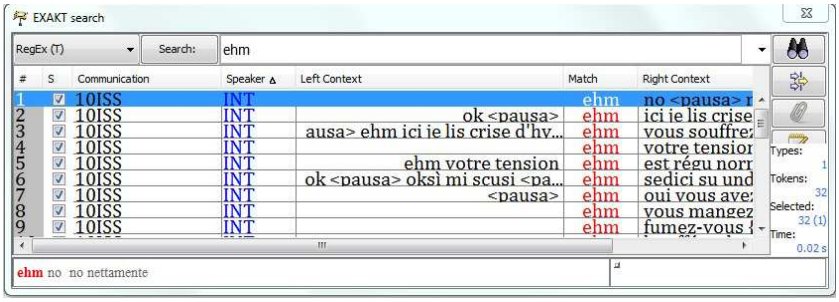


Figure 3. L'outil d'interrogation.

Le dernier avantage, et non des moindres, est la possibilité d'exporter les transcriptions en plusieurs formats, choisissant ainsi *comment* traiter les données en fonction du type d'analyse. On peut choisir, entre autres : un format RTF, pour visualiser en notation musicale sur des ordinateurs qui n'ont pas le logiciel installé ; un format TXT, pour

visualiser ce que les participants se disent tour par tour ; un format HTML, pour le partage des données sur la toile suivant soit la progression temporelle soit l'enchainement des tours de l'interaction ; et un format XML, pour l'interrogation du corpus avec d'autres logiciels. Compte tenu de tous ces formats et de la distinction proposée par Straniero Sergio et Falbo entre "spoken corpus" et "speech corpus" (2012 : 31-32), EMARaLDA donne la possibilité de créer les deux : à la fois une collection de transcriptions de données audio ou vidéo (*spoken corpus*) et un corpus multimodal où les transcriptions et les fichiers audio / vidéo font partie du corpus lui-même (*speech corpus*).

3. La linguistique informe la didactique

D'après une perspective pragmatique, le *contexte* de l'interaction contribue largement à la construction du sens. Un énoncé comme "les poissons levez la main"[2] assume une signification différente en fonction du contexte dans lequel il est prononcé : dans le contexte d'une école, il pourrait signifier que tous les enfants du signe des poissons doivent lever la main ; dans le contexte d'un restaurant, il pourrait signifier que ceux qui mangeront du poisson, et non pas de la viande, doivent faire autant.

Puisque comprendre ce qui est dit est fondamental pour communiquer les uns avec les autres, les recherches sur la langue parlée ont essayé d'expliquer le fonctionnement de la *compréhension*. Ils l'ont d'abord traitée comme un fait cognitif, privé, individuel et puis, suite aux études ethnométhodologiques, comme un accomplissement collectif et public (Mondada 2011 : 542). En

2 Nous payons notre double dette envers Jacques Cosnier, qui a l'habitude d'utiliser cet exemple, et Véronique Traverso, qui l'a emprunté lors d'un cours qu'elle a tenu à nos étudiants de traduction et médiation française à l'Université de Modène et de Reggio Emilia le 16 mars 2012.

d'autres termes, on a cessé de faire des hypothèses sur ce qu'un locuteur pouvait bien vouloir dire avec son énoncé en se basant sur le contexte dans lequel celui-ci était prononcé (cf. l'école ou le restaurant), et on a commencé à concevoir la compréhension comme quelque chose qui a à faire avec ce qu'un deuxième participant *fait* en réponse à ce que le premier a dit. La compréhension n'a donc plus été traitée comme un procès mental, mais elle a été mise en relation avec l'*action successive* accomplie par le co-participant.

Cette *succession* a été mieux précisée par l'Analyse Conversationnelle (AC : Sacks *et al*. 1974, Traverso 1999), qui a contribué à élargir le domaine d'enquête traditionnel de la linguistique et fourni une théorie du contexte qui lui faisait défaut.

> Car si le rôle de la dépendance contextuelle des énoncés a fini par être reconnu par la plupart des théories linguistiques qui s'intéressaient au sens des énoncés, ces dernières manquaient d'une conception du contexte qui soit accessible à l'enquête empirique. (De Fornel / Léon 2000: 145)

Cela a été possible grâce à l'idée de *contexte séquentiel*, selon laquelle l'interprétation d'énoncés elliptiques tels que "les poissons levez la main" dépendrait de leur collocation dans des séquences d'actions où l'action accomplie par un premier locuteur *projette* une action adéquate de la part de l'interlocuteur et où la compréhension est naturellement montrée par le fait que les participants continuent la conversation et par la manière choisie. En observant les *(ré)actions* des interlocuteurs aux tours de parole qui précèdent (dans l'exemple susmentionné, le fait que quelqu'un lève la main), l'AC explique donc comment les participants interprètent les actions de leurs pairs (par ex. comme un ordre adressé à certains) et comment ils aboutissent à une co-construction de l'interaction et à une co-compréhension de ce qui se passe (par ex. vérifier la commande avant de la passer).

Cette même *séquentialité*, qui est particulièrement visible dans des couples adjacents de tours comme Question-Réponse ou Invitation-Acceptation, existe aussi dans les entretiens cliniques en présence d'un interprète. Mais entre les *actions et (ré)actions* des interlocuteurs primaires (médecin / patient) il y a la compréhension de l'interprète de dialogue : tout en devant idéalement limiter son

intrusion dans la conversation, ce dernier ne peut pas se passer de donner sa propre interprétation de la parole à traduire, ce qui ouvre un débat sur la place de l'interprète dans l'entretien.

Avec sa distinction théorique entre *traduction* et *coordination*, Wadensjö (1998) a été la première à montrer que cette place n'est pas uniquement celle d'un traducteur qui répète, dans une autre langue, ce qui a été dit par autrui. C'est également celle d'un participant à part entière qui, en vertu de son bilinguisme, coordonne l'échange entre deux ou plusieurs personnes qui ne partagent pas la même langue, jouant ainsi un double rôle dans la construction de la compréhension (Wadensjö 2002). Depuis les études de cette chercheuse suédoise, un intérêt croissant a été accordé à la place des interprètes de dialogue dans l'interaction (Bolden 2000, Davidson 2000 et 2002, Traverso 2003, Biagini 2011, Merlino en cours de publication) et à la façon dont ils exercent leur fonction de coordinateurs de l'échange (Baraldi / Gavioli 2012).

Les pages qui suivent voudraient apporter une contribution à ce vif débat en présentant les quelques résultats de nos analyses conversationnelles. Nous montrerons :

- comment l'analyse d'entretiens simulés peut aider à comprendre quelle est la place de étudiant-interprète dans l'interaction ;
- comment l'analyse d'entretiens réels peut montrer quelle est la place de l'interprète-professionnel dans l'interaction ;
- comment les deux peuvent être combinées pour réfléchir à des parcours de formation plus adéquats.

Nous passerons enfin de la réflexion à la pratique, en proposant un jeu de rôle qui pourrait mieux préparer les étudiants à prendre leur place dans les interactions réelles.

3.1. Données didactiques

L'enseignement de l'interprétation de dialogue a été fortement influencé par la littérature sur l'interprétation de conférence (cf. Gile 1995, Falbo *et al.* 1999, Seleskovitch / Lederer 2002). Traditionnellement divisée entre textes à caractère descriptif et

prescriptif, cette littérature est diversement présente dans les programmes des cours universitaires (Benelli / Raccanello 2004), qui tout en étant profondément différents semblent d'accord sur un point : on apprend l'interprétation *en faisant* de l'interprétation. En effet, les cours et les examens d'interprétation de conférence se structurent pour la plupart comme des simulations de simultanées et de consécutives où l'étudiant-interprète met en œuvre des stratégies qui vont vraisemblablement se révéler utiles dans son futur professionnel.

Ce qui fonctionne depuis plusieurs décennies dans la formation des interprètes de conférence a été progressivement utilisé dans la formation des interprètes de dialogue. Toujours partant de l'idée que l'on apprend *en faisant*, la didactique de l'interprétation de dialogue se base sur l'utilisation, aussi bien en classe qu'en salle d'examen, de simulations d'interactions où l'étudiant *fait* l'interprète pour deux professeurs qui feignent de ne pas se comprendre. Tout en allant de pair avec d'autres modes d'enseignement et d'évaluation[3], ces simulations représentent le "noyau dur" des unités didactiques (Ballardini 2006 : 66). C'est justement en raison de leur fréquence d'utilisation, ainsi que de leur apparente proximité à la réalité qu'elles mettent en scène, que nous avons choisi les simulations en contexte d'examen comme premier objet d'analyse[4].

Dans les données examinées, deux professeurs et un étudiant simulent un entretien clinique entre un médecin italophone et un patient francophone communiquant à travers un interprète bilingue. Cette simulation se base sur un scénario qui généralement s'inspire d'une situation réelle et qui présente tous les tours de parole des interlocuteurs primaires. Pour que le jeu se déroule comme prévu, l'étudiant-interprète doit donc traduire fidèlement tout ce qu'ils disent.

---

3 Dont les plus répandus semblent être la traduction à vue de et vers la langue étrangère et la discussion des rôles de l'interprète sur la base de la littérature (cf. par ex. Mack / Russo 2005, Rudvin 2006, Bancroft / Rubio-Fitzpatrick 2011).

4 Les raisons qui nous ont amené à choisir les simulations en salle d'examen plutôt que celles en salle de classe, que nous avions également enregistrées, sont détaillées dans Niemants (2012a : 83).

Le corpus que nous avons sélectionné pour l'analyse se compose de 12 interactions à durée variable, pour un total d'à peu-près 3 heures d'enregistrement.

Interactions simulées						
N°	*Nom fichier*	*Sexe*	*Niveau*	*Autre langue*	*Durée*	*Sujet*
1ISS	090527_010	F	1ère Langue	Russe	13.54 minutes	Abdomen A
2ISS	090527_003	F	1ère Langue	Anglais	14.37 minutes	Cyanose
3ISS	090527_004	F	1ère Langue	Anglais	16.54 minutes	Céphalée
4ISS	090610_000	F	1ère Langue	Espagnol	14.29 minutes	Jambes enflées
5ISS	090610_003	F	1ère Langue	Anglais	14.38 minutes	Toux
6ISS	090527_005	F	1ère Langue	Espagnol	20.45 minutes	Palpitations
7ISS	090527_008	F	1ère Langue	Russe	17.11 minutes	Dyspnée
8ISS	090527_012	F	1ère Langue	Anglais	12.41 minutes	Évanouissement
9ISS	090529_003	F	2ème Langue	Anglais	17.14 minutes	Abdomen B
10ISS	090529_007	F	2ème Langue	Anglais	13.59 minutes	Hypertension
11ISS	090610_004	M	1ère Langue	Anglais	11.02 minutes	Céphalée
12ISS	090529_006	F	2ème Langue	Anglais	14.46 minutes	Toux

Tableau 1. Les entretiens cliniques simulés.

En observant ces 12 entretiens simulés avec les lentilles analytiques de l'AC, nous avons remarqué deux phénomènes récurrents, que nous allons résumer en présentant quelques exemples représentatifs.

Primo, les simulations se déroulent suivant le modèle "idéal" de l'interprétation de dialogue (cf. Wadensjö 1998 : 104). D'après ce modèle, appelé AIBIA, l'interprète (I) traduirait, tour-par-tour, les messages du faux médecin (A) et du faux patient (B) en utilisant la première (EX 1) ou la troisième (EX 2) personne.

(EX 1)

01 A mm hm (.) va bene lei fuma signora ?
02 I est-ce que vous fumez ?
03 B non je ne fume pas
04 I no non fumo

(EX 2)

01	A	signora lei lavora al computer ?
02	**I**	**est-ce que vous travaillez avec l'ordinateur?**
03	B	(1) ah non non non je ne travaille pas docteur
04	**I**	**non lavora**

Comme le montrent les exemples ci-dessus, l'interprétation équivaut à une activité de traduction d'énoncés isolés, qui sont reproduits sans en négocier la compréhension et sans interférer avec les "unités de traduction" (Martin 2009) proposées par A et B. L'étudiant ne fait généralement pas de demandes d'éclaircissement s'il n'a pas bien compris (préférant une mauvaise traduction à une négociation de sa compréhension) et il ne commence pas à parler avant que le tour de parole du locuteur qui précède n'ait atteint un "point de complétude" (Mondada 2007).

Secundo, les quelques éloignements du modèle AIBIA que nous avons repérés dans nos données didactiques montrent une orientation à l'évaluation de l'étudiant-interprète. Pas de quoi s'émerveiller, diront certains, étant donné le contexte d'examen dans lequel les simulations se déroulent. Et pourtant, quelque chose d'intéressant se dégage, nous semble-t-il, des trois exemples qui suivent.

Dans l'EX 3, A prend la parole immédiatement après la traduction, dans une place qui serait *idéalement* de B, afin d'énoncer d'autres unités à traduire. Ses tours commencent par un *mm hm* qui revient fréquemment dans nos données didactiques et qui paraît correspondre à une évaluation positive de ce qui vient d'être traduit.

(EX 3)

01	A	signora respiri profondamente (.) a bocca aperta
02	I	ehm respirez profondément la bouche ouverte
03	**A**	**(.) mm hm (.) adesso respiri solo con il naso**
04	I	maintenant seulement par le nez
05	**A**	**mm hm (.) tossisca**
06	I	toussez
07	**A**	**mm hm (.) apra la bocca**
08	I	ouvrez la bouche

Par contre, dans l'EX 4, A occupe la place qui serait *idéalement* de B pour corriger la *réaction traductive* de l'étudiant, qu'il considère comme erronée.

```
(EX 4)
      01  A  mm hm va bene ehm senta e (2) ha la febbre ?
      02  I  (.) est-ce que vous avez (.) vous êtes fièvreux ?
      03     (.)
      04  A  [euse]
      05  B  [ehm ] je
      06     (.)
      07  I  [fièvreuse fièvreuse]
      08  A  [fièvreuse è una donna]
      09     (1)
```

Encore, dans l'EX 5, où c'est B qui s'éloigne du modèle pour adresser une question à l'interprète une fois la simulation terminée (cf. le *va bene* conclusif de A au tour 03), A se réinsère, en chevauchant l'interprète, pour la stimuler à traduire (*cosa dice ?* aux tours 06 et 08).

```
(EX 5)
      01  A  mm hm l'accompagnerà l'infermiera e anche
                     l'interprete
      02  I  l'infirmière et moi aussi ehm nous allons vous
                     accompagner
      03  A  va bene
      04  B  ah vous faites partie du personnel de service ?
      05  I  [ehm désormais] {ride}
      06  A  [cosa dice] ?
      07  I  [{esala}]
      08  A  [cosa dice] ?
      09  I  mi chiede se faccio parte del ehm
      10  A  ah ora capisco
      11  I  dello [staff ehm medico]
      12  A        [dello staff ah ormai]
```

A un premier regard, la question répétée par A paraît renforcer la responsabilité communicative de l'interprète bilingue : le médecin admet ne pas comprendre ses interlocuteurs et avoir donc besoin de

l'interprète pour communiquer. Mais si on y regarde de plus près, c'est le contraire : tout en étant, dans la fiction du jeu de rôle, le seul locuteur bilingue, l'interprète n'a pas la responsabilité de la communication. C'est A, le faux médecin, le véritable *metteur en scène* de la conversation (cf. Orletti 2000) : c'est lui qui contrôle l'allocation des tours et c'est lui qui coordonne l'interaction en bloquant toute initiative qui s'éloigne de la traduction tour-par-tour et en décidant ainsi de la place de l'étudiant-interprète dans l'entretien.

En l'absence du cadre du jeu, dans les données réelles, l'interprète-professionnel occupe une place bien différente.

3.2. Données réelles

Le corpus de données réelles que nous avons sélectionnées pour l'analyse se compose de 12 entretiens cliniques à durée variable, pour un total d'à peu près 3 heures d'interaction en présence d'un intermédiaire linguistique professionnel[5].

Interactions réelles						
N°	*Nom fichier*	*Sexe*	*ID interprète*	*Autre langue*	*Durée*	*Suiet*
1BCH	100531_000	F	INT-B1	-	32.15 minutes	Infection
2BCH	100512_000	F	INT-B1	-	14.44 minutes	Orthopédie
3BCH	100330_000	F	INT-B1	-	17.34 minutes	Logement
4BCH	100315_000	F	INT-B2	-	7.47 minutes	Carte Bleue
5BCH	100322_000	F	INT-B2	-	3.35 minutes	Données
6BCH	100420_000	F	INT-B2	-	4.17 minutes	Prise de sang
7IREa **7IREb**	A0020313 + A0090330	F	INT-I1	Arabe	15.00 minutes 6.11 minutes	Problème aux yeux (A et B)
8IRE	A0030322	F	INT-I1	Arabe	42.30 minutes	Grossesse
9IFC	D0091019	F	INT-I2	Arabe	3.32 minutes	Écographie
10IFCa **10IFCb**	D0111021 + D0121021	F	INT-I2	Arabe	26.58 minutes 5.33 minutes	Grossesse
11IRN	02 Francese	F	INT-I3	Allem., Russe	8.30 minutes	Péricardite
12IRN	05 Francese	F	INT-I3	Allem., Russe	3.03 minutes	Péricardite

Table 2. Les entretiens cliniques réels.

5 Nous qualifions de "professionnel" toute personne préposée *à* et payée *pour* un service d'interprétation de dialogue, et cela indépendamment de sa formation (qui peut être universitaire, comme dans le cas de l'INT-I3, ou autre, comme dans les autres cas).

En faisant une AC de ces 12 entretiens, nous avons observé que le modèle *idéalisé* d'alternance des tours est tellement peu reconnaissable que l'on pourrait même douter de son applicabilité à la vie réelle. En analysant tous les mouvements d'éloignement du modèle AIBIA nous avons remarqué que :

- L'activité de traduction s'insère dans un mécanisme de négociation de la compréhension entre les interlocuteurs ;
- La traduction tour-par-tour n'assure pas nécessairement la compréhension ;
- L'interprète bilingue peut promouvoir la compréhension et donc la participation des interlocuteurs puisque c'est lui le *metteur en scène* qui coordonne l'entretien et qui exerce la responsabilité de la communication ;
- La traduction s'entrelace avec d'autres activités qui construisent ce qu'il y a à dire, à faire et donc à traduire.

Ces quatre remarques peuvent être faites à partir de l'extrait suivant, que nous proposons à titre d'exemple. Le A correspond toujours au médecin, le B au patient et le I à l'interprète.

```
(EX 6)
      01  A  menarca ?
      02  I  allora la la première fois que tu as eu la (.) les
                      règles ?
      03     (.)
      04  B  mm hm
      05  I  la première fois (.) à quelle età ?
      06     (.)
      07  I  onze ans douze ans treize ans ?
      08     (1)
      09  I  les règles (.) le sang ?
      10  B  ehm (.) en ce moment j' avais
      11     (2)
      12  I  allora lei [è sana nega patologie]
      13  B             [treize ans comme ça]
      14  I  treize ?
      15  B  mm [hm]
      16  I     [tredici] più o meno
      17  A  ok tredici anni
```

Dans les premiers tours de cet extrait, la traduction tour-par-tour de I n'est pas suivie de la réponse de la patiente, qui signale sa réception (cf. *mm hm*) mais garde le silence (cf. les pauses aux tours 03, 06, 08). I pose donc plusieurs questions de promotion de la narration afin que B réponde, enfin, à la question posée par le médecin au premier tour. Nous observons, en outre, que les activités de traduction et de promotion de la narration s'entrelacent avec une activité de "co-interviewing" (Davidson 2000). N'ayant pas encore obtenu la réponse qu'elle est censée traduire, I récupère une information qu'elle avait collectée auparavant, pendant l'interview du patient que le médecin lui-même l'avait précédemment déléguée à faire.

En analysant les entretiens réels nous avons également remarqué que l'idée sur laquelle se base l'enseignement de l'interprétation de dialogue, à savoir le fait que les deux interlocuteurs primaires ne se comprennent point, ne correspond pas à la réalité. En effet : le patient, qui vit depuis un certain temps dans le pays d'accueil, a une certaine compétence active et passive dans la langue des institutions ; et le médecin, qui a de plus en plus à faire avec des étrangers, a une certaine compétence, du moins passive, dans la langue du patient. Ces compétences se manifestent de deux manières : parfois les interlocuteurs primaires n'attendent pas la traduction de I et réagissent directement au tour de l'autre manifestant ainsi leur compréhension (cf. tour 09) ; d'autres fois ils prononcent des mots dans la langue de l'autre faisant de véritables commutations de code (cf. le signe §, aux tours 03 et 11).

(EX 7)

01	A	raffreddore ?
02	I	hm
03	**A**	**§ rhume et et lacrimation ?**
04	B	si si
05	I	hm
06		(.)
07	A	da quanto tempo ?
08		(.)
09	**B**	**ehm (.) § due giorno tre giorno**

(55 secondes omises)

```
10  I  tu as [tu as déjà fait ehm la visite] ?
11  B        [si j'ai fait § ho fatto visita]
12  A  e cosa t'han detto ?
13  B  hein
14     (.)
15  A  cosa ti hanno detto ?
16  I  qu'est-ce qu'il a dit le docteur ?
17  B  (.) ehm il dottor niente
18  A  [non hai non hai la risposta] ?
19  I  [pour les yeux pour les yeux le feuil] ? la
                réponse du docteur tu es ehm tu as
                porté la
```

Ces initiatives des interlocuteurs primaires ouvrent des scénarios inédits pour l'interprète de dialogue : du moment que sa place traditionnelle de traducteur n'est pas nécessaire, l'interprète doit négocier de nouveaux rôles dans l'interaction. Comme cela se dégage de l'extrait qui précède, un nouveau rôle paraît consister à *veiller* sur le déroulement de l'entretien et sur la compréhension des participants et à intervenir lorsqu'ils manifestent des problèmes de communication, soit parce qu'ils ont mal compris, soit parce qu'ils ne parviennent pas à s'exprimer ou à répondre directement à l'autre. C'est par exemple le cas du tour 16, où I se réinsère dans l'interaction pour traduire la question formulée par le médecin au tour 12, puisque contrairement à ce qui s'était passé avant (tours 04 et 09), le patient ne répond pas de façon autonome.

3.3. Quelques réflexions sur la formation

Si l'on rapproche les résultats de nos analyses sur les sous-corpus didactique et réel, on s'aperçoit des limites pédagogiques de simulations où :

- Les étudiants n'exercent qu'une activité de traduction tour-par-tour, ce qui, comme nous l'avons montré dans les exemples, n'assure pas la compréhension ;
- Les étudiants ne s'exercent pas à (ré)agir autrement qu'en traduisant, alors qu'il peut y avoir des cas où la traduction n'est pas nécessaire ni fonctionnelle à la communication (cf. les

manifestations de plurilinguisme du médecin et du patient) et des cas où d'autres actions peuvent contribuer à la compréhension et participation des interlocuteurs (cf. les questions de promotion) ;

- Les étudiants ne s'exercent pas à la responsabilité de la communication entre médecin et patient puisque (a) les deux se comprennent parfaitement, même s'ils font semblant de ne pas se comprendre (cf. *cosa dice ?*) et (b) ce n'est pas l'étudiant qui coordonne l'interaction bilingue.

Nous proposons donc une approche didactique partiellement différente, où l'activité du jeu de rôle se base sur notre corpus d'entretiens réels et où les interprètes s'entraînent à réagir aux tours de parole de véritables médecins et patients, réfléchissant à la responsabilité communicative qu'ils exerceront dans leur profession future.

4. Une proposition d'utilisation pédagogique

Dans ce qui suit, nous allons montrer comment un corpus construit pour des fins de recherche telles que les nôtres pourrait être réutilisé pour des fins pédagogiques. Ce que nous proposons a été imaginé *pour* et testé *dans* des classes d'interprétation de dialogue, mais cela n'empêche pas de l'appliquer à d'autres contextes didactiques, dont l'enseignement des langues étrangères et secondes nous semble, à présent, le premier candidat.

4.1. Un jeu-de-rôle particulier

Le type d'activité que nous présentons s'inspire du *Conversation Analytic Role Play Method* ou CARM (Stokoe 2011a). Cette méthode a été mise au point pour former des professionnels pour qui l'interaction orale revêt une importance considérable. Elle consiste à

montrer aux professionnels la vidéo d'une interaction réelle et à la bloquer à un tour particulier en leur demandant de :

- réfléchir aux problèmes qui se sont éventuellement générés dans l'interaction jusqu'à ce moment-là ;
- produire des *actions successives* au tour bloqué.

L'utilisation de cette méthode avec, par exemple, les médiateurs de conflits a montré que même ceux qui avaient une certaine expérience professionnelle avaient des difficultés à prévoir et à fournir les *actions successives* de leurs collègues. Cela nous semble avoir des implications pour la didactique qui se base sur le jeu de rôle traditionnel, car si l'on ne sait pas ce qui se passe dans une interaction réelle, comment peut-on la simuler de façon authentique ? (cf. Stokoe 2011b : 5). Si cela est vrai pour des médiateurs qui exercent déjà la profession, cela est d'autant plus vrai pour des étudiants-interprètes qui doivent passer de la formation à une profession qu'ils n'ont généralement pas encore exercée. C'est pourquoi nous proposons une simulation particulière, qui puisse les aider à prendre leur place dans l'interaction et à passer ainsi du jeu à l'exercice de leur(s) rôle(s).

Cette activité consiste à faire écouter / lire[6] aux étudiants quelques tours d'un entretien réel en présence d'un interprète et à bloquer la reproduction à un certain point, qui sera indiqué ci-dessous par un astérisque, en leur demandant de produire l'*action successive*. L'on montre ensuite aux étudiants comment a (ré)agit l'interprète professionnel et l'on réfléchit aux conséquences de cette (ré)action pour le déroulement de l'entretien.

Le choix des extraits à utiliser dépend de l'approche du formateur, qui peut être qualitative et / ou quantitative : dans le premier cas, il choisira des exemples qu'il considère comme qualitativement intéressants ; dans le second, il choisira des exemples qui reviennent fréquemment dans le corpus. Pour ce qui est des deux extraits que nous présentons à titre d'exemple, et que nous avons

6 Ne disposant pas de données vidéo, nous nous sommes limité à montrer la transcription des tours en même temps que l'enregistrement audio correspondant.

testés dans nos cours, la sélection a été informée par notre recherche qualitative.

Le premier extrait est déjà connu du lecteur : il s'agit du début d'un entretien entre un médecin et une femme enceinte qui doit répondre à des questions concernant son histoire médicale. L'une de ces questions porte sur la date de ses premières règles, et le médecin l'introduit de la façon suivante :

(EX 8)

01	I	ok niente
02		(.)
03	A	dimmi
04	I	no no parlava del vomito
05	A	ah
06	I	però è una cosa del
07*	**A**	**menarca ?**

Après avoir fait lire / écouter aux étudiants ces premiers tours, nous leur avons demandé de *(ré)agir* à la première partie du couple adjacent Question-Réponse qui est produite par le médecin au tour 07. Nous avons ensuite comparé leurs *actions successives* avec celles de l'interprète professionnelle qui, comme nous l'avons déjà montré dans l'EX 6, tout d'abord réagit avec une traduction tour-par-tour et puis avec une série de questions qui stimulent la patiente à produire la réponse projetée par le médecin.

08	**I**	**allora la la première fois que tu as eu la (.) les règles ?**
09		(.)
10	B	mm hm
11	**I**	**la première fois (.) à quelle età** ?
12		(.)
13	**I**	**onze ans douze ans treize ans ?**
14		(1)
15	**I**	**les règles (.) le sang ?**
16	B	ehm (.) en ce moment j' avais
17		(2)
18	**I**	**allora lei [è sana nega patologie]**
19	B	[treize ans comme ça]
20	**I**	**treize ?**

```
21  B  mm [hm]
22  I     [tredici] più o meno
23  A  ok tredici anni
```

L'utilisation de cet extrait permet aux étudiants de réfléchir sur : (1) la construction des tours en fonction de l'interlocuteur, ce que Sacks *et al*. (1974 : 727) appellent *recipient design* (cf. la simplification terminologique de "règles" à "sang") ; (2) les pauses à valeur de tour comme possibles signaux de non-compréhension demandant l'intervention du responsable de la communication (cf. l'interprète qui réagit en reformulant quatre fois, de manière de plus en plus simple, la question du médecin) ; (3) l'intersection de plusieurs activités (cf. la traduction, la promotion de la narration et le *co-interviewing*). Cet extrait permet en outre de réfléchir sur une difficulté qui semble caractériser les patients provenant de certaines régions africaines : celle de placer les événements dans le temps et, de ce fait, de fournir les réponses précises que les médecins occidentaux souhaitent (cf. Mazzetti 2003). L'extrait peut donc servir de point de départ pour une discussion plus ample, ainsi que d'incitatif à la lecture d'approfondissement.

Le deuxième extrait qui se prête au CARM est tiré d'une longue interaction entre un (autre) médecin et une (autre) femme enceinte. Comme cela se dégage des premiers tours que nous avons montrés aux étudiants, tout en ayant un permis de séjour régulier, cette femme n'a pas de médecin dans la ville de Castelfranco[7]. Le médecin lui dit donc de se rendre dans un bureau qui s'appelle *Saub*.

```
(EX 9)
      01  I  tu as le permis de séjour ici
      02  B  mm hm
      03  I  sì lei è regolare [vedi]
      04  A                    [ah]
      05  I  solo che non ha il medico qua a Castelfranco
      06     (.)
      07* A  allora le diciamo di andare al Saub è
```

7 Tous les noms propres ont été changés pour des raisons de respect de l'intimité de la vie privée des participants.

Après avoir demandé aux étudiants de *(ré)agir* au tour 07, nous avons de nouveau comparé leurs *actions successives* avec celles de l'interprète professionnelle, qui tout d'abord négocie avec le médecin la validité du permis de séjour de la patiente et qui produit ensuite une traduction "élargie" du tour original (cf. les "expanded renditions" de Wadensjö 1998), précisant la rue du bureau auquel la patiente devra se rendre ainsi que la raison pour ce faire.

08 I (.) allora aspetta (.) sto guardando qua (1) valido fino al dodici dieci duemiladieci
09 (.)
10 A quindi va bene quest'anno ci rientra (1) dodici dieci ottobre
11 I hm parce que tu dois aller (1) à (1) al Saub (.) via Mandorla (1) pour choisir le médecin de base
12 (1)
13 I ok ?
14 B hm

L'action de l'interprète au tour 11 montre encore une fois son orientation à la construction du tour pour l'interlocuteur et elle invite les étudiants à réfléchir sur les conséquences de ce type de *(ré)action traductive* pour la compréhension de la patiente, qui est d'ailleurs immédiatement vérifiée par un *ok ?*. Dans un souci de clarté, l'interprète va jusqu'à profiter d'un silence de 3 secondes qui se crée un peu plus tard dans l'interaction, pour s'assurer que la patiente comprenne comment arriver à ce bureau. La continuation de l'extrait montre en effet que l'interprète donne à la patiente toutes les indications (verbales et écrites) nécessaires pour s'y rendre seule, en construisant avec elle une séquence d'instructions :

15 (3)
16 I tu demandes à quel à quelqu'un qui (1) qui connaît bien ehm Castelfranco de t'accompagner là-bas
17 B mm hm

```
18  I  (.) hein (.) c'est pas lointain c'est c'est via
                    Flaminia (1) ehm avec le l'autobus num
                    numéro deux (1) et tu vas là Saub Sob
19     (.)
20  B  [Saub]
21  I  [scri] écris si c'est écrire ça ehm allora (2) je te
                    l'écris ici (.) via Mandorla (.) ehm
                    autobus (.) numéro deux (.) pour aller
                    (.) ça c'est l'autobus qui porte à
                    Modena
22  B  mm hm
23  I  de la direction de Re de Reggio
24  B  hm
25  I  pour (.) aller (.) à (.) Reggio je je le mets ici
                    (.) ehm (1) ok donc (.) Saub (1) pour
                    la (.) tessera ehm si pour la tessera
                    (.) sanitaria (3) choisir (2) le (.)
                    médecin (1) de base (.) ou de famille
                    (.) ok (.) donc tu vas tu tu peux aller
                    toute seule (.) aussi sur l'autobus
                    c'est pas difficile tu fai tu fais voir
                    cette cette adresse là
26  B  mm hm
27  I  et te peut il te porte le chauffeur il te porte ehm
                    là-bas (.) ehm puis c'est pas difficile
                    (.) tu entres tu tu demandes là-bas (.)
                    ok ?
28  B  si (.) ok
```

La présentation de cet exemple amorce une réflexion sur ce que *être responsable de la communication* peut effectivement signifier. Les étudiants observent que dans *la vie réelle* la compréhension entre les interlocuteurs n'est pas assurée par la traduction tour-par-tour et que les éloignements du modèle AIBIA peuvent justement servir à les mettre dans les conditions de se comprendre (cf. l'*ok ?* qui sert à vérifier leur compréhension ou la traduction élargie et les indications qui servent à la promouvoir). La dernière partie de l'exemple permet en outre de sensibiliser les étudiants aux signaux qui peuvent manifester la compréhension ou la non-compréhension des participants (cf. la présence de *mm hm* ou de pauses à valeur de tour), ce qui peut s'avérer utile dans des milieux où les interlocuteurs

primaires ont une certaine compétence active et passive dans la langue de l'autre et où les interprètes pourraient donc être appelés à jouer un nouveau rôle de *vigilants* de la communication (cf. 2.2 ci-dessus).

En somme, l'activité que nous proposons nous semble assurer une plus grande "authenticité" de la simulation (Ellis 2003), puisque celle-ci reproduit les mêmes situations et les mêmes formats d'interaction que les entretiens réels. Et si l'on en croit Widdowson, pour qui l'authenticité des activités didactiques réside plus dans la réaction de l'étudiant que dans la reproduction de la réalité (1978 : 90), ce type d'activité nous semble offrir au futur interprète l'opportunité de *(ré)agir* plus authentiquement aux tours de vrais médecins et patients.

5. En guise de conclusion

Les dimensions de nos sous-corpus ainsi que le type d'analyse que nous voulions en faire nous ont dirigée vers une approche qualitative. Cela n'empêche que les résultats et les premières applications de notre étude peuvent servir de "rampe de lancement" pour des recherches quantitatives (cf. Straniero Sergio / Falbo 2012 : 36).

Prenons le cas des commutations de code, qui peuvent être aisément comptées si elles ont été préalablement annotées (cf. Niemants 2012b). Il serait sans doute souhaitable de vérifier la fréquence de ce phénomène dans une grande banque de données afin de justifier des changements dans la formation des étudiants aux multiples rôles de l'interprète de dialogue. Si, comme nous avons raison de le croire (cf. Baraldi / Gavioli 2012), ce phénomène revenait dans d'autres sous-corpus d'entretiens cliniques, et si ceux-ci était transcrits avec EXMARaLDA ou d'autres logiciels compatibles, on pourrait utiliser EXAKT pour localiser des séquences quantitativement intéressantes à jouer avec le CARM. Une fois les transcriptions achevées, cet outil permet de visualiser les concordances et, en double cliquant sur l'une d'entre elles, de lire et

d'écouter la séquence dans laquelle elle se trouve, soit en notation musicale, comme dans la figure ci-dessous, soit tour-par-tour.

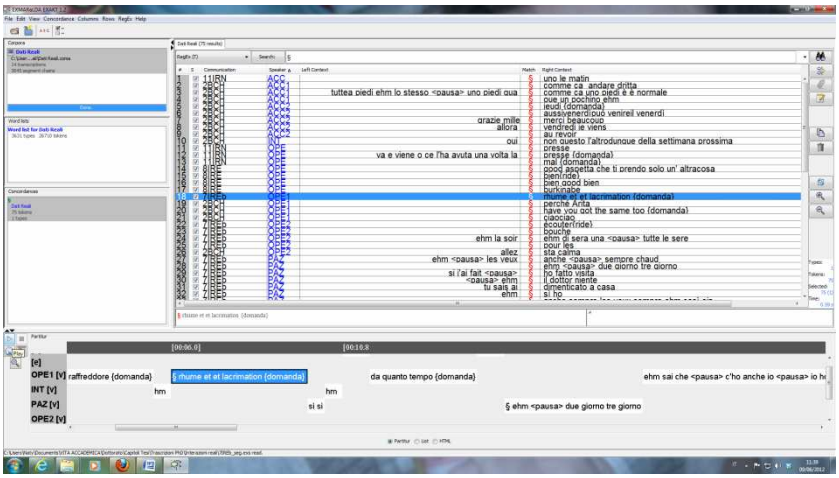


Figure 4. Les commutations de code : symbole-clé en contexte.

Voilà pourquoi nous avons construit un sous-corpus qui se prête à un partage global et à des enquêtes quantitatives, dont deux exemples ultérieurs pourraient être : (1) l'analyse des tours régulateurs comme *mm hm*, qui paraissent jouer un rôle dans la transition entre la parole à traduire et la parole traduite ; (2) l'analyse des *wordlists*, qui dans notre cas témoignent de la fréquence de termes qui ont à faire avec la grossesse, une phase particulièrement délicate dans la vie des familles immigrées, à laquelle les autorités de santé prêtent beaucoup d'attention (Baraldi *et al*. 2008) et à laquelle le futur interprète devrait probablement se préparer.

Nous n'avons mentionné que trois exemples de possibles enquêtes quantitatives. Or, une approche didactique basée sur les corpus pourrait s'avérer "décloisonnante" (Hédiard 2007), ouvrant maintes pistes de réflexion et montrant les multiples réalités de l'interprétation de dialogue en milieu médical. Il faudra peut-être attendre quelques années, puisque le développement des études informatisées d'interprétation est ralenti par la difficulté de transcrire

et de comparer les données orales (Straniero Sergio / Falbo 2012), mais un jour les corpus rendront accessibles des faits dont on ignorait l'existence. Et une fois que nous saurons comment se déroule l'interprétation *réelle*, on n'aura plus d'excuses pour ne pas l'enseigner (cf. Widdowson 1996). La question de l'authenticité se posera alors avec toute sa force, car le fait d'utiliser du *matériel* authentique n'assure pas l'authenticité des *activités* qu'on propose.

Références bibliographiques

Ballardini, Elio 2006. Pour un enseignement universitaire de l'interprétation en milieu médical. *Etudes de Linguistique Appliquée* 141, 43-50.

Bancroft, Marjorie / Rubio-Fitzpatrick, Lourdes 2011. *The community interpreter: a comprehensive training manual* (5th edition). Columbia : Cross Cultural Communications.

Baraldi, Claudio / Barbieri, Viola / Giarelli, Guido (eds.) 2008. *Immigrazione, mediazione culturale e salute*. Milano : Franco Angeli.

Baraldi, Claudio / Gavioli, Laura (eds.) 2012. *Coordinating participation in dialogue interpreting*. Amsterdam / Philadelphia: John Benjamins.

Benelli, Graziano / Raccanello, Manuela 2004. Relazione sull'insegnamento delle traduzione e dell'interpretazione da e verso il francese. In Cigada, Sergio / Galazzi, Enrica (eds.) *L'insegnamento del francese nell'università italiana*. Brescia : Editrice La Scuola, 45-58.

Biagini, Marta 2011. Quelles questions pour quelles réponses? L'interprète de dialogue dans l'interrogatoire au tribunal. In Dufiet, Jean-Paul *et al*. (eds.) *Cahiers de recherche de l'école doctorale en linguistique française*. Milano : Lampi di stampa, 93-106.

Bolden, Galina 2000. Toward understanding practices of medical interpreting: Interpreters' involvement in history taking. *Discourse Studies* 2 / 4, 387-419.

Davidson, Brad 2000. The interpreter as institutional gatekeeper: The social-linguistic role of interpreters in Spanish-English medical discourse. *Journal of Sociolinguistics* 4 / 3, 378-405.

Davidson, Brad 2002. A model for the construction of conversational common ground in interpreted discourse". *Journal of Pragmatics* 34, 1273-1300.

De Fornel, Michel / Léon, Jaqueline 2000. L'analyse de conversation. De l'ethnométhodologie à la linguistique interactionnelle. *Histoire Epistémologie Langage* 22 / 1, 131-155.

Ellis, Rod 2003. *Task-based language learning and teaching*. Oxford : Oxford University Press.

Falbo, Caterina / Russo, Mariachiara / Straniero Sergio, Francesco (eds.) 1999. *L'interpretazione simultanea e consecutiva. Problemi teorici e metodologie didattiche*. Milano: Hoepli.

Galazzi, Enrica 2002. *Le son à l'école*. Brescia : La Scuola.

Gile, Daniel 1995. *Basic concepts and models for interpreter and translator training*. Amsterdam / Philadelphia: John Benjamins.

Hédiard, Marie 2007. L'analyse guidée sur corpus : Une approche « décloisonnante » pour accéder au sens. In Bellati, Giovanna *et. al* (eds.) *Un paysage choisi. Mélanges de linguistique française offerts à Leo Schena*, Torino : L'Harmattan, 225-238

Mack, Gabi / Russo, Mariachiara (eds.) 2005. *Interpretazione di trattativa : La mediazione linguistico-culturale nel contesto formativo e professionale*. Milano : Hoepli.

Martin, Philippe 2009. Grand corpus oraux : Classification, transcription, exploitation des données. In Paissa, Paola / Biagini, Marta (eds.) *Cahiers de recherche de l'école doctorale en linguistique française*. Milano : Lampi di stampa, 121-134.

Mazzetti, Marco 2003. *Il dialogo transculturale*. Roma : Carocci.

Merlino, Sara forthcoming. Traduction orale et organisation de la parole: La gestion multimodale des transitions. In Mondada, Lorenza (ed.) *Corps en interaction. Participation, spatialité, mobilité*. Lyon : ENS éditions.

Mondada, Lorenza 2007. L'interprétation online par les co-participants de la structuration du tour in fieri en TCUs : Evidences multimodales. *TRANEL* 48, 7-38.

Mondada, Lorenza 2011. Understanding as an embodied, situated and sequential achievement in interaction. *Journal of Pragmatics* 43, 542-552.

Niemants, Natacha S.A 2012a. *Traduzione e mediazione nell'interpretazione dialogica in ambito sanitario: Ruolo o responsabilità? Una risposta interazionista*. Tesi di dottorato non ancora pubblicata, Università di Modena e Reggio Emilia. http://www.dailynterpreter.com/phd-thesis.pdf

Niemants, Natacha S.A 2012b. The transcription of interpreting data. *Interpreting* 14 / 2, 165-191.

Orletti, Franca 2000. *La conversazione diseguale*. Roma : Carocci.

Rudvin, Mette 2006. Issues of culture and language in the training of language mediators for public services in Bologna : Matching market needs and training. In Londei, Danielle / Miller, Donna Rose / Puccini, Paola (eds.) *Insegnare le lingue / culture oggi : il contributo dell'interdisciplinarietà*. Bologna : Asterisco, 57-72.

Sacks, Harvey / Schegloff, Emmanuel / Jefferson, Gail 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696-735.

Schmidt, Thomas 2004. Transcribing and annotating spoken language with EXMARaLDA. *Proceedings of the LREC-Workshop on XML based richly annotated corpora*, Lisbon 2004, Paris : ELRA.

Seleskovitch, Danika / Lederer, Marianne 2002. *Pédagogie raisonnée de l'interprétation*. Paris : Didier Erudition/ Knlincksieck.

Stokoe, Elisabeth 2011a. Simulated interaction and communication skills training: The 'Conversation Analytic Role-play Method'. In Antaki, Charles (ed) *Applied conversation analysis*. New York *et al*. : Palgrave Macmillan, 119-139.

Stokoe, Elisabeth 2011b. Overcoming roadblocks to mediation : Training mediators using the 'Conversation Analytic Role-play Method'. *Mediation Digest Online*.

Straniero, Sergio, Francesco / Falbo, Caterina 2012. Studying interpreting through corpora. An introduction. In Straniero Sergio, Francesco / Falbo, Caterina (eds.) *Breaking ground in corpus-based interpretino studies*. Frankfurt : Peter Lang, 9-52.

Traverso, Véronique 1999. *L'analyse des conversations*. Paris : Armand Colin.

Traverso, Véronique 2003. Rencontres interculturelles à l'hôpital : La consultation médicale avec interprète. *Tranel* 36, 81-100.

Wadensjö, Cecilia 1998. *Interpreting as Interaction*. London / New York : Longman.

Wadensjö, Cecilia 2002. The double role of a dialogue interpreter. In Pöchhacker, Franz / Shlesinger, Miriam (eds.) *The interpreting studies reader*. London : Routledge, 355-370.

Widdowson, Henry G. 1978. *Teaching language as communication*. Oxford : Oxford University Press.

Widdowson, Henry G. 1996 Comment : Authenticity and autonomy in ELT. *ELT Journal* 50 / 1, 67-68.

I corpora in contesti legali

MARCELLO SOFFRITTI

9. Konjunktiv in deutschsprachigen Gesetzbüchern

1. Einführung und Fragestellung

In der juristischen Kommunikation scheint der Konjunktiv erforderlich zu sein:

> Inzwischen ist es so, dass wir sagen, 50 Prozent unserer Studienanfänger beherrschen den Konjunktiv nicht mehr sicher, und das ist beispielsweise, wenn man juristische Sachverhalte darstellen will, ein Kernproblem, denn dann kann ich nicht mehr ausdrücken, was ist nun wahr und was wird bestritten, also was ist streitig zwischen den Parteien.[1]

In den Redaktionsrichtlinien für Gesetzgeber der Bundesrepublik Deutschland, der Schweiz und Österreichs ist jedoch keine direkte Empfehlung bezüglich des Konjunktivs enthalten[2]. Anhand einzelner Redaktionsnormen kann man gelegentlich vermuten, dass man den Konjunktiv als solchen gar nicht erst in Erwägung ziehen sollte, wie etwa beim Gebrauch von Modalverben in der Formulierung von Geboten, Verboten, Erlaubnissen und Ermächtigungen. Ansonsten ist der Konjunktiv in der Redaktion technischer Anleitungen generell unerwünscht (Übersicht in Höfler 2012: 142).

---

1 <http://www.dradio.de/dlf/sendungen/campus/2174610/> (Interview mit einer deutschen Rechtsprofessorin – letzter Abruf September 2013).

2 Für Deutschland: *Handbuch der Rechtsförmlichkeit* <http://hdr.bmj.de/page_b.1.index.html> (letzter Abruf September 2013); für Österreich: *Handbuch der Rechtssetzungsstechnik*: <http://www.bka.gv.at/DocView.axd?CobId=1656> (letzter Abruf September 2013). Für Südtirol sind noch keine offiziellen Richtlinien vorhanden. Für die Schweiz scheinen gegenwärtig unterschiedliche Ansätze vorzuliegen.

Offensichtlich ändern sich also die Forderungen nach Bestimmtheit, Genauigkeit und Differenziertheit je nachdem, ob man argumentiert, berichtet oder Normen erlässt. Wenn man sich auf Norm setzende Texte bezieht, sollen vereinzelte Sachverhalte und deren subjektive Darstellung vermieden werden, wie im folgenden Abschnitt gefordert:

> Rechtsvorschriften sind abstrakt zu formulieren. Kasuistische Regelungen sind zu vermeiden. Beispiele sind in einer Rechtsvorschrift nur dann anzuführen, wenn sie Begriffe verdeutlichen oder zur Konkretisierung von Generalklauseln beitragen. (Handbuch der Rechtssetzungsstechnik: 7)

Das dürfte in erster Linie die Wiedergabe von Behauptungen oder Meinungen, und somit Formen der Redewiedergabe im Konjunktiv I ausschließen. Die Hauptfrage dieses Beitrags ist jedoch, in welchen Kombinationen und Funktionen der Konjunktiv II, und insbesondere der Irrealis, in Gesetzbüchern tatsächlich verwendet wird. In Texten ergibt sich ein Irrealis generell aus der Nichterfüllung einer Wahrheitsbedingung. Bei der Formulierung gesetzlicher Normen wird aber meistens die reine Verkettung von Bedingungen und Folgen aufgezeigt, ohne jeden Spielraum für die etwaige Unwahrscheinlichkeit oder Irrealität einer Annahme. Und in der Tat sind Formulierungen wie: „wer X getan hätte, würde mit Z bestraft“, oder „falls Y einträfe, wäre Z die Folge“ in den rein normierenden Teilen von Absätzen oder Paragraphen nicht denkbar. Durch eine Präsupposition der Irrealität würden die gesetzlichen Normen unbrauchbar und müßig. Irrealis-Formen sind also eher in nicht-performativen Teilen von Gesetzestexten zu erwarten. Trifft diese Annahme zu, und unter welchen Umständen?

Bei der empirischen Untersuchung dieser Problematik sollen gleichzeitig Möglichkeiten und Grenzen einer korpusbasierten Analyse gesetzlicher Normen aufgezeigt werden. Es werden vorrangig folgende Fragen erörtert: Wie weit reichen die Ergebnisse, die man mithilfe einer gängigen Softwareanwendung aus diesen Korpora gewinnen kann? Wie viel „manueller“ Aufwand ist zusätzlich für Einzelauswertungen notwendig? Wie können die beiden Verfahren opti-

mal miteinander kombiniert werden? Welche Vor- und Nachteile haben die verwendeten Korpora?

2. Datenerhebung

Um eine brauchbare empirische Basis für die Untersuchung zu gewinnen, wurden jeweils das Bürgerliche (bzw. Zivilrechtliche) Gesetzbuch (BGB), die Strafprozessordnung (StPO) und die Zivilprozessordnung (ZPO) aus Deutschland, Österreich, der Schweiz und Südtirol in einem Gesamtkorpus zusammengeführt. Das Südtiroler Material ist bekanntlich eine Übersetzung (vgl. die gedruckten Fassungen in Bauer *et al*. 1989, 1991, 1996) des italienischen Originals, das bei strittiger Auslegung immer als entscheidende Referenz heranzuziehen ist. Dieses Korpus bildete bereits die Grundlage für einige Studien zur Verwendung von konditionalen Konnektoren (zuletzt Soffritti 2013). Es sind hier nochmals kurz die wichtigsten Implikationen und Feststellungen zu erwähnen, die sich aus dieser Wahl ergeben.

Gesetzbücher besitzen eine hohe Relevanz im jeweiligen juristischen und linguistischen Nationalraum. Es gilt insgesamt für die Relevanz des hier verwendeten Materials das, was in Soffritti (2009: 51) ausgeführt wurde. Da diese Gesetzbücher teilweise seit über 100 Jahren bestehen, enthalten sie Spuren eines veralteten Sprachgebrauchs, die gelegentlich auch in der Verwendung von Konjunktivformen (und in der entsprechenden syntaktischen Umgebung) erkennbar sind. Es ist insgesamt nicht angebracht, diese geschichtlich stratifizierten Gesetzbücher unter Berufung auf erst in den letzten Jahrzehnten entwickelte Vereinfachungsprinzipien systematisch zu kritisieren. Erst bei neueren Gesetzestexten wäre es sinnvoll, die Entsprechung zu den gegenwärtigen Formulierungsrichtlinien zu überprüfen und gegebenenfalls einzufordern. Die Frage ist umso interessanter, als Gesetzbücher einen grundlegenden Anhaltspunkt für die laufende Gesetzgebung bilden und stellenweise bis auf die (komplexe) Syntax in diese übernommen werden.

Gesetzbücher im deutschsprachigen Raum sind das Ergebnis nationaler Traditionen in der Strukturierung und Formulierung dieser Textsorte. Aus diesem Korpus sind also nicht nur allgemeine Schlüsse auf den Sprachgebrauch zu ziehen, sondern auch wichtige Erkenntnisse über die diatopische Variationsbreite der deutschen Gesetzessprache zu gewinnen. Zu den variierenden Erscheinungen könnte durchaus der Konjunktiv gehören, und bei der Analyse der Belege sollen eventuell auftretende Eigentümlichkeiten festgehalten werden. Das verwendete Gesamtkorpus, das insgesamt 1278458 Wörter (30853 Types) umfasst, setzt sich quantitativ wie folgt zusammen:

	KORPUS D	KORPUS A	KORPUS ST	KORPUS CH
TOKENS	347907	225480	532858	172213
TYPES	13342	13099	15633	9531

Abb. 1.

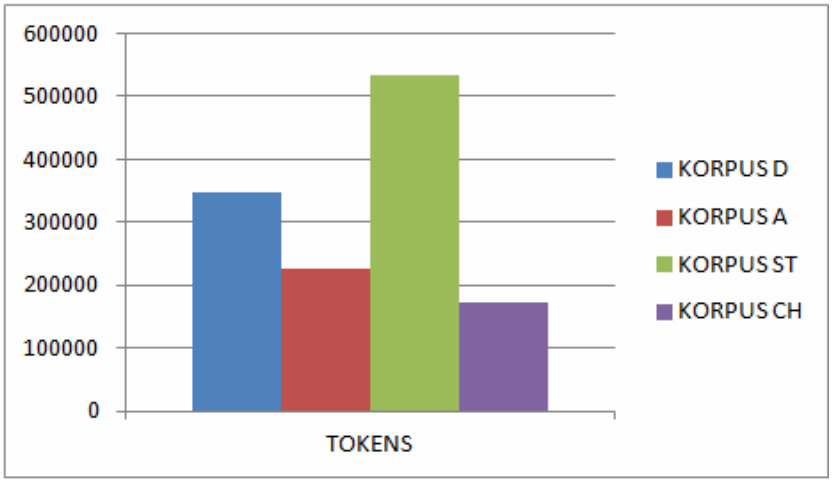


Abb. 2.

Da es sich um ein fachsprachliches Spezialkorpus mit gemeinsamer Thematik und einer einzigen Textform handelt, ist die Erfordernis der qualitativen Ausgewogenheit der Bestandteile unschwer erfüllt. Jedes Teilkorpus setzt sich, wie schon erwähnt, aus drei thematisch identi-

schen Gesetzbüchern zusammen, die jeweils das Zivilrecht, die zivilrechtliche und die strafrechtliche Prozessordnung regeln. Andererseits sind unter den einzelnen Korpora erhebliche quantitative Unterschiede festzustellen, weswegen die jeweils erhobenen quantitativen Werte nur prozentual verglichen werden. Es sind damit in wesentlichen Teilen auch die Voraussetzungen erfüllt, die Gries & Newman (i.D.) für eine sinnvolle Auswertung sprachlicher Daten anhand großer, nicht fachsprachlicher Korpora ausgearbeitet hat.

3. Der linguistische Untersuchungsgegenstand

Bei der Abfrage wurden nur folgende Konjunktivformen berücksichtigt: *hätte, wäre, würde, könnte, sollte, dürfte, müsste*. Andere Konjunktivformen wie *ginge* oder *liefe* wurden zwar probeweise in Betracht gezogen, aber die entsprechende Abfrage ergab keine Belege, so dass man davon ausgehen darf, dass diese Art von synthetischen Konjunktivformen im Korpus eine seltene Ausnahme sind.

Die Distribution der erhobenen Konjunktivformen im Gesamtkorpus sieht folgendermaßen aus:

	Anzahl	Frequenz
wäre(n)	356	0,027846
hätte(n)	214	0,016739
würde(n)	344	0,026907
könnte(n)	120	0,009386
sollte(n)	33	0,002581
müsste(n)	16	0,001252
dürfte(n)	9	0,000704
GESAMT	1092	0,085415

Abb. 3.

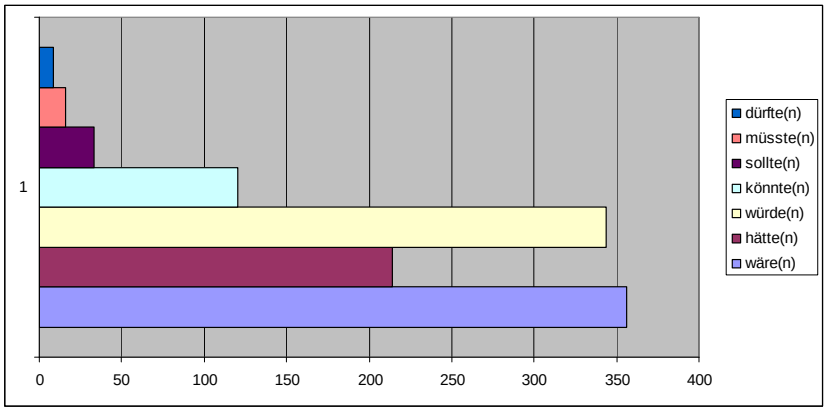


Abb. 4.

Es folgt die Gesamtaufstellung mit einem entsprechenden Diagramm. Die Tabelle enthält für jedes Teilkorpus die reine Anzahl der Belege von Konjunktivformen und die jeweiligen prozentualen Frequenzwerte (im Verhältnis zur Gesamtzahl der Wörter im entsprechenden Teilkorpus). Das Diagramm stellt die prozentualen Frequenzwerte für jedes Teilkorpus dreidimensional dar.

	D	D	A	A	ST	ST	CH	CH
wäre(n)	132	0,0379	127	0,0563	57	0,0107	40	0,0232
hätte(n)	40	0,0115	60	0,0266	96	0,0180	18	0,0105
würde(n)	198	0,0569	84	0,0373	30	0,0056	32	0,0186
könnte(n)	23	0,0066	42	0,0186	35	0,0066	20	0,0116
sollte(n)	8	0,0023	6	0,0027	12	0,0023	7	0,0041
müsste(n)	6	0,0017	5	0,0022	4	0,0008	1	0,0006
dürfte(n)	6	0,0017	3	0,0013	0	0,0000	0	0
GESAMT	413	0,1187	327	0,1450	234	0,0439	118	0,0685

Abb. 5.

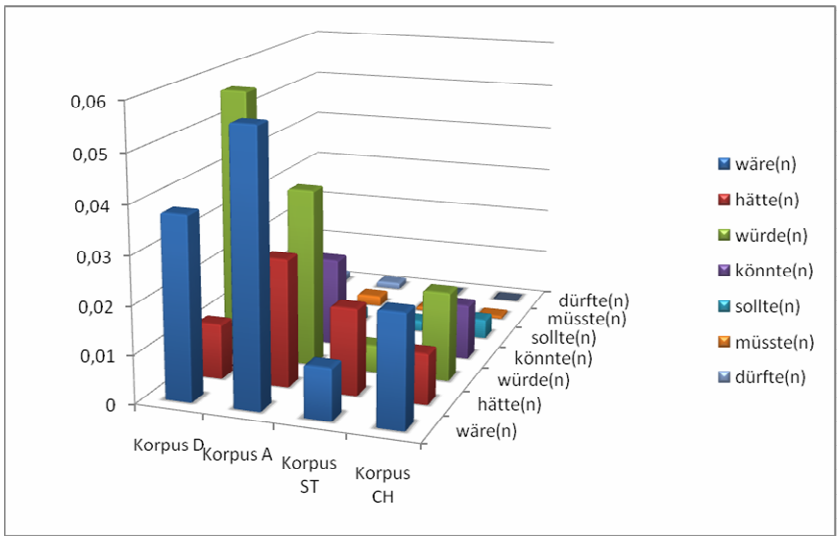


Abb. 6

Die Gesamtfrequenz dieser Formen mit einem Wert von 0,085415 ist an sich zuerst nichtssagend. Sie kann aber eine gewisse Bedeutung erhalten, wenn man einen vorläufigen Vergleich zwischen Indikativ- und Konjunktivformen durchführt. Aus naheliegenden Gründen muss man sich bei diesem Vergleich auf die reinen Singularformen (*wäre, hätte, könnte, sollte, müsste, dürfte*) beschränken, um sie mit *ist, hat, kann, soll, muss, darf* zu vergleichen. Dabei ergibt sich das folgende Bild: die neu berechnete Frequenz der so eingeschränkten 618 Konjunktivformen beträgt 0,048339, während diejenige der insgesamt 27705 Indikativformen den Wert 2,167064 erreicht. Das Verhältnis ist dabei etwa 1:37. Diese Werte kann man wiederum mit denjenigen vergleichen, die sich aus einem großen Korpus der deutschen Sprache (Corpus: deWaC, 1,7 Md. Wörter)[3] gewinnen lassen:

3 Frei zugänglich unter der Adresse: <http://nl.ijs.si/tei/teiHeaders/dewac-teiHeader-en.html> (letzter Abruf September 2013) . Für eine umfassendere Beschreibung dieser Art Korpora vgl. Baroni et al. (2008).

	Indikativ	Konjunktiv
Anzahl	21082713	2447591
Frequenz	1,24016	0,143976

Abb. 7.

In deWaC liegt also das Verhältnis bei 1:8,61, d.h. die Dichte dieser Konjunktivformen ist mehr als 4 mal höher als in den analysierten Gesetzbüchern.

Man kann also Folgendes festhalten: einerseits sind in Gesetzbüchern mehr Belege da, als man anhand prinzipieller Überlegungen über die Zulässigkeit des Konjunktivs erwartet hätte. Andererseits liegt deren Frequenz weit unter derjenigen, die man im allgemeinen schriftlichen Sprachgebrauch feststellen kann. Auch die Distribution der bisher berücksichtigten einzelnen Konjunktivformen stellt sich in den zwei Korpustypen anders dar.

4. Die Ermittlung relevanter Kontexte anhand der Cluster

Mit der verwendeten Software (AntConc 3.3.5) ist es möglich, für jede Konjunktivform zu prüfen, ob sie sich mit bevorzugten Begleitern in Clustern kombinieren, die ihrerseits auf syntaktische, semantische oder stilistische Tendenzen schließen lassen. Die Methodik der Cluster-basierten Korpusanalysen ist m. E. noch nicht vollständig ausgereift, so dass an dieser Stelle eine gewisse Flexibilität in der Einstellung der entsprechenden Parameter (links oder rechtsgerichtete Clusters, Mindestfrequenz, minimale und maximale Anzahl der Cluster-Komponenten) notwendig erscheint, um spezifische morphosyntaktische oder semantische Fragen zu erhellen. Um beispielsweise Wortstellungsfragen zu klären, genügt oft eine Einstellung mit 2 bis 3 links liegenden Elementen, die im Falle von *wäre* (bei minimaler Frequenz = 20) folgendes Bild ergibt:

Anzahl der Belege	Textsegment
32	gewesen wäre
32	worden wäre
29	erschwert wäre
27	wesentlich erschwert wäre
21	zuständig wäre
20	oder wesentlich erschwert wäre
20	unbillig wäre

Abb. 8.

Um signifikante Aspekte der sog. semantischen Prosodie zu erarbeiten und darauf gestützte textsemantisch Verallgemeinerungen vorschlagen zu können, ist es dagegen sinnvoller, eine höhere Anzahl von Clusterelementen einzustellen. Im Fall von *wäre* kann man dadurch semantische Information erhalten, die nicht schon im regierten Infinitiv oder Partizip steckt. Die Tabelle der 4 Elemente-Clusters für *wäre(n)*[4] sieht folgendermaßen aus:

Anzahl der Belege	Textsegment
20	oder wesentlich erschwert wäre
5	Ermittlungen sonst aussichtslos wären
4	ansonsten wesentlich erschwert wäre
4	der Ermittlungen gefährdet wäre
4	soweit es unbillig wäre
3	des Kindes gefährdet wäre
3	eine unbillige Härte wäre
3	so vorzugehen, als wäre
2	abs. 1 gefällt worden wäre
2	als gering anzusehen wäre
2	andernfalls wesentlich erschwert wäre
2	der Gesundheit verbunden wäre
2	die gesetzliche Erben wären

4 Es wurde in diesem Fall eine maximale Anzahl von 4 Begleitern auf der linken Seite der Konjunktivform und eine minimale Frequenz von 2 im Gesamtkorpus eingestellt.

2	eine Person berufen wäre
2	Entwicklung zu besorgen wäre
2	Erheblichen Unannehmlichkeiten verbunden wären
2	erklärte Person berufen wäre
2	Ermittlungsverfahren zuständig gewesen wäre
2	erschwert oder aussichtslos wäre
2	Gefährdung wiedererlangt worden wäre
2	gemeinschaftlichen Kindes unbillig wäre
2	Gütergemeinschaft nicht eingetreten wäre
2	hohen Aufwand verbunden wäre
2	ist oder aufzunehmen wäre
2	Kind bereits geboren wäre
2	Kindes grob unbillig wäre
2	nicht erreicht worden wäre
2	Sache selbst zuständig wäre
2	sie nicht geeignet wäre
2	so vorzugehen gewesen wäre
2	Streitverfahren zuständig gewesen wäre
2	wirtschaftlichen Verhältnisse unbillig wäre
2	Zeit rechtshängig geworden wäre
2	öffentlichen Klage vorzulegen wären

Abb. 9.

Ab dieser Schwelle kann man Hypothesen anhand semantischer Ähnlichkeiten oder Regelmäßigkeiten unter den Clustern leichter erstellen. Diese Hypothesen müssen dann durch genauere Auswertung größerer Textausschnitte überprüft werden. In diesem Fall legen die Clusters in erster Linie die Vermutung nahe, dass das Nichteintreten bestimmter Voraussetzungen (als Bedingung für den Gebrauch des Konjunktivs) damit zusammenhängt, dass aus der angedachten Situation sich etwas Negatives (Ungerechtigkeit, Unzumutbarkeit, Schwierigkeiten oder Gefahr für eine Ermittlung, usw.) für ein Verfahren, eine Ermittlung oder einen Betroffenen ergäbe. Die Bestätigung ergibt sich aus Beispielen wie dem folgenden:

> (2) Der Ehegatte, der den gemeinsamen Haushalt führt, leistet dadurch seinen Beitrag im Sinn des Abs. 1; er hat an den anderen einen Anspruch auf Unterhalt, wobei eigene Einkünfte angemessen zu berücksichtigen sind. Dies gilt nach der Aufhebung des gemeinsamen Haushalts zugunsten des bisher

> Unterhaltsberechtigten weiter, sofern nicht die Geltendmachung des Unterhaltsanspruchs, besonders wegen der Gründe, die zur Aufhebung des gemeinsamen Haushalts geführt haben, **ein Mißbrauch des Rechtes wäre**. Ein Unterhaltsanspruch steht einem Ehegatten auch zu, soweit er seinen Beitrag nach Abs. 1 nicht zu leisten vermag. (94 BGB A)

Aus derselben Clusterliste könnten natürlich weitere Hypothesen hervorgehen, für denen Überprüfung der vorliegende Rahmen leider nicht ausreicht.

Bei hätte erhält man bei gleich bleibenden Einstellung der Parameter folgendes Bild:

Anzahl der Belege	Kontext
3	auf deren Rückgabe hätte
3	Erbfall nicht erlebt hätte
3	überhaupt nicht zugestimmt hätte
2	angeordnet wurde oder hätte
2	des Vertrages bestanden hätte
2	des Vertragsabschlusses bestanden hätte
2	eine verdeckte Ermittlung hätte
2	einen Angehörigen einzuschreiten hätte
2	Erbfalls nicht gelebt hätte
2	höhere Prämie zugestimmt hätte
2	ist oder eintreten hätte
2	Möglichkeit dazu gehabt hätte
2	nach diesem Gesetz hätte
2	nicht erhoben werden hätte
2	sei denn, dieser hätte
2	Vollziehung der Auflage hätte

Abb. 10.

Unter den zahlreichen Hinweisen, die sich aus der Tabelle ergeben, lohnt es sich beispielsweise, der scheinbar nicht grammatikalischen Kombination *nicht erhoben werden hätte* nachzugehen. Die Überprüfung der entsprechenden KWIC-Belege bringt folgende Konkordanzzeilen zutage:

Wenn die Anklage nicht erhoben werden hätte dürfen oder nicht weite…	CPP_1991_de.txt
Wenn die Anklage nicht erhoben werden hätte dürfen oder nicht weite…	CPP_1991_de.txt

Abb. 11.

Wie aus der Benennung der Ursprungsdateien hervorgeht, handelt es sich bei den letzten zwei Belegen um Südtiroler Wortstellungsvarianten im Gebrauch der Modalverben. Die quantitative Relevanz dieser Varianten kann man deutlich machen, wenn man eine zusätzliche Tabelle mit rechtsliegenden Clusters heranzieht:

Anzahl der Belege	Kontext
9	hätte angeordnet
9	hätte angeordnet werden
9	hätte können
6	hätte müssen
4	hätte angeordnet werden dürfen
4	hätte werden
3	hätte angeordnet werden können
3	hätte kennen
3	hätte kennen müssen
3	hätte sollen
1	hätte dürfen

Abb. 12.

Die direkte Reihenfolge <hätte + Modalverb> ist also ausreichend belegt und alterniert mit der „bundesdeutschen“ Reihenfolge, in der das Partizip und eventuelle andere Komponenten dazwischen liegen. Sie ist nur im Südtiroler und – in wenigen Fällen – im österreichischen Korpus zu finden.

Für semantische und inhaltliche Analysen sind sowohl die rechts als auch die links gestellten Clusters zu berücksichtigen. Aus einer ersten Auswertung, die hier nicht detailliert vorgelegt werden kann, ergibt sich eine gewisse Dichte von Normen, die thematisch mit Gerichtsentscheidungen, Erbfällen, Versicherungen und Vertragsklau-

seln zu tun haben, bei denen sich bestimmte Voraussetzungen nicht erfüllt haben.

Die reichhaltige *würde*-Clustertabelle, die man bei einer Einstellung mit min. 4 und max. 5 Elementen und mindestens 2 Belegen erhält, sieht folgendermaßen aus:

Anzahl der Belege	Kontext
5	des Mündels gefährden würde
5	eine Härte bedeuten würde
4	Bundesrepublik Deutschland herbeiführen würde
4	Interesse des Mündels gefährden würde
4	vereiteln oder beeinträchtigen würde
3	dazu berechtigt sein würde
3	der Zeuge nicht würde
3	deutschen Landes Nachteile bereiten würde
3	die Bundesrepublik Deutschland herbeiführen würde
3	dieser Sorgfalt entstanden sein würde
3	erheblichen Schwierigkeiten unterliegen würde
3	gesetzlichen Erben sein würden
3	Landes Nachteile bereiten würde
3	Sorgfalt entstanden sein würde
3	unmittelbar zustatten kommen würde
2	an, welcher berufen sein würde
2	beschwerten unmittelbar zustatten kommen würde
2	den italienischen Staat vornehmen würde
2	den Untersuchungserfolg gefährden würde
2	des angeklagten führen würde
2	des Verfahrens erheblich verzögern würde
2	die Gefahr zuziehen würde
2	die Wegnahme gefährdet würde
2	durch die Wegnahme gefährdet würde
2	einen unverhältnismäßigen Aufwand erfordern würde
2	eines Haftbefehles rechtfertigen würden
2	Entscheidung herbeigeführt haben würde
2	Erbteil gemindert werden würde
2	Erlaß eines Haftbefehls rechtfertigen würden
2	ersetzenden Nachteil bringen würde
2	Gefahr strafgerichtlicher Verfolgung zuziehen würde
2	gewähren, was sie haben würden
2	gründen erheblichen Schwierigkeiten unterliegen würde
2	Haft des angeklagten führen würde

2	Haushalts eine Härte bedeuten würde
2	italienischen Staat vornehmen würde
2	Nacherben vereiteln oder beeinträchtigen würde
2	nicht abgegeben haben würde
2	Rücktritts dazu berechtigt sein würde
2	schwere Härte darstellen würde
2	sein Erbteil gemindert werden würde
2	seine gesetzlichen Erben sein würden
2	sie den Untersuchungserfolg gefährden würde
2	so schwere Härte darstellen würde
2	strafgerichtlicher Verfolgung zuziehen würde
2	unverhältnismäßigem Nachteil gereichen würde
2	unverhältnismäßigen Aufwand erfordern würde
2	Verfahrens erheblich verzögern würde
2	was sie haben würden
2	welche der Zeuge nicht würde
2	welcher berufen sein würde
2	wert gestanden haben würde
2	wirklichen Wert gestanden haben würde
2	zu ersetzenden Nachteil bringen würde
2	zu unverhältnismäßigem Nachteil gereichen würde

Abb. 13.

Unter morphosyntaktischen Gesichtspunkten fallen hier wieder einige (vereinzelte) diatopische oder diachronische Varianten auf. Sie sind in den folgenden Konkordanzzeilen deutlicher einzuordnen und erweisen sich als Zeugnisse der Tradition der österreichischen Rechtssprache:

in Bezug auf Thatsachen, über welche der Zeuge nicht würde aussagen können,	A ZPO.txt
über Fragen, welche der Zeuge nicht würde beantworten können,	A ZPO.txt
über Fragen, die der Zeuge nicht würde beantworten können,	STPO.txt

Abb. 14.

Darüber hinaus sind die sehr zahlreichen Kombinationen mit Hilfsverben zu erwähnen (*sein würde*, *haben würde*, *werden würde*), deren Berechtigung und Verteilung in diesem Beitrag leider nur kurz ana-

lysiert werden kann. Die Kombination *sein würde* scheint fast ausschließlich im deutschen BGB vorzukommen. Unter den Belegen der Kombination *haben würde* stammen dagegen 7 aus dem österreichischen, 10 aus dem deutschen und 1 aus dem Südtiroler Teilkorpus. Eine ähnliche Verteilung besteht auch im Fall der Kombination *werden würde*. Insgesamt kann man also vermuten, dass solche Kombinationen lediglich im Schweizer Sprachraum der Gesetzgebung bewusst vermieden werden.

Unter semantischen Gesichtspunkten tritt auch in diesen Clusters eine „negative" Charakterisierung von Umständen verhältnismäßig häufig auf, und zwar in Bezug auf noch nicht abgeschlossene Situationen, die ihre Relevanz in der Gegenwart entfalten. Diese negative Färbung entspricht größtenteils derjenigen, die schon in Verbindung mit den *wäre*-Clusters hervorgetreten war. Dazu exemplarisch ein vollständiger Beleg:

> Der dingliche Arrest findet statt, wenn zu besorgen ist, dass ohne dessen Verhängung die Vollstreckung des Urteils vereitelt oder wesentlich erschwert werden würde. (917 ZPO ST).

In der Clustertabelle von *könnte(n)* sind verhältnismäßig wenige Clusters enthalten, weil das Korpus nur 120 entsprechende Belege enthält. Zur genaueren Veranschaulichung ist es hilfreich, auch die 2er Frequenzen aufzulisten:

Anzahl der Belege	Kontext
4	verantwortlich gemacht werden könnten
4	verletzt worden sein könnten
3	Untersuchungszweck gefährdet werden könnte
3	verletzt worden sein könnte
2	Beeinflussung ausgesetzt sein könnte
2	beeinträchtigt worden sein könnte
2	angeordnet werden könnte
2	das Kind ausüben könnte
2	das Zeugnis verweigern könnte
2	den Untersuchungszweck gefährden könnte
2	geschädigt worden sein könnte
2	psychischen Belastung führen könnte
2	schwerer Schaden entstehen könnte

2	verantwortlich gemacht werden könnte
2	vollständigen Aussage beeinflussen könnte
2	zuständig gemacht werden könnte

Abb. 15.

Die Clusteranalyse könnte mehrere weitere Ansätze zur Diskussion und Vertiefung von Regularitäten liefern. Quantitativ müsste jedoch auch der Frage nachgegangen werden, welche Konsequenzen sich für die Tragfähigkeit der Hypothesen aus relativ niedrigen Frequenzwerten wie unseren ergeben können[5].

5. Syntaktische Distribution

Die Clusters liefern viele wertvolle Ansätze zur Analyse der unmittelbaren syntaktischen und semantischen Umgebung der Konjunktivformen als solche. Zur Erforschung der eigentlichen Satzstrukturen und der Satzverknüpfungen muss das Verfahren dagegen erweitert und differenziert werden, um ein angemessenes Bild der entsprechenden syntaktischen Distribution zu gewinnen. Auch die Frage der syntaktischen Komplexität wird sich präsentieren: Ergeben sich in der Umgebung der Konjunktivformen besonders komplexe Strukturen?

Es wurde oben in der Einführung festgehalten, dass Irrealis Formen in nicht-performativen Teilen von Gesetzestexten zu erwarten sind. Demnach müssten Konjunktive grundsätzlich nur in Nebensätzen zu finden sein. Mangels eines Parsers, mit dem man zuverlässig Nebensatzstrukturen erkennen kann, mussten die Konjunktivbelege anhand der KWIC-Tabellen aufwändig geprüft werden. Aus dieser

5 Da für die verbleibenden Konjunktivformen (*sollte, müsste, dürfte*) die Anzahl der Belege noch niedriger ist, erscheint ihre Clusteranalyse nicht mehr sinnvoll.

Auswertung ergeben sich folgende syntaktische Umgebungen, die tatsächlich ausnahmslos aus untergeordneten Strukturen bestehen:

- Relativsätze:

  (3) Wenn sich ein Zeuge der Beantwortung von Fragen nicht entschlägt, **hinsichtlich deren er die Aussage gemäß §. 321 Z 1 und 2, zu verweigern berechtigt wäre**, kann sich das erkennende Gericht oder der die Vernehmung leitende beauftragte oder ersuchte Richter gleichfalls vorbehalten, über die Ablegung des Eides erst nach erfolgter Abhörung des Zeugen zu entscheiden.

- Konditionalsätze (auch mit einschränkendem Konnektor wie *sofern, soweit, insofern, insoweit, solange, sobald*). Bei letzteren kommt immer wieder die Tatsache zum Vorschein, dass eine semantische Einschränkung der Bedingungswahrscheinlichkeit im speziellen Verwendungskontext der Konjunktivformen kaum noch erkennbar ist:

  (2) Erfolgt die Zustellung im Ausland durch Behörden des Zustellstaates, so genügt die Einhaltung jener Vorschriften, die das Recht dieses Staates für die Zustellung entsprechender Schriftstücke vorsieht. Das gilt nicht, **wenn die Anwendung dieser Vorschriften mit Art. 6 der Europäischen Konvention zum Schutze der Menschenrechte und Grundfreiheiten, BGBl. Nr. 210/ 1958, unvereinbar wäre.** (106 ZPO A)

  (3) Unmündigen kann der Zutritt als Zuhörer verweigert werden, **sofern durch ihre Anwesenheit eine Gefährdung ihrer persönlichen Entwicklung zu besorgen wäre**. (171 ZPO A)

- Konzessive Sätze (*obwohl, obschon*):

  Wer eine Schuld bezahlt hat, **obwohl er sich auf die Aufrechnung hätte berufen können**, kann zum Nachteil Dritter die zugunsten seiner Forderung bestehenden Vorzugsrechte und Sicherheiten nicht mehr geltend machen, es sei denn, daß ihm das Bestehen der eigenen Forderung aus berechtigten Gründen nicht bekannt gewesen ist. (1251 CC ST)

  Die Anordnung, wodurch jemanden [sic] unter einer aufschiebenden unmöglichen Bedingung ein Recht ertheilt wird, ist ungültig, **obschon die Erfüllung der Bedingung erst in der Folge unmöglich, und die Unmöglichkeit dem Erblasser bekannt geworden wäre**. Eine auflösende unmögliche Bedingung

wird als nicht beygesetzt angesehen. Alles dieses gilt auch von den unerlaubten Bedingungen. (698 BGB A)

- Nebensätze, in denen schon der Konnektor einen Konjunktiv „verlangt“ (*als/wie ob/wenn, ohne dass*):

2 Der Anteil des Enterbten fällt, sofern der Erblasser nicht anders verfügt hat, an die gesetzlichen Erben des Erblassers, **wie wenn der Enterbte den Erbfall nicht erlebt hätte**. (478 BGB CH)

Das Gericht hat auch dann, wenn die Partei zwar obsiegt, aber keinen Kostenersatz beansprucht, darüber zu entscheiden, ob und wieweit der Gegner zum Ersatz der im § 64 Abs. 1 Z 1 und Z 5 genannten Beträge verpflichtet ist. Ist der Gegner der Partei zum Kostenersatz verpflichtet, so ist bei der Kostenfestsetzung so vorzugehen, **als wäre der Rechtsanwalt der Partei nicht vorläufig unentgeltlich beigegeben worden**. (70 ZPO A)

Im Fall einer kontrollierten Lieferung, das ist der Transport von verkehrsbeschränkten oder verbotenen Waren aus dem oder durch das Bundesgebiet, **ohne dass die Staatsanwaltschaft verpflichtet wäre**, nach § 2 Abs. 1 vorzugehen, gelten die Bestimmungen der §§ 71 und 72 des Bundesgesetzes über die justizielle Zusammenarbeit in Strafsachen mit den Mitgliedstaaten der Europäischen Union (EU-JZG) sinngemäß. (99 StPO A)

- Exzeptive Sätze (außer, es sei denn, es wäre denn):

Das Vermögen, welches nach Abzug aller Kosten und erlittenen Nachtheile über den Hauptstamm zurück bleibt, ist der Gewinn. Der Hauptstamm selbst bleibt ein Eigenthum derjenigen, welche dazu beygetragen haben; **außer es wäre** der Werth der Arbeiten zum Capitale geschlagen und alles als ein gemeinschaftliches Gut erklärt worden. (1192 BGB A)

Die Voraussetzungen und das Verfahren für das Wirksamwerden oder Unwirksamwerden von Vaterschaftsanerkenntnissen, über die die Niederschrift vor dem Inkrafttreten dieses Bundesgesetzes aufgenommen worden ist, bestimmen sich nach dem bisher geltenden Recht, **es sei denn, das Anerkenntnis wäre nach diesem Bundesgesetz rechtswirksam**. (5 BGB A)

- Vereinzelte *dass*-Sätze (auch konsekutive):

Im Übrigen darf Akteneinsicht nur vor Beendigung des Ermittlungsverfahrens und nur insoweit beschränkt werden, als besondere Umstände befürchten las-

> sen, **dass durch eine sofortige Kenntnisnahme von bestimmten Aktenstücken der Zweck der Ermittlungen gefährdet wäre**. (96 StPO A).
>
> (2) Ist eine Zeit nicht bestimmt, so ist die Sache zurückzugeben, nachdem der Entleiher den sich aus dem Zweck der Leihe ergebenden Gebrauch gemacht hat. Der Verleiher kann die Sache schon vorher zurückfordern, wenn so viel Zeit verstrichen ist, **dass der Entleiher den Gebrauch hätte machen können**. (604 BGB D)

- Ein konditionales Gefüge, in dem sowohl ein Obersatz als auch ein untergeordneter Bedingungssatz einen Konjunktiv enthalten. Dieses Gefüge kann in Einzelfällen auch einen uneingeleiteten Bedingungssatz enthalten:

> (2) Wird ein nichtiger Vertrag von den Parteien bestätigt, so sind diese im Zweifel verpflichtet, einander zu gewähren, **was sie haben würden, wenn der Vertrag von Anfang an gültig gewesen wäre**. (141 BGB D)
>
> Den unehelichen Kindern, die gemäß Artikel 279 ein Recht auf Unterhalt, Ausbildung und Erziehung haben, gebührt eine lebenslange Rente in Höhe des Ertrags des Erbteils, auf den sie Anspruch hätten, **wäre die Abstammung festgestellt oder anerkannt worden** (594). (580 CC ST).

Es wurden vereinzelt auch uneingeleitete Verb 1-Konditionalsätze, auf die aber ein Hauptsatz mit Verb im Indikativ folgt, gefunden. Solche Belege sind typisch für ältere Teile des österreichischen Korpus, und stellen unerwartete Ausnahmen dar:

> Auch eine Person, die sonst unfähig ist, ihr Vermögen zu verwalten, kann eine richtige und verfallene Schuld rechtmäßig abtragen, und sich ihrer Verbindlichkeit entledigen. **Hätte sie aber eine noch ungewisse, oder nicht verfallene Schuld abgetragen,** so sind die mit der Obsorge betrauten Personen, ihr Sachwalter oder Kurator berechtigt, das Geleistete zurückzufordern. (1421 BGB A).

Durch die Verknüpfung der nicht eingetretenen Voraussetzung mit einem weiteren Gefüge von Bedingung und Folge im selben Absatz entstehen gelegentlich komplexe syntaktische Gebilde wie das folgende, in dem Indikativ und Konjunktiv in derselben Umgebung zusammen vorkommen:

> (2) **Bezieht** sich in den Fällen des § 2078 der Irrtum nur auf eine bestimmte Person und ist diese anfechtungsberechtigt oder **würde** sie anfechtungsberechtigt sein, wenn sie zur Zeit des Erbfalls gelebt **hätte**, so **ist** ein anderer zur Anfechtung nicht berechtigt. (2080 BGB D)

Es ist schließlich noch auf eine scheinbare Ausnahme hinzuweisen, die sich im folgenden Beleg als weiterführender, erklärender Zusatz präsentiert:

> (1) Das Familiengericht stundet auf Antrag eine Ausgleichsforderung, soweit sie vom Schuldner nicht bestritten wird, wenn die sofortige Zahlung auch unter Berücksichtigung der Interessen des Gläubigers zur Unzeit **erfolgen würde**. Die sofortige Zahlung **würde** auch dann zur Unzeit erfolgen, wenn sie die Wohnverhältnisse oder sonstigen Lebensverhältnisse gemeinschaftlicher Kinder nachhaltig verschlechtern **würde**. (1382 BGB D)

6. Der Ausdruck der nicht eingetretenen Voraussetzung

Die normierenden Kontexte, in denen Konjunktivformen verwendet werden, enthalten oft, wie schon gezeigt, ein explizites oder mitzudenkendes Konditionalkonstrukt. Dabei entstehen semantische Relationen unter Aussagen bzw. Propositionen, auf die man in diesem Abschnitt näher eingehen muss.

Damit ein Konjunktiv II (als Irrealis) eingesetzt wird, bedarf es in einem normierenden Text einer logisch-semantischen Voraussetzung: Etwas traf nicht, oder trifft (noch) nicht zu/ein. Diese nicht eingetretene oder nicht eintretende Situation wird als nicht erfüllte Bedingung zugrunde gelegt, aus der sich eine Konsequenz ergibt, die im Konjunktiv II formuliert wird. Es wurden schon mehrere Belege wie der folgende präsentiert, in denen diese nichterfüllte Bedingung in einem Konditionalsatz explizit dargelegt ist:

> Diejenigen, die testamentarische oder gesetzliche Erben wären, **wenn der Verschollene am Tage, auf den die letzte Nachricht über ihn zurückgeht, verstorben wäre**, oder deren Erben (…) können die Einweisung in den einstweiligen Besitz des Vermögens beantragen (…). (50 CC ST).

Die Bedingung, die in diesem Fall nicht (mit Sicherheit) erfüllt ist, ist das Sterben des Verschollenen an einem bestimmten Tag. Es wird auf sie referiert, um eine Analogie zu bilden: Die Regel, die man anzuwenden hat, ist analog zu einer, die für einen hier nicht aktuellen Fall greift. Diese Analogie kann, wie schon dargelegt, eine weitere explizite Formulierung in hypothetischen Vergleichssätzen finden (s. o.) oder eine implizite in alternativen Konstrukten bzw. Ausdrucksmitteln finden. Der entsprechende Bezug kann in einem Adverb liegen, wie im folgenden Beispiel, in dem mit *ansonsten* das nicht zustande Kommen des Datenabgleichs gemeint ist:

> (2) Datenabgleich ist zulässig, wenn die Aufklärung eines Verbrechens (§ 17 Abs. 1 StGB) **ansonsten wesentlich erschwert wäre** und nur solche Daten einbezogen werden, die Gerichte, Staatsanwaltschaften und Sicherheitsbehörden für Zwecke eines bereits anhängigen Strafverfahrens oder sonst auf Grund bestehender Bundes- oder Landesgesetze ermittelt oder verarbeitet haben. (141 StPO A)

Dieselbe Art von Bezug kann gelegentlich in einem hypothetischen oder vergleichenden Zusammenhang eingebettet sein. Im folgenden Beispiel bezieht man sich gleich auf zwei nicht eingetretene Umstände (ein Sachverhalt bestand nicht, eine Prämie wurde nicht festgesetzt):

> Wenn die Änderungen so beschaffen sind, daß der Versicherer, wenn der neue Sachverhalt zur Zeit des Vertrages bestanden hätte, der Versicherung nur gegen eine höhere Prämie zugestimmt hätte, ist die Zahlung der Versicherungssumme **im selben Verhältnis** zu kürzen, in dem die vereinbarte Prämie zu jener steht, die **sonst festgesetzt worden wäre**. (1926 CC ST)

Es zeigt sich bei diesen negativen Voraussetzungen im Prinzip eine ähnliche Variationsbreite, wie sie schon in der Verwendung allgemeiner konditionaler Ausdrücke festgestellt wurde (Hoffmann 1998; Soffritti 2009). In Frage kommen, außer Adverbien, vereinzelte präpositionale Syntagmen (*bei*, *ohne*). Im folgenden Beispiel wird eine relevante negative Voraussetzung mithilfe von *ohne* in einem exzeptiven Zusammenhang genannt:

> Anstelle des Schadensersatzes statt der Leistung kann der Gläubiger Ersatz der Aufwendungen verlangen, die er im Vertrauen auf den Erhalt der Leistung

> gemacht hat und billigerweise machen durfte, **es sei denn, deren Zweck wäre auch ohne die Pflichtverletzung des Schuldners nicht erreicht worden**. (284 BGB D)

Die Begründung für den Gebrauch des Irrealis kann sich erst im Nachhinein erschließen, und zwar in einem Satz, aus dem hervorgeht, dass von der sonstigen Norm abgewichen wird:

> Rechtsstreitigkeiten gegenüber mehreren Personen, **die gemäß den Artikeln 18 und 19 vor verschiedenen Gerichten einzuleiten wären**, können, wenn sie in einem tatsächlichen oder rechtlichen Zusammenhang (40) stehen, vor dem Gericht eingeleitet werden, in dessen Sprengel eine dieser Personen ihren Wohnsitz oder ihr Domizil (43 , 47 ZGB.) hat, um im selben Verfahren entschieden zu werden (103 , 274). (33 CPC ST)

In diesem Kontext muss man eine implizierte Voraussetzung rekonstruieren, und sie lautet: Im vorliegenden Artikel gelten die Normen der Artikel 18 und 19 nicht, und es wird folgerichtig anders vorgegangen. Implizite Voraussetzungen dieser Art sind im Korpus nur spärlich vertreten und können dem Laien gelegentlich einen gewissen Interpretationsaufwand abfordern.

7. Irrealis und Analogie

Bei Nichtvorhandensein eines Sachverhalts in einem Gesamtzusammenhang, der eigentlich vorgeregelt wäre, entsteht oft die Notwendigkeit, eine bereits bestehende Regelung nur in Teilen anzupassen. Es wird dabei auf die ansonsten anzuwendende Regel Bezug genommen, die in der vorliegenden Situation nur „entsprechend“ oder „sinngemäß“ greift. Ein allgemeiner Verweis wäre in den neuen Redaktionsrichtlinien generell nicht mehr erlaubt:

> Eine „sinngemäße“ oder „entsprechende“ Anwendung anderer Rechtsvorschriften darf nicht angeordnet werden; es ist entweder uneingeschränkt auf die anderen Rechtsvorschriften in ihrer bestehenden Fassung zu verweisen

> oder aber anzugeben, mit welcher Maßgabe sie angewendet werden sollen.[6] (Handbuch der Rechtssetzungsstechnik: 19 – Unterstreichung im Original)

Analogie ist überhaupt ein möglicherweise unverzichtbarer Mechanismus im konzeptuellen Raum von Normgebilden. Es sind im Korpus, wie schon mehrmals gezeigt, viele Beispiele vorhanden, in denen der Gesetzgeber Analogien dadurch Rechnung trägt, dass er hypothetische Vergleiche heranzieht und sie als solche mit Konjunktiv bzw. besonderen hypothetischen Konnektoren markiert:

> Der Gläubiger einer Hypothek kann von dem Eigentümer verlangen, dass dieser eine vorrangige oder gleichrangige Hypothek löschen lässt, wenn sie im Zeitpunkt der Eintragung der Hypothek des Gläubigers mit dem Eigentum in einer Person vereinigt ist oder eine solche Vereinigung später eintritt. Ist das Eigentum nach der Eintragung der nach Satz 1 begünstigten Hypothek durch Sondernachfolge auf einen anderen übergegangen, so ist jeder Eigentümer wegen der zur Zeit seines Eigentums bestehenden Vereinigungen zur Löschung verpflichtet. **Der Löschungsanspruch ist in gleicher Weise gesichert, als wenn zu seiner Sicherung gleichzeitig mit der begünstigten Hypothek eine Vormerkung in das Grundbuch eingetragen worden wäre.** (1179a BGB D)

Die mehr oder weniger großen Ähnlichkeiten, die in solchen Normen angesprochen werden, stellen sich also in einen komplexen Zusammenhang mit der negativen Voraussetzung, die die Verwendung des Konjunktivs steuert. Es entsteht dabei das Muster: „dieser Sachverhalt stimmt zwar nicht mit dem anderen überein, aber es sind genug Ähnlichkeiten da, um faktisch nur eine Norm darauf anzuwenden“.

---

6 Dieses Verbot hat natürlich lange Zeit nicht bestanden, wie die zahlreichen Belege von *entsprechend* und *sinngemäß* in den Korpora zeigen.

8. Fachsprachliche Schlussfolgerungen

Es wurde versucht, anhand rein morphologischer Kriterien einen Bestand von Konjunktiv II-Verbformen aus einem Korpus zu gewinnen. Dieser Bestand enthält möglicherweise nicht alle im Korpus vorhandenen Formen, aber man kann sicher sein, dass der eventuell noch fehlende Anteil quantitativ sehr wenig bedeutend wäre. Aus der Analyse ergibt sich, trotz der verhältnismäßig überschaubaren Belegmasse, ein komplexes Bild. Sprachgeschichtlich scheint zwar eine Epoche der Regulierung angebrochen zu sein, in der der Konjunktiv als unzulässige Komplikation nicht zu verwenden und durch einfache, lineare Aussagen (ausschließlich im Indikativ?) zu ersetzen wäre. In einer deskriptiven Perspektive liegen aber deutliche Anzeichen dafür vor, dass eine solche Bereinigung sehr schwierig wäre und einen tiefgreifenden Einschnitt in Sprache, Begrifflichkeit und Stil bedeuten würde. Zu breitgefächert und konstitutiv scheinen dafür in den Gesetzbüchern die Funktionen zu sein, die mit dem Gebrauch des Konjunktivs und den entsprechenden syntaktischen und semantischen Strukturen eng zusammenhängen.

Wie insgesamt dargelegt, ergibt sich ein relativ kleiner Teil der verwendeten Konjunktive aus syntaktischen Regeln in Verbindung mit besonderen Konnektoren, ohne viel Spielraum für Abweichungen.

Ein beträchtlicher Teil ist semantisch und textuell unvermeidlich, weil sich bestimmte Szenarien nur (oder nicht ohne makrostrukturelle Eingriffe in die Textgestaltung) als irreal vor- und darstellen lassen. Selbst wenn man die Forderung, dass jeder Satz grundsätzlich nur eine Proposition bzw. eine Aussage enthalten sollte (Höfler/Bünzli 2010; Handbuch der Rechtsförmlichkeit: 97), tatsächlich erfüllen wollte, ergäbe sich in den noch erlaubten Satzverknüpfungen höchstwahrscheinlich kein syntaktischer Spielraum, in dem man konjunktivische Nebensätze samt ihren regierenden Obersätzen unterbringen könnte.

Ein weiterer, winziger Teil der Konjunktivformen ist möglicherweise schwach begründet oder muss in seiner Zweideutigkeit als Relikt vergangener Rechtsprachepochen akzeptiert werden. Die diato-

pische und diachronische Variation beim Konjunktivgebrauch ist vor allem in morphosyntaktischer Hinsicht feststellbar. Sie reflektiert Besonderheiten, die an sich nicht typisch für die Rechtssprache sind, wie man unter Heranziehung weiterer Korpora der österreichischen und Südtiroler Schriftsprache unschwer dokumentieren könnte.

Nachuntersuchung wären nun angebracht, um herauszufinden, ob in neueren Gesetzen Konjunktivformen in ähnlicher Menge und mit ähnlicher Distribution in den jeweiligen Ländern nachweisbar sind. Die in den letzten Jahren immer wieder erlassenen Richtlinien scheinen auf den ersten Blick nur punktuell befolgt zu werden und keine nachhaltigen Konsequenzen mit sich zu bringen, außer vielleicht in der Schweiz, wo politischer Druck und Kontrolle stärker die Abläufe beeinflussen und Projekte mit großer Tragweite (Höfler/Bünzli 2010) entstehen lassen.

9. Bilanz der korpusbasierten Untersuchung

Die analysierten Korpora wurden relativ einfach aus öffentlich zugänglichen Internetquellen gewonnen und ohne zusätzliche Indexierung aufbereitet. Einzig für das Südtiroler Korpus war eine komplexere Prozedur notwendig, da gegenwärtig nur die deutschsprachige Fassung des *Codice Civile* als pdf-Datei für das Publikum verfügbar ist. Bei fachsprachlichen, thematisch eingeschränkten Korpora dieser Art ist es nicht immer möglich, ähnliche Abfragemöglichkeiten zu realisieren wie bei allgemeinen Korpora, die PoS-Tagging und Parser integrieren.[7] Für diese Untersuchung wurde aus zeitlichen und praktischen Gründen darauf verzichtet, eine verhältnismäßig aufwändige morphosyntaktische Indexierung zu implementieren. Die zu erwartenden Vorteile wären jedoch nicht unbeträchtlich, da man möglicher-

7 Diesbezüglich lohnt sich ein Vergleich mit Hornung (2009), einer korpusbasierten Untersuchung von Konjunktiv I-Formen in einem Fachkorpus soziologischer Texte. Dabei wurde die Sketch-Engine-Software benutzt.

weise viel feinere Unterscheidungen etwa in der Aktantenstruktur der konjunktivhaltigen Sätze erzielen könnte. Demgegenüber stehen nun die Ergebnisse, die man mithilfe einer gängigen, stabilen und kostenlosen Softwareanwendung aus diesen Korpora gewinnen konnte. Kollokationen und vor allem Clusters, die unproblematisch erstellt wurden, haben sich als aufschlussreich erwiesen und haben viele wertvolle Anhaltspunkte für wichtige Hypothesen geliefert. Da bisher deskriptive Untersuchungen zum Gebrauch des Konjunktivs in Gesetzestexten kaum durchgeführt worden sind, ergaben sich schon aus diesen ersten Erhebungen und Analysen m.E. brauchbare Erkenntnisse und stimulierende Anregungen. Bei der Überprüfung einiger Hypothesen hat sich jedoch gezeigt, dass nicht nur mehrere sukzessive Abfragerunden, sondern vor allem geduldige Überprüfungen langer Ergebnistabellen unumgänglich bleiben. Dies nicht nur um bestimmte banale Zweifelsfälle zu klären, wie z.B. den morpholexikalischen Status des Wortes *Würde* am Satzanfang, sondern um viele kontextuelle Verhältnisse zu überprüfen, die sich gelegentlich erst nach mehrmaligem Lesen erschließen lassen. In diesen Fällen wären möglicherweise auch die bisher verfügbaren syntaktischen Parser überfragt. Der zusätzliche Aufwand ist also - je nach Fragestellung - nicht nur beträchtlich, sondern auch nicht immer genau vorhersehbar.

Darüber hinaus hat sich die Notwendigkeit bestätigt, für bestimmte Fragestellungen Ergebnisse aus beschränkten fachsprachlichen Korpora mit entsprechenden Ergebnissen aus großen Korpora der deutschen Sprache zu vergleichen. Auch dieser Schritt kann wegen der unterschiedlichen Funktionen der jeweiligen Abfragesoftware problematisch werden. Ob bedeutende Verbesserungen durch Integration immer raffinierterer Anwendungen zur Erstellung, Indexierung und Abfrage von Korpora möglich sind, muss noch abgewartet werden. Die entsprechende Methodik ist jedenfalls im Wandel begriffen, und auch in der korpusbasierten Erforschung der juristischen Textsorten wären breit angelegte, systematischere Untersuchungen sehr vielversprechend.

Bibliographie

Baroni, Marco *et al*. 2008. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora <http://wacky.sslmit.unibo.it/lib/exe/fetch.php?media=papers:wacky_2008.pdf> (letzter Abruf September 2013).

Bauer, Max W. *et al*. (Hgg.) 1989. *Italienisches Zivilgesetzbuch*. Bozen: Athesia.

Bauer, Max W. *et al*. (Hgg.) 1991. *Strafprozeßordnung mit Nebengesetzen*. Bozen: Athesia.

Bauer, Max W. *et al*. (Hgg.) 1996. *Italienische Zivilprozeßordnung*. Bozen: Athesia.

Bundesministerium der Justiz. Handbuch der Rechtsförmlichkeit. 3. Auflage <http://hdr.bmj.de/page_b.1.index.html> (letzter Abruf September 2013).

Gries, Stefan Th. & John Newman. Creating and using corpora (i.E.). In Podesva, Robert J. / Sharma, Devyani (Hgg.), Research methods in linguistics. Cambridge: Cambridge University Press. <http://www.linguistics.ucsb.edu/faculty/stgries/research/ToApp_STG-JN_CreatingUsingCorpora_ResMethLing.pdf> (letzter Abruf September 2013).

Hoffmann, Ludger 1998. Fachtextsorten der Institutionensprache I: das Gesetz. In Hoffmann, Ludger / Kalverkämper, Hartwig / Wiegand, Herbert E. (Hgg.) *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft.* Berlin/New York: de Gruyter, 522-528.

Höfler, Stefan / Bünzli, Alexandra 2010. Designing a controlled natural language for the representation of legal norms. *Zurich Open Repository and Archive*, University of Zurich <http://ceur-ws.org/Vol-622/paper8.pdf> (letzter Abruf September 2013).

Höfler, Stefan 2012. Legislative drafting guidelines: How different are they from controlled language rules for technical writing? In Kuhn, Tobias / Fuchs, Norbert E. (Hgg.) *Controlled Natural Language - Third International Workshop, CNL 2012.* Berlin -

Heidelberg, 138-151. <http://www.zora.uzh.ch/63553/> (letzter Abruf September 2013).

Hornung, Antonie 2009. Soziologische Fachtexte aus Korpusperspektive. Zum Auffinden indirekter Rede in den Annali di Sociologia / Jahrbüchern für Soziologie. *Fortschritte in Sprach- und Textkorpusdesign und linguistischer Korpusanalyse II – Linguistik online* 39, 3 <http://www.linguistik-online.de/39_09/hornung.html> (letzter Abruf September 2013).

Soffritti, Marcello 2009. Genau wenn oder nur insoweit? Einschränkende Bedingungen in deutschen und italienischen Gesetzbüchern. *Moderne Sprachen* 53, 49-80.

Soffritti, Marcello 2013. *Diatopische Unterschiede im Ausdruck von Bedingungen in deutschsprachigen Gesetzbüchern* (i.D).

DORIS HÖHMANN

10. Zur Untersuchung erweiterter Nominalgruppen mit Hilfe von Concgrams. Eine sprachvergleichende Studie zum deutschen und italienischen Umweltrecht

1. Vorbemerkungen zum zweifachen Untersuchungsgegenstand der Studie

1.1. Korpuslinguistische Untersuchungsperspektiven

Es ist unbestritten, dass die sprachwissenschaftliche Forschung von den heutigen technischen Möglichkeiten profitiert: Forschungsarbeiten können nicht nur auf eine größere und damit empirisch besser abgesicherte Datengrundlage gestellt werden, sondern es können auch mikroskopisch feine Regelmäßigkeiten im Sprachgebrauch herausgearbeitet werden, deren Erfassung ohne korpuslinguistische Tools kaum möglich wäre.

Trotz aller Fortschritte sind die praktischen Schwierigkeiten, mit denen man sich bei einer quantitativ-qualitativen Korpusanalyse konfrontiert sieht, jedoch nach wie vor sehr groß. So kann der Aufbau von hoch komplexen Nominalgruppen wie in den beiden folgenden Beispielbelegen, bisher noch nicht bewältigt werden:

(1) Für die Rechtmäßigkeit der Abwägung kommt es auf die erst nachträglich angestellten planerischen Erwägungen und gutachtlichen Untersuchungen zum Umfang der Auswirkungen eines Schadensfalls mit Stofffreisetzung, insbesondere zu dem von den Antragstellern geltend gemachten, von der Beigeladenen und vom Antragsgegner aber bestrittenen Schadensszenario einer sich bodennah ansammelnden Ethylengaswolke mit späterer Zündung,

nicht an. (Teilkorpus Gerichtsentscheidungen, VGH Mannheim, Beschluss vom 14.11.2011 – 8 S 1281/11)

(2) Nella specie, quanto alla VIA statale, le delibere citate del 15 marzo 2001 e del 6 dicembre 2001 risultano essere state precedute, tra l'altro, da uno studio di impatto ambientale, riferito alle opere di regolazione e strutturato in conformità al DPCM 27 dicembre 1988, che, pur se avviato in epoca anteriore all'entrata in vigore anche della versione iniziale del d.P.R. n. 357/1997, ha compreso, alla luce della straordinarietà degli interventi e della procedura, un'analisi globale di tutti i profili ambientali, incluse flora, fauna e, più in generale, habitat naturale (cfr. SIA -quadro di riferimento ambientale). (Teilkorpus Gerichtsentscheidungen, Consiglio Stato sez. VI, 16 marzo 2005 n. 1102)

Ein Desideratum der Korpuslinguistik besteht folglich darin, Untersuchungsstrategien zu entwickeln, die die Organisations- und Funktionsprinzipien der historischen gewachsenen Kultursprachen angemessener nachbilden und dabei der Komplexität fachbezogenerer Äußerungen Rechnung tragen können.

Der Umstand, dass die praktischen Schwierigkeiten der Korpusanalyse wesentlich an grundlegende Eigenschaften der Sprache gebunden sind – zu nennen sind insbesondere Polyfunkionalität und Polylexikalität der Sprachmittel sowie der weite Bereich des Ausgelassenen und Mitgemeinten – legt es nahe, eine Fokusveränderung in der Herangehensweise vorzunehmen und neben der linearen Abfolge auch andere Organisationsprinzipien der Sprache (die Vereindeutigung durch Kontextelemente, die Bildung von Konstituenten u.a.) zum Ausgangspunkt qualitativ-quantitativer Untersuchungen werden zu lassen. Bisher ist es jedoch nur unzureichend gelungen, noch andere Organisationsprinzipien als das der linearen Abfolge in die Funktionsweise der korpuslinguistischen Tools einzubauen.

Neue Forschungsmöglichkeiten und -perspektiven ergeben sich hier aus der Untersuchung von Concgrams (bzw. Concgrammen), mit deren Hilfe das gemeinsame Auftreten von Wörtern oder Wortbildungseinheiten unabhängig von ihrer genauen Abfolge und Verbindung innerhalb einer bestimmten Kollokationsspanne angezeigt wer-

den kann[1]. Sie erweisen sich daher für die korpuslinguistische Untersuchung komplexer fach(sprach)licher Äußerungen als besonders wertvoll, in denen grammatikalisch und/oder inhaltlich eng zusammengehörende Sprachmittel nicht unmittelbar aufeinanderfolgen[2], sondern etwa wie in der folgenden Beispielreihe durch das Vorkommen weiterer Äußerungskomponenten voneinander getrennt sind:

(3) Concgramme Auswirkung + auf, Schutzgut

a. oder anderen erheblichen **Auswirkungen auf** in § 1a genannte **Schutzgüter** zu besorgen
b. keine erheblichen nachteiligen **Auswirkungen auf** ein in § 2 Absatz 1 Satz 2 des Gesetzes über die Umweltverträglichkeitsprüfung genanntes **Schutzgut** haben können
c. mögliche Verlagerungen von nachteiligen **Auswirkungen** von einem **Schutzgut auf** ein anderes zu berücksichtigen
d. bedeutsamen **Auswirkungen** des Vorhabens **auf** in § 1a genannte **Schutzgüter** einschließlich der Wechselwirkungen
e. eine Beschreibung der **Auswirkungen** der unter Buchstabe e dargestellten Direktstrahlung und Abgabe radioaktiver Stoffe **auf** die in § 1a dargelegten **Schutzgüter**

(Teilkorpus „Einzelkommentierungen zum UVPG“)

Concgrams erweisen sich bei der empirischen, quantitativ ausgerichteten Untersuchung von teils festgefügten, teils weniger festgefügten Kombinationsmöglichkeiten von Ausdrucksmitteln wie auch bei der die Erfassung und Verwendung sprachlicher Varianten als ein wichtiges Hilfsmittel[3].

---

1 Der Terminus leitet sich von dem Namen des Korpustools ConcGram1.0 (Greaves 2009) ab, das in der Lage ist, die Kookkurrenzen von bis zu fünf Worteinheiten innerhalb einer bestimmten Kollokationsspanne automatisch zu erfassen.

2 Die Entwicklung dieses Tools (bzw. das ihm zugrundeliegende Prinzip) stellt insbesondere für die korpuslinguistische Untersuchung so genannter Klammersprachen, die gerade wie das Deutsche vornehmlich durch diskontinuierliche Konstituenten gekennzeichnet sind, einen entscheidenden Fortschritt dar (vgl. Höhmann 2011).

3 S. dazu u.a. Greaves (2009) und Höhmann (2011).

Mit Hilfe von Concgrams kann, wie in dem vorliegenden Beitrag gezeigt werden soll, auch die Verknüpfung von Prädikationen und im Zusammenhang damit u.a. der Auf- und Ausbau von Nominalgruppen in größeren Textvorkommen näher untersucht werden. Damit ist zugleich der zweite Ausgangspunkt bzw. Untersuchungsgegenstand der Studie angesprochen, der im folgenden Abschnitt erläutert wird.

Für sprachwissenschaftliche Untersuchungen mit Hilfe korpuslinguistischer Ansätze ist es besonders wichtig, dass die vereindeutigende Wirkung von Kontextelementen bei der Untersuchung großer Datenmengen dazu eingesetzt werden kann, die Suchergebnisse auf bestimmte Verwendungsweisen von Worteinheiten oder auf bestimmte kommunikative Zusammenhänge einzugrenzen. Dies kann wie in Beispiel (3) dadurch geschehen, dass von vornherein nach einer bestimmten Kombination von Sprachmitteln gesucht wird. Vorteilhaft ist darüber hinaus die Arbeit mit „erweiterten Congrams". Dazu werden etwas umfangreichere Kontextausschnitte von einzelnen Lexemen zum Ausgangspunkt genommen und halbmanuell auf ihr Vorkommen in Kombination mit anderen Worteinheiten hin untersucht. Auf diese Weise können Wörter und Wortbildungselemente zumeist nicht nur relativ leicht in ihren verschiedenen Verwendungsweisen vereindeutigt und unterschieden werden, sondern es wird auch einfacher möglich, die lexikalischen Füllungen der Ergänzungen und Angaben der Lexeme zu erfassen und die Belege herauszufiltern, die bestimmten semantischen Klassen (hier: die von nachteiligen Umweltauswirkungen betroffenen Menschen und Schutzgüter) zugeordnet werden können. Dazu jeweils eine Beispielreihe für beide der untersuchten Sprachen, wobei durch die Eingrenzung des Korpus auf Texte aus dem Umweltrecht[4] von vornherein ein gemeinsamer thematischer Rahmen gegeben ist:

(4) Erweiterte Concgrams zu *Folge* + lexikalische Füllungen der semantischen Klasse ‚von nachteiligen Umweltauswirkungen betroffene Menschen und Schutzgüter'

4 Zum Korpusaufbau s. Abschnitt 1.3.

a. verpflichtet, geeignete Vorsorgemaßnahmen zum Schutz vor nachteiligen Hochwasser**folgen** und zur Schadensminderung zu treffen, insbesondere die Nutzung von Grundstücken den möglichen nachteiligen **Folgen für Mensch**, **Umwelt** oder **Sachwerte** durch Hochwasser anzupassen.
b. In weiteren Kapiteln werden sodann die **Folgen** der Bergsenkungen **für einzelne Schutzgüter** dargestellt.
c. Der Betreiber einer Anlage der Kategorie A hat im Fall eines schweren Unfalls der zuständigen Behörde unverzüglich alle erforderlichen Informationen zur Verfügung zu stellen, um die **Folgen** des Unfalls **für das Wohl der Allgemeinheit** zu minimieren.
d. Die persönliche Betroffenheit des Klägers durch die Inanspruchnahme von Flächen und die daraus resultierenden **Folgen für seinen landwirtschaftlichen Betrieb** sind ebenfalls abwägungsfehlerfrei behandelt worden.
e. Dasselbe gilt mit Blick auf die **Folgen** des Baus einer neuen Überführung der Bundesstraße B 19 über die A 3 **für die Verkehrssituation im Raum Würzburg**.

(aus den Teilkorpora zum bundesdeutschen Umweltrecht)

(5) Erweiterte Concgrams zu *danno* + lexikalische Füllungen der semantischen Klasse ‚von nachteiligen Umweltauswirkungen betroffene Menschen und Schutzgüter'

a. tenendo conto del possibile **rischio di danni all'ambiente** e **alla salute** derivanti dall'utilizzo o dal trasporto del materiale, della sostanza o del prodotto secondario
b. in presenza del **pericolo di un danno per il patrimonio artistico e culturale**, **per l'ambiente**, **per la salute**, **per la sicurezza pubblica** o **la difesa nazionale**
c. creando fenomeni di biomagnificazione **a danno dei depredatori al vertice della catena alimentare (volpi, gatti, uccelli rapaci)**
d. Anche in questo secondo ricorso viene chiesta la condanna delle Amministrazioni al **risarcimento del** "**danno esistenziale** (…) unitamente al **danno patrimoniale**" **derivante ai membri del Comitato** ed **ai residenti nella zona** dall'ampliamento della discarica.
e. Salvo il caso di omessa richiesta del parere, il responsabile del procedimento non può essere chiamato a rispondere degli eventuali **danni derivanti dalla mancata espressione dei pareri di cui al presente comma**

(aus den Teilkorpora zum italienischen Umweltrecht)

Einschränkend muss hinzugefügt werden, dass auch mit Hilfe von einfachen und erweiterten Concgrams nur ein Teil der praktischen

Schwierigkeiten der quantitativen Korpusanalyse überwunden werden können, die sich aus der sprachlichen Ausdrucksvielfalt und Kontextabhängigkeit von Bedeutungen ergeben.

Zwar wird es durch die beschriebene Vorgehensweise möglich, die verschiedenen konkret vorkommenden Elemente von semantischen Klassen leichter und vollständiger zu erfassen und einzuordnen. Die dabei feststellbare sprachliche Vielfalt wie auch die Schwierigkeiten, die bei der Zuordnung der Lexeme zu einer bestimmten inhaltlichen Kategorie (s.u.) und bei der Abgrenzung der semantischen Klassen auftreten, machen jedoch zugleich deutlich, wie problematisch die Kategorienbildung bei der Erstellung fachbezogener Ontologien ist. Die Frage etwa, ob in diesem Zusammenhang der landwirtschaftliche Betrieb im Beispiel (5b) als Schutzgut einzuordnen ist oder nicht, erfordert juristisches Fachwissen. Ähnlich verhält es sich bei den grammatikalischen Formen. Während grammatikalisch gänzlich anders geartete Belege (wie bei *Folge* als Bestandteil des Stützverb- bzw. Funktionsverbgefüges *Folge leisten* oder *danno* als Verbform) relativ leicht aus den Trefferlisten herausgefiltert werden können, kommt es im Fall von Fügungen, die in einigen Kontexten einen Prozess der Lexikalisierung bzw. Grammatikalisierung unterlaufen und/oder einen formelhaften Charakter annehmen (wie z.B. bei *zur Folge* oder *per l'effetto*) mitunter zu erheblichen Abgrenzungsschwierigkeiten.

1.2. Zur Untersuchung von Nominalgruppen im Sprachvergleich

Wie die beiden eingangs angeführten Beispielbelege (1-2) zeigen, ist dem Deutschen und dem Italienischen die starke Komprimierung von Ausdrucksmitteln gemeinsam. Dabei steht das Italienische dem Deutschen, für das die Bildung hoch komplexer Nominalgruppen als charakteristisch gilt, in nichts nach. Doch wie lassen sich die zweifelsfrei bestehenden Unterschiede und Gemeinsamkeiten näher bestimmen und erfassen? Bei dem Versuch, hier zu mehr Klarheit zu gelan-

gen, liegt in der vorliegenden Studie das Hauptaugenmerk auf der Frage, wie Inhalte bzw. Prädikationen miteinander verknüpft werden[5].

In der Feststellung, dass es sich bei Nominalisierungen, Komposita und Mehrwortlexemen sowie Attribuierungen um eine verdichtete Verknüpfung von Prädikationen handelt[6], kann ein Schlüssel zum besseren Verständnis des Aufbaus von komplexen Nominalphrasen und ihren einzelnen Bestandteilen gesehen werden. Festzuhalten ist dabei insbesondere, dass aus diesem Blickwinkel heraus die Bildung von komplexeren Lexemen in einen größeren kommunikativen Zusammenhang gestellt wird. Während bei der Untersuchung von Nominalgruppen herkömmlicherweise die verschiedenen Wortbildungstypen sowie die Bestimmung der zur ihrer Abgrenzung notwendigen Kriterien[7] im Vordergrund stehen, rückt nun die Kombination der zum Einsatz kommenden Sprachmittel in den Mittelpunkt des Interesses. Dies bringt es u.a. mit sich, dass bei der Sprachbetrachtung die Rolle der Wortgrenze relativiert wird und lexikalische Einheiten, die aus mehreren Wörtern gebildet sind, an Bedeutung gewinnen. Dafür sind in Bezug auf den Untersuchungsgegenstand der vorliegenden Studie vor allem die beiden folgenden Gründe ausschlaggebend:

- Im Allgemeinen bestehen verschiedene Möglichkeiten der Kombination, die sich u.a. durch einen unterschiedlichen Grad der Explizitheit oder auch nur durch die Befolgung unterschiedlicher Schreibkonventionen auszeichnen und sich dabei nicht notwendigerweise auf den Terminologiecharakter der Sprachmittel auswirken; man denke etwa an die Varianten *legislative Gewalt / Legislative / gesetzgebende Gewalt / rechtsetzende Gewalt / Recht setzende Gewalt / Rechtsetzungsgewalt*[8] oder beispielsweise an die unterschiedlichen Benennungen von Gesetzen wie bei den Ausdrucksvarianten *BAföG / Bundesaus-*

5 Die Fragestellung knüpft u.a. an die Arbeiten von v. Polenz (2008), Schmidt (1999) und Prandi (2006) an.

6 Siehe dazu v. Polenz (2008: 40), der allerdings nicht auf die Bildung von Mehrwortlexemen eingeht.

7 Stellvertretend sei auf Elsen (2007) verwiesen.

8 Alle angeführten Varianten sind in einschlägigen Fachtexten belegt, wie eine am 27.10.2012 durchgeführte Google-Suche gezeigt hat.

bildungsförderungsgesetz / Bundesgesetzes über individuelle Förderung der Ausbildung.

- In der zwei- bzw. mehrsprachigen Kommunikation tritt deutlich zutage, dass der Aufbau der verschiedenen historisch gewachsenen Kultursprachen durch zahlreiche Asymmetrien geprägt ist,[9] die nicht nur das Sprachsystem, sondern auch den Sprachgebrauch betreffen. In Bezug auf die Gestaltung von Nominalgruppen ist hier insbesondere relevant, dass die Grenzen zwischen verschiedenen Wortbildungstypen in verschiedenen Sprachen unterschiedlich verlaufen und u.a. der Wortartenwechsel („Transposition") aufgrund der vorhandenen Asymmetrien zu den grundlegenden Übersetzungsverfahren gehört und sich daher etwa auch in den mehrsprachigen Fassungen europarechtlicher Gesetzestexte findet. Als Beispiel seien die Benennungen des Vertrags zur Gründung der Europäischen Atomgemeinschaft (1957) in seiner deutschen, italienischen, englischen und französischen Fassung genannt, deren grammatikalische Gestaltung jeweils unterschiedlich ausfällt bzw. nur zum Teil deckungsgleich ist:

de.	*Vertrag zur Gründung der Europäischen Atomgemeinschaft*
it.	*Trattato che istituisce la Comunità europea dell'energia atomica*
en.	*Treaty establishing the European Atomic Energy Community*
fr.	*Traité instituant la Communauté européenne de l'énergie atomique.*

In der großen Bedeutung, die den Wortgrenzen bisher bei der Behandlung und Klassifizierung lexikalischer Einheiten beigemessen wurde, dürfte der Grund dafür liegen, dass die für die Beschreibung von Nominalgruppen grundlegenden Bezeichnungen *prädeterminierende Attribute* und *Linkserweiterung* nicht selbstverständlich auch die Bestimmungsformen in den Komposita mit umfassen. In der vorliegenden Studie wird daher der Bezeichnung *Ausbau* der Vorzug

9 Zum unterschiedlichen Stellenwert der verschiedenen Wortbildungstypen im Deutschen und im Italienischen siehe u.a. die Ausführungen auf der Website <http://www.tedescoinrete.it> (letzter Zugriff am 27.10.2012).

gegeben, wenn es bei den jeweils untersuchten Sprachmitteln auch um die Bildung von Komposita geht.

Die inner- und zwischensprachliche Untersuchung der Frage, wie Inhalte in authentischen Texten[10] miteinander verknüpft werden, gehört (wie die Weiterentwicklung korpuslinguistischer Suchstrategien) ebenfalls zu den aktuellen sprachwissenschaftlichen Desiderata. Dass hier noch ein großer Forschungsbedarf besteht, macht sich etwa bei dem Versuch bemerkbar, halbwegs komplexe Äußerungen vorhersagbarer zu machen oder automatisch zu generieren.

Die Entscheidung, die Untersuchung einzelner Prädikationen zunächst auf Nominalgruppen zu beschränken, erklärt sich daraus, dass Nominalgruppen im Gegensatz zu den Verbalkomplexen relativ leicht umrissen werden können[11] und ihre Erfassung nur in begrenztem Maße durch weit voneinander getrennte Bestandteile polylexikalischer Lexeme oder anderweitig eng zusammengehörender Wortverknüpfungen erschwert wird[12]. Darüber hinaus fungieren Nominalgruppen als Baustein komplexerer syntaktischer Einheiten, so dass ihre Erfassung auch aus diesem Grund vorangestellt werden sollte.

Die Aufschlüsselung der verschiedenen miteinander verwobenen Prädikationen führt, wie bereits an dieser Stelle vorweggenommen werden soll, schließlich über die Eingrenzung auf Nominalgruppen hinaus.

---

10 Der Ausdruck „authentische Texte“ hebt darauf ab, dass als Datengrundlage vorgefundene, komplex belassene Sprachvorkommen und nicht selbst erfundene oder vereinfachte Beispielbelege dienen, die bei der Verfolgung anderer Untersuchungsansätze bzw. -ziele verwendet werden.

11 Dies gilt in den beiden untersuchten Sprachen zumindest für die oft durch ein Artikelwort markierten Grenzen der Linkserweiterung.

12 Bei Verbalklammern in Rechtstexten, in denen die einzelsprachspezifischen Ausdrucksmöglichkeiten oft in extremem Maße ausgereizt werden, sind Kollokationsspannen, die mehrere Dutzend Tokens umfassen, keine Seltenheit.

1.3. Zum Korpus- und Untersuchungsdesign

Der Untersuchungsansatz der Einzelfallstudie, die in Abschnitt 2 vorgestellt werden soll, beruht insbesondere auf den folgenden Ergebnissen[13] vorausgegangener einschlägiger korpuslinguistischer Studien:

- Die Voraussetzung für die qualitative Untersuchung einer großen Datenmenge ist offenbar die genaue Festlegung einer sehr kleinen Anzahl sprachlicher Phänomene, deren Frequenz und Distribution mit Hilfe elektronischer Suchfunktionen erfasst werden kann. Die unmittelbare (halb)automatische Erfassung der Sprachstrukturen beruht dabei notwendig auf der Identifizierung wiederkehrender Buchstaben- und Zeichenkombinationen.
- Die Auswertung größerer hochgradig spezialisierter Sprachvorkommen[14] lässt den Sprachreichtum der Kultur- und Ausbausprachen in aller Deutlichkeit hervortreten, der sich u.a. im Neben- und Miteinander zahlreicher Ausdrucksvarianten[15] und konkurrierender Begriffsbestimmungen äußert. Das Vorkommen sprachlicher Varianten stellt somit keine Randerscheinung dar, von der bei der Prozessierung natürlicher Sprachbestände und bei der Erstellung von Ontologien, die der Komplexität fachlicher Handlungsbereiche Rechnung zu tragen versuchen,

---

13 S. u.a. Soffritti (1999, 2000); Wichmann/Nielsen (2000); Heller/Engberg (2002); Rovere (2005). Die Studie baut insbesondere auf den Vorgängerarbeiten der Verfasserin auf (vgl. Höhmann 2005, 2007, 2009, 2011, 2013a und 2013b).

14 Zum Korpusdesign s. weiter unten.

15 Die intra- und intertextuelle Variation, die oftmals den kommunikativen Austausch erschwert und eine Standardisierung des Sprachgebrauchs als wünschenswert erscheinen lässt, kann nicht allein als Störfaktor kommunikativer Abläufe gesehen werden. Es können mindestens drei Funktionen der intra- und intertextuellen Variation für die Versprachlichung fachlicher Inhalte unterschieden werden: a) der fachsprachliche Ausbau kommunikativer Handlungsräume durch die Schaffung neuer Oppositionsbildungen, b) die Vertextung fachlicher Inhalte und c) die Wissensvermittlung durch Explikation, variierende Wiederholung bzw. Verdichtung der verschiedenen sinnrelevanten semantischen Merkmale (vgl. Höhmann 2005).

abgesehen werden könnte. Sinnvoll scheint es daher zu sein, die Gegenüberstellung einzelner Sprachstrukturen auf den Vergleich bestimmter Ausschnitte aus dem semantischen Netz auszuweiten. Die (halb)automatische Erfassung von Ausdrucksvarianten bringt indes Schwierigkeiten mit sich, die bisher erst ansatzweise gelöst werden können und in kontrastiv angelegten Untersuchungen noch ausgeprägter sind.

- Über die systematische korpuslinguistische Untersuchung der Kontexte einzelner Lexeme kann nicht nur eine gezielte Suche nach potentiell vorkommenden Äußerungskomponenten (Ergänzungen und Angaben) relativ leicht bewerkstelligt werden, sondern es wird auch möglich, die jeweiligen Realisierungs- und Auslassungsmuster zu erfassen. Sofern sich die erfassten Ausdrucksvarianten in Hinblick auf den Grad ihrer Explizitheit voneinander unterscheiden, können ausgelassene bzw. mitgemeinte Äußerungskomponenten leichter ermittelt werden.
- Die Gebrauchsprofile der einzelnen Lexeme weisen deutliche Unterschiede auf, die nicht mit den herkömmlichen Kategorien (Wortart, Fachterminus im engeren Sinn vs. allgemeine Fachlexik u.a.) gefasst werden können und die, wie zu vermuten ist, sowohl sprachlich als auch außersprachlich bedingt sein dürften. Der zuletzt genannte Befund begründet die Notwendigkeit, detaillierte Einzelfallstudien durchzuführen, die die konkrete Ausgestaltung von Vorkommensweisen einzelner Lexeme untersuchen.

Vor diesem Hintergrund wird für die Untersuchung der Frage, inwieweit die Verknüpfung von Inhalten im inner- und zwischensprachlichen Vergleich über die Aufschlüsselung und Erfassung von Prädikationen erhellt werden kann, der folgende Ansatz gewählt: Die qualitativ-quantitative Erfassung der Prädikationen erfolgt über die Untersuchung der sprachlichen Umgebungen einzelner Lexeme, wobei von vornherein der Bedeutung von Ausdrucksvarianten Rechnung getragen wird. Aufgrund ihrer Vorkommenshäufigkeit und relativ un-

ausgeprägten Terminologisierung[16] fiel im Rahmen der hier vorgestellten Einzelfallstudie die Wahl auf Lexeme der Wortfamilie ‚Umweltauswirkungen'. Von den Belegen und den Strukturgegebenheiten des Deutschen ausgehend werden die betreffenden Nominalgruppen – ebenfalls exemplarisch – hinsichtlich ihrer Erweiterung a) durch adjektivische Attribute als Beispiel für eine Linkserweiterung und b) durch Genitivattribute, die die unmittelbare Ursache bzw. den Ursprung der Auswirkungen näher bezeichnen, als Beispiel für eine Rechtserweiterung untersucht.

Datengrundlage ist ein knapp 2 Mio. Tokens großes zweisprachiges Korpus aus Texten des europäischen und nationalen Umweltrechts. Um inner- und zwischensprachliche Vergleichsmöglichkeiten leichter zu ermöglichen, wurden die einzelnen Texte anhand der Kriterien ‚Rechtssprache', ‚Rechtsordnung', ‚Textart' und ‚Rechtsgebiet' in verschiedenen Teilkorpora von etwa gleichem Umfang zusammengefasst. Für den innersprachlichen deutschen Sprachvergleich wurde außerdem ein weiteres Teilkorpus aus Gesetzeskommentaren zusammengestellt.

Bei der Zusammenstellung der Teilkorpora wurde versucht, den außersprachlichen Strukturprinzipien kommunikativer Handlungsräume so weit wie möglich Rechnung zu tragen und auf diese Weise ein Korpus zu schaffen, das sich durch eine hohe kommunikative Dichte[17] und Gegenstandsgemäßheit auszeichnet[18]. Insbesondere wird

---

16 Dies hat den Vorteil, dass die Vorkommen des Lexems nicht von vornherein auf einige wenige spezialisierte Verwendungskontexte beschränkt ist. Als stark terminologisierte Verwendungsweise hat die Wortverbindung „mittelbare Auswirkungen" zu gelten, deren Bedeutung für das Umweltrecht in der Rechtsprechung näher festgelegt wurde.

17 Damit wird zugleich der sozialen Dimension Rechnung getragen, die kommunikatives Handeln notwendig prägt.

18 Zu diesem Zweck wurden Textarten ausgewählt, die zu den grundlegenden Instrumenten juristischen Handelns gehören und von JuristInnen in ihrem Berufsalltag oft gemeinsam verwendet werden. Des Weiteren wurde darauf geachtet, dass bei der Textauswahl möglichst viele Faktoren der Kommunikationssituation konstant gehalten wurden (im Einzelnen Teilfachgebiet, Textsorte/n, Leser- bzw. Nutzerkreis, Kommunikationsgegenstand, Fachlichkeitsgrad, Rechtssprache/n, Medium und Entstehungszeitraum der Texte). Auch

durch diesen Zuschnitt der Datengrundlage der Umstand berücksichtigt, dass die Versprachlichung fachbezogener Zusammenhänge hochgradig spezialisiert verläuft und sich in Abhängigkeit von den Faktoren der Kommunikationssituation Regelmäßigkeiten im Sprachgebrauch beobachten lassen.

Die hohe kommunikative Dichte ist dabei zusammen mit der hochgradigen Spezialisierung des Korpus für die Untersuchung fachsprachlicher Sprachbestände in mehrfacher Hinsicht von Bedeutung:

- Da die potentiellen Bedeutungs- und Gebrauchsmöglichkeiten von Lexemen in bestimmten Handlungszusammenhängen nur beschränkt aktualisiert werden, wird die Erfassung der feststellbaren, zum Teil mikroskopisch feinen Vorkommensregelmäßigkeiten von Sprachmitteln erleichtert und fällt zugleich ergiebiger aus.
- Dadurch, dass die vorkommenden Sprachmittel größtenteils in einem engen fachlichen Zusammenhang stehen und schriftsprachliche Texte darüber hinaus durch einen besonders hohen Anteil expliziter Informationen bzw. Äußerungskomponenten gekennzeichnet sind, wird die Untersuchung fachsprachlicher semantischer Netze in ihrer komplexen konkreten Ausgestaltung vereinfacht. Insbesondere wird leichter feststellbar, welche Bedeutungs- bzw. Äußerungskomponenten zur Konstituierung fachbezogener Wissensrahmen – die notwendig immer auch sprachlich vermittelt sind – beitragen, zumal im Allgemeinen sowohl stärker komprimierte als auch explizitere Formulierungen zu den gleichen Sachverhalten aufgefunden werden können.

---

wurden nur vollständige, in sich abgeschlossene Texte in das Korpus aufgenommen. Fach(sprach)lich gesehen ist schließlich die Einbeziehung unionsrechtlicher Texte wichtig, die angesichts ihres großen Einflusses auf das innerstaatliche Verwaltungsrecht ebenfalls zu einer größeren Gegenstandsgemäßheit des Korpus beiträgt. Die verwendeten Texte weisen zudem fachbedingt eine starke Intertextualität auf.

- Zugleich werden damit günstige Voraussetzungen dafür geschaffen, ausgelassene bzw. mitgemeinte Äußerungskomponenten zu rekonstruieren.

Für die Untersuchung deutscher und italienischer Texte des Umweltrechts ist zudem von Belang, dass sie im Spannungsfeld zwischen innerstaatlichem, unionsrechtlichem und internationalem Recht entstanden sind bzw. entstehen. Das innerstaatliche Recht wird sowohl in juristischer als auch insbesondere in sprachlicher Hinsicht wesentlich vom Unionsrecht beeinflusst, das seinerseits unter den Bedingungen einer hochkomplexe Sprachkontaktsituation geschaffen wurde bzw. weiterentwickelt wird, in der verschiedene, zum Teil eng miteinander verwobene Rechtssprachen aufeinander treffen[19]. Diese Ausgangskonstellation führt einerseits dazu, dass bei vergleichenden Untersuchungen keine sprachpuristischen Standpunkte eingenommen werden können, da im Allgemeinen kaum zurückverfolgt werden kann, inwieweit die einzelnen fach- und nicht-fachsprachlichen Ausdrucksmittel durch Übersetzungen oder andere Formen der mehrsprachigen Kommunikation in den Institutionen der Europäischen Union geprägt worden sind. Zugleich eröffnet sie die Möglichkeit zu untersuchen, inwiefern sich der Sprachgebrauch im innerstaatlichen Verwaltungsrecht trotz der konstanten Bezugnahme auf das Unionsrecht von dem der europarechtlichen Rechtsvarietäten unterscheidet, und in einer diachronischen Perspektive beispielsweise danach zu fragen, inwieweit die Verwendung bestimmter Sprachmittel und -muster auf einzelsprachliche Gebrauchregelmäßigkeiten zurückgeführt werden kann.

Damit ist der Rahmen näher umrissen, in dem sich die im nächsten Abschnitt vorgestellte Einzelfallstudie bewegt.

---

19 Vgl. Höhmann (2013a, 2013b). Zur Spracharbeit im mehrsprachigen Recht der Europäischen Union s. u.a. Kjær (1999), Cosmai (2007) und die im Themenheft Rechtssprache der Vierteljahreszeitschrift *Muttersprache* veröffentlichten Beiträge von Engberg (2009), Muhr (2009), Schade (2009) und Wimmer (2009).

2. Einzelne Prädikationen im Sprachvergleich

Trotz der Versprachlichung gleich gelagerter inhaltlicher Zusammenhänge treten, wie im Verlauf der folgenden Erläuterungen deutlich wird, bei der feinmaschigen Untersuchung von Lexemen aus dem Wortfeld bzw. semantischen Teilnetz ‚Umweltauswirkungen' im inner- und zwischensprachlichen Vergleich erhebliche Unterschiede auf.

Erklärungsbedürftig sind zunächst die stark voneinander abweichenden Belegzahlen für die einzelnen Lexeme (s. Tab. 1), die noch einmal veranschaulichen, dass der sprachliche Auf- und Ausbau der historisch gewachsenen Kultursprachen trotz der vereinheitlichenden Wirkung des gemeinsamen europarechtlichen Bezugrahmens zahlreiche Asymmetrien aufweist. Die feststellbaren Unterschiede betreffen – neben Gründen wie der unterschiedlichen inhaltlichen Gewichtung, deren Untersuchung den Rahmen der vorliegenden Studie sprengen würde, – u.a. den fachsprachlichen Stellenwert bzw. Status der einzelnen Lexeme und Wortbildungselemente. So tritt der dominante Stellenwert des Lexems *Auswirkung* bereits aufgrund seiner größeren Verwendungshäufigkeit deutlich zutage, wie im erweiterten innersprachlichen Vergleich zusätzlich anhand der Komposita *Umweltauswirkung, Umweltbeeinträchtigung, Umweltfolge, Umweltrisiko* und *Umweltschaden* aufgezeigt werden kann (s. Tab. 2).

Die Lexempaare *Auswirkung – Folge* und *impatto – effetto* in den untersuchten Korpustexten	**EU-Recht**	**bundesdeutsches / italienisches Recht**	
	Richtlinien	Gesetzestexte	Gerichts-entscheidungen
Auswirkung[20]	423	251	283
Folge	82	44	99
Impatto	267	293	685
Effetto	447	159	267

Tab. 1.

Auch im Italienischen wird, wie die Ausweitung des innersprachlichen Vergleichs auf andere Lexeme des Wortfeldes bzw. semantischen Teilnetzes bestätigt (s. Tab. 3), der dominante Stellenwert des Lexems *impatto* bereits durch seine häufigere Verwendung angezeigt. Dabei dürfte die vergleichsweise hohe Gebrauchsfrequenz von *impatto* vor allem auch dadurch bedingt sein, dass dieses Wort überwiegend als Bestandteil grundlegender umweltrechtlicher Termini vorkommt, die im juristischen Berufsalltag eine herausragende Rolle spielen (wie z.B. im Terminus *valutazione dell'/di/d'impatto ambientale*, der italienischen Entsprechung für *Umweltverträglichkeitsprüfung*). In diesem Umstand dürfte auch einer der Gründe für die insgesamt fast doppelt so hohe Anzahl der italienischen Belege liegen, da im Fall von *Umweltverträglichkeitsprüfung* oder auch *Umweltverträglichkeitsstudie* anders als im Italienischen die „Auswirkungen" auf die Umwelt nicht benannt werden.

20 Wie auch in den folgenden Tabellen wird jeweils die Anzahl der einzelnen Belege in den ungefähr gleich großen, jeweils ca. 300000 Tokens umfassenden Teilkorpora genannt. Eine Ausnahme bildet lediglich das in Tab. 2 hinzugezogene Teilkorpus ‚Einzelkommentierungen zum Gesetz über die Umweltverträglichkeitsprüfung', das mit ca. 148000 Tokens knapp halb so groß wie die anderen Teilkorpora ist. Konsequent mitberücksichtigt wurde das Vorkommen der Lexeme in Komposita und Mehrwortlexemen.

Das Vorkommen von *Umwelt*-Komposita im innersprachlichen Vergleich	**bundesdeutsches Recht**		
	Gesetzestexte	Gerichts-entscheidungen	Einzelkom-mentierungen
Umweltauswirkung[21]	68	74	459
Umweltbeeinträchtigung	0	0	11
Umweltfolge	0	0	42
Umweltrisiko	0	3	4
Umweltschaden	37	1	0

Tab. 2.

Das Vorkommen von Lexemen der italienischen Wortfamilie ‚Umweltauswirkungen' im innersprachlichen Vergleich	**italienisches Recht**	
	Gesetzestexte	Gerichts-entscheidungen
Impatto	293	685
Effetto	159	267
Incidenza	26	87
ripercussioni[22]	4	4
Danno	45	61

Tab. 3.

Darüber hinaus ist das Verhältnis der untersuchten Lexeme zueinander insbesondere durch sich mehrfach überlagernde oder überkreuzende

21 S. die Erläuterungen in der Fußnote zu Tabelle 1.

22 Das Lexem ist in den untersuchten Teilkorpora nur in der Pluralform belegt.

Eins-zu-viele-Entsprechungen[23] gekennzeichnet. Um ein Beispiel zu nennen, kann *Auswirkung* in der umweltrechtlichen Fachdiskussion im Italienischen je nachdem u.a. durch *impatto, effetto, ripercussione* und *incidenza* wiedergegeben werden, wobei die untersuchten Lexeme in beiden Sprachen nur einen Ausschnitt des betreffenden Wortfeldes bzw. semantischen Teilnetzes darstellen.

Beim Ausbau der Nominalgruppen zeichnen sich hinsichtlich ihrer Erweiterung durch einzelne Sprachstrukturen deutliche Tendenzen ab, die die Anzahl der – nur theoretisch – nach oben offenen Kombinationsmöglichkeiten stark begrenzen und somit die Voraussagbarkeit des Sprachgebrauchs selbst in syntaktisch hoch komplexen und schwer lesbaren Textsorten wie Gerichtsentscheidungen steigern.

Festzuhalten ist zudem, dass der Sprachgebrauch in den untersuchten Korpustexten nationalen Rechts von dem der EU-Richtlinien zum Teil deutlich abweicht. Bei der Übernahme europarechtlicher Regelungsgehalte in das nationale Recht scheinen einzelsprachspezifische Charakteristika an Boden zu gewinnen bzw. umgekehrt sich bestimmte Sprachmuster nicht durchsetzen zu können. Wie die behandelten Sprachmuster zeigen, können die Unterschiede auch inhaltliche Aspekte betreffen.

2.1. Die Verwendung der Substantive in Verbindung mit adjektivischen Attributen

Die Unterschiede, die im inner- und zwischensprachlichen Vergleich bei der Untersuchung der Adjektive festgestellt werden können, mit denen die verschiedenen Lexeme des Wortfeldes bzw. semantischen Teilnetzes ‚Umweltauswirkungen' näher gekennzeichnet werden, lassen sich folgendermaßen zusammenfassen:

Schon im innersprachlichen Vergleich treten deutliche Unterschiede in der Verteilung der adjektivischen Attribute in Abhängigkeit von den einzelnen Lexemen auf: Während beispielsweise 53% der Belege des Lexems *Auswirkung* im Teilkorpus ‚Gesetzestexte' zu-

23 Rovere (1998) spricht in diesem Zusammenhang von „polisemie incrociate".

sammen mit den untersuchten Adjektiven *erheblich, signifikant, nachteilig, negativ, schädlich* vorkommen, liegt der entsprechende Anteil bei *Folge* mit 6,8 % weit darunter. Bei den beiden italienischen Lexemen *impatto* und *effetto* zeichnet sich mit einem jeweiligen Anteil von 17,9 % und 24,4 % ein zwar ebenfalls deutlicher, jedoch weniger stark ausgeprägter Unterschied im Sprachgebrauch ab (s. Tab.4 und 5). Unterschiede bestehen auch hinsichtlich der semantischen Klassen der Adjektive. In den beiden Teilkorpora ‚Gesetzestexte' und ‚Gerichtsentscheidungen', des innerstaatlichen italienischen Rechts beträgt bei dem Lexem *impatto* der Anteil der Adjektive, die das Ausmaß der Auswirkungen (*apprezzabile, importante, rilevante, significativo*) bezeichnen, ca. 76,1 %, bei dem Lexem *effetto* hingegen ca. 43,8 %, so dass bei Letzterem die Anzahl der Adjektive, die eine Wertung beinhalten (*dannoso, negativo, nocivo, sfavorevole*) eindeutig überwiegt.

Im zwischensprachlichen Vergleich wird deutlich, dass sich mit einem Verhältnis von annähernd 5:3 der Sprachgebrauch im bundesdeutschen Recht durch eine weitaus häufigere Verwendung der untersuchten attributiven Adjektive auszeichnet. Auch hier lassen sich bezüglich der beiden semantischen Klassen der Adjektive Unterschiede feststellen. Während im Deutschen das Verhältnis der Adjektive zueinander relativ ausgewogen ist, überwiegt im Italienischen deutlich der Anteil der Belege für Adjektive, die dazu dienen, das Ausmaß der Auswirkungen anzugeben. Diese beiden Befunde sind insofern überraschend, als dass die Notwendigkeit, Umweltauswirkungen einzuschätzen, schon rechtlich gesehen in den beiden EU-Staaten von der gleichen Bedeutung ist. Es ist daher anzunehmen, dass bei der Versprachlichung der inhaltlichen Zusammenhänge in den beiden Sprachen unterschiedliche Wege beschritten werden, die sehr wahrscheinlich auch den Bereich des Ausgelassenen und Mitgemeinten betreffen[24].

---

24 Eine nähere Untersuchung dieser Annahme würde eine anderes Untersuchungsdesign erforderlich machen und kann damit nicht im Rahmen der vorliegenden Studie geleistet werden.

Die Lexeme *Auswirkung* / *Folge* im Sprachvergleich (I)	**EU-Recht**	**bundesdeutsches Recht**	
	Richtlinien	Gesetzestexte	Gerichts-entscheidungen
Auswirkung	**423**	**251**	**283**
davon in Verbindung mit	158 (37,4 %)	133 (53 %)	111 (39,2 %)
erheblich/ signifikant*	60 / 10	56 / 4	57 / 0
negativ/ nachteilig / schädlich*	38/41/9	1/55/17	10/44/2
davon in den Kombinationen			
erheblich / signifikant* + negativ* / nachteilig**	24	30	22[25]
Folge	**82**	**44**	**99**
davon in Verbindung mit	12 (14,6%)	3 (6,8 %)	1 (1 %)
erheblich/ signifikant*	0 / 1	0 / 0	0 / 0
negativ/ nachteilig / schädlich*	1 / 8 / 2	0 / 3 / 0	0 / 1 / 0
davon in den Kombinationen			
erheblich / signifikant* + negativ* / nachteilig**	1	0	0

Tab. 4.

Wichtige einzelsprachliche Unterschiede zeichnen sich schließlich in der gemeinsamen Verwendung der Adjektive beider semantischer Klassen ab: hier ist der Anteil der Belege aus den bundesdeutschen Teilkorpora 2,6-mal so hoch wie in den italienischen Teilkorpora. Bezeichnend ist in diesem Zusammenhang auch, dass die untersuchten Nominalgruppen im Deutschen häufiger als im Italienischen durch eine Kombination mehrerer Adjektive erweitert werden. Festzuhalten ist des Weiteren, dass im Italienischen ein breiteres Spektrum von Adjektiven zur Bezeichnung des Ausmaßes und der Bewertung von Umweltauswirkungen gebraucht wird.

25 Es kommt lediglich die Kombination *erheblich* + nachteilig** vor.

Die Lexeme *impatto – effetto* im Sprachvergleich (I)	***EU-Recht***	***italienisches Recht***	
	Richtlinien	Gesetzestexte	Gerichts-entscheidungen
Impatto	**267**	**296**	**650**
davon in Verbindung mit	49 (18,4 %)	53 (17,9 %)	35 (5,4 %)
apprezzabil/ significativ*/ rilevant*/ important**	0/11/3/2	0/32/7/2	0/10/7/9
dannos / negativ*/ nociv* / sfavorevol**	0/33/0/0	0/12/0/0	1/5/0/3
davon in den Kombinationen			
apprezzabil/ significativ* / rilevant* + negativ** [26]	3	7	0
Effetto	**257**	**159**	**267**
davon in Verbindung mit	130 (50,6 %)	39 (24,5 %)	18 (6,7 %)
apprezzabil/ significativ*/ rilevant*/ important**	0/41/9/0	2/15/2/0	0/4/2/0
dannos / negativ*/ nociv* / sfavorevol**	3/52/25/0	2/12/6/0	0/10/0/2
davon in den Kombinationen			
apprezzabil/ significativ* / rilevant* + negativ**	17	10	3

Tab. 5.

Als Beispiel für die Unterschiede im Sprachgebrauch, die sich beim Vergleich der europarechtlichen Teilkorpora mit denen aus Texten nationalen Rechts feststellen lassen, sei angeführt, dass *effetto* in den italienischen Teilkorpora nicht halb so oft mit den untersuchten Adjektiven verknüpft wird wie in den auf Italienisch abgefassten

26 Anders als im Deutschen fällt die Wortstellung der Adjektive im Italienischen sehr unterschiedlich aus, was möglicherweise als Indiz dafür gewertet werden kann, dass hier kein weit verbreitetes, ‚eingefahrenes' oder ‚eingeschliffenes' Sprachmuster vorliegt.

Richtlinien[27]. Besonders deutliche einzelsprachspezifische Tendenzen zeichnen sich in der Rechtserweiterung durch Genitivattribute ab, um die es im folgenden Abschnitt gehen soll.

2.2. Die Verwendung der Substantive in Verbindung mit einem Genitivattribut zur Angabe der Ursache bzw. des Ursprungs von Umweltauswirkungen

Die Versprachlichung der Ursache bzw. des Ursprungs von Umweltauswirkungen liefert ein weiteres Beispiel dafür, dass trotz des Vorliegens gleicher syntaktischer Anschlussmöglichkeiten deutliche Unterschiede im Sprachgebrauch auftreten.

Im innersprachlichen Vergleich bestätigt sich die Erfahrung, dass die einzelnen Lexeme ein sehr unterschiedliches Gebrauchsprofil aufweisen. So kommt *Auswirkung* weitaus häufiger als *Folge* in Verbindung mit einem Genitivattribut zur Angabe der Ursache bzw. des Ursprungs von Umweltauswirkungen vor und insbesondere auch mit den Lexemen, die hier Oberbegriffe (wie z.B. *Projekt* oder *Vorhaben*) bezeichnen. *Folge* verbindet sich hingegen weitaus öfter mit Lexemen, die der juristischen Sphäre angehören.

---

27 Ausgenommen wurden 189 Belege für die Verwendung von *effetto serra* („Treibhauseffekt“) als Bestandteil umfangreicherer Mehrwortbenennungen (wie z.B. in *gas ad effetto serra*). Da diese Art der Verwendungsweise als solche keine Erweiterung durch Attribute zulässt und ausschließlich in dem europarechtlichen Teilkorpus vorkommt, würde ihre Mitberücksichtigung die sich herauskristallisierenden Tendenzen im Sprachgebrauch eher verdecken als aufzeigen.

Die Lexeme *Auswirkung* / *Folge* im Sprachvergleich (II)	**EU-Recht**	**bundesdeutsches Recht**	
	Richtlinien	Gesetzestexte	Gerichts-entscheidungen
Auswirkung	**423**	**251**	**283**
davon in Verbindung mit	169 (40 %)	75 (29,9 %)	125 (44 %)
**Plan*, *Pläne*[28]	8	2	9
Programm	8	2	3
Projekt	7	1	2
Vorhaben	1	26	54
in Verbindung mit anderen Lexemen, die die Ursache und/oder den Ursprung von Umweltfolgen bezeichnen[29]	145	44	57
Folge	**82**	**44**	**99**
davon in Verbindung mit	25 (30,5 %)	9 (20,5 %)	14 (14,1 %)
**Plan*, *Pläne*	0	0	0
Programm	0	0	0
Projekt	0	0	0
Vorhaben	0	1	0
in Verbindung mit anderen Lexeme, die die Ursache und/oder den Ursprung von Umweltfolgen bezeichnen	25	8	14

Tab. 6.

28 Die beiden Lexeme *Plan/Pläne* und *Programm** kommen in diesem Sprachmuster oft gemeinsam vor.

29 Zu den Abgrenzungsschwierigkeiten bei der Zuordnung der Belege zu bestimmten semantischen Klassen s. die Ausführungen in Abschnitt 1.1.

Die Lexeme *impatto / effetto* im Sprachvergleich (II)	**EU-Recht**	**italienisches Recht**	
	Richtlinien	Gesetzestexte	Gerichts-entscheidungen
impatto	**267**	**296**	**650**
davon in Verbindung mit	103 (38,6 %)	25 (8,8 %)	67 (10,3 %)
piano/i	0	24	1
programma/i	0	2	1
progetto/i	14	11	21
opera/e	0	4	18
in Verbindung mit anderen Lexemen, die die Ursache und/oder den Ursprung von Umweltfolgen bezeichnen	89	6	26
effetto	**257**	**159**	**168**
davon in Verbindung mit	60 (23,3 %)	25 (12 %)	26 (15,5 %)
piano/i	7	0	0
programma/i	7	0	0
progetto/i	6	5	4
opera/e	0	1	9
andere Lexeme, die die Ursache und/oder den Ursprung von Umweltfolgen bezeichnen	40	19	13

Tab. 7.

Eine ähnliche Beobachtung lässt sich im Italienischen in Bezug auf das Verhältnis der lexikalischen Varianten *impatto* und *effetto* machen. Im Verhalten des Lexems *effetto* ist hinsichtlich der Bezeichnung des Grundes oder Ursprungs bestimmter Wirkungen in diesem Zusammenhang besonders interessant, dass bei gesetzlichen Bestimmungen und bei verwaltungsrechtlichem Handeln u. Ä. öfter das Gegenstück zum deutschen Genitivattribut verwendet wird (z.B. *effetti dell'articolo, effetti dell'approvazione del progetto*), während diese Möglichkeit des syntaktischen Anschlusses bei Ursachen, die für Auswirkungen auf Umweltschutzgüter verantwortlich sind, seltener genutzt wird.

Zwischensprachlich gesehen fällt hingegen auf, dass zum Ausdruck der Ursache bzw. des Ursprungs von Umweltauswirkungen das Genitivattribut weitaus häufiger als sein italienisches Pendant vorkommt, auf das bei dem Lexem *impatto* 16,1 % entfallen, während der Anteil der entsprechenden Belege von *Auswirkung* 36,6 % beträgt.

Mit Hilfe von Concgrams kann zumindest ein Teil der syntaktischen Anschlussmöglichkeiten, die stattdessen zum Einsatz kommen, zügig erfasst werden. Verwendet werden insbesondere unterschiedliche Partizipialstrukturen und Relativsätze, wie die folgenden Beispielbelege zeigen:

(6) Erweiterte Concgramme zu *impatto* + lexikalische Füllungen der semantischen Klasse ‚Ursache und/oder Ursprung von Umweltfolgen'

a. possono essere interessate agli **impatti** sull'ambiente **dovuti all'attuazione dei piani, programmi o progetti**
b. gli **impatti** significativi sull'ambiente **derivanti dall'attuazione di detti piani e programmi**
c. gli **impatti indotti dall'opera** sul sistema ambientale,
d. non è inoltre da trascurare l'**impatto** ambientale, **causato dalla prevedibile intensificazione del traffico**
e. valutati gli **impatti** significativi **che l'attuazione del piano o del programma proposto potrebbe avere** sull'ambiente e sul patrimonio culturale
f. i principali **impatti** sull'ambiente e sul patrimonio culturale **che il progetto può produrre**

(aus den Teilkorpora zum italienischen Umweltrecht)

Bei der Ausweitung des Vergleichs auf den europarechtlichen Teilkorpus lässt sich schließlich feststellen, dass – mit Ausnahme des Gebrauchs von *Auswirkung* – das Genitivattribut und sein italienisches Gegenstück hier sehr viel häufiger als in den Teilkorpora des innersprachlichen Rechts vorkommen. In den einzelsprachlichen Korpora werden hier offensichtlich die Einflüsse der Sprachkontaktsituation, in der europarechtliche Texte entstehen, zurückgedrängt. Dass sich Verwendungsweisen in diesen Korpora durchsetzen, die für den Sprachgebrauch im Deutschen und Italienischen (d.h. in der bundesdeutschen und italienischen Sprachgemeinschaft) charakteristisch sind, kann auch an der Verwendung der Oberbegriffe *Vorhaben* und *opera* abge-

lesen werden, die in den untersuchten Kontexten hier vermehrt auftreten. Dieser Befund ist besonders auffällig bei dem Lexem *Vorhaben*, das in dem europarechtlichen Teilkorpus nicht belegt ist.

3. Schlussbemerkung und Ausblick

Es konnte gezeigt werden, dass mit Hilfe von Concgrams die semantischen und syntaktischen Bezüge in Nominalgruppen feinmaschig untersucht werden können. Zum einen ist festzuhalten, dass sich bei ihrer qualitativ-quantitativen Untersuchung mikroskopisch feine Regelmäßigkeiten bzw. Tendenzen im Zusammenspiel sprachlicher Ausdrucksmittel herausarbeiten lassen. Zum anderen erweisen sich Concgrams als wichtiges Hilfsmittel zur Erfassung der sprachlich zum Teil sehr unterschiedlich gestalteten Ausdrucksmittel und Verknüpfungsmöglichkeiten, die bei der Versprachlichung der inhaltlichen Zusammenhänge zwischen zwei oder mehreren Lexemen bestehen (können). Welches Potential diese korpuslinguistischen Untersuchungsstrategie für empirische Studien in sich birgt, soll abschließend noch einmal anhand einer weiteren Beispielreihe veranschaulicht werden. In den Belegen werden die Informationen darüber, wer oder was von Umweltauswirkungen betroffen ist und wodurch sie hervorgerufen werden oder entstanden sind, zusammen angezeigt:

(7) Erweiterte Concgramme zu Auswirkung / *impatto* + lexikalische Füllungen der semantischen Klassen ‚von nachteiligen Umweltauswirkungen betroffene Menschen und Schutzgüter' und ‚Ursache und/oder Ursprung von Umweltfolgen' *auf, Umwelt, Vorhaben*

a. Durchführung einer **U**VP hat den Sinn, die **Auswirkungen** eines **Vorhabens auf** die **Umwelt** zu ermitteln und diese
b. kein „Suchverfahren" ist, in dem alle nur erdenklichen **Auswirkungen** eines **Vorhabens auf Umwelt**güter und deren Wertigkeit
c. und einzelner Fachbeiträge zu den **Umweltauswirkungen** des **Vorhabens** greifen
d. Die **Auswirkungen** der in der Verordnung zu regelnden **Vorhaben**art **auf** die **Umwelt** müssen

Bei der Untersuchung der Variantenpaare aus dem Wortfeld ‚Umweltauswirkungen' wurde noch einmal deutlich, wie unterschiedlich die Gebrauchsregelmäßigkeiten bei einzelnen Lexemen ausfallen. Der Umstand, dass Lexeme wie *Umweltrisiko* und *ripercussioni*, die in anderen Kommunikationszusammenhängen recht häufig[30] vorkommen, in den untersuchten, hochgradig spezialisierten Teilkorpora zum bundesdeutschen und italienischen Umweltrecht nur 7- bzw. 8-mal belegt sind, lässt zudem erkennen, wie wichtig die Datengrundlage für die Auswertung und nur sehr begrenzt mögliche Verallgemeinerbarkeit empirischer Ergebnisse ist. Zugleich zeigt dieser Befund, dass in der Eingrenzung von Sprachbeständen anhand außersprachlicher Faktoren die Chance liegt, die in den so umrissenen Teilbereichen hochkomplexer Fachkommunikation (hier: Gesetzestexte und Gerichtsentscheidungen) vorkommenden Sprachmittel und feststellbaren Tendenzen im Sprachgebrauch leichter zu erfassen.

In der durchgeführten Einzelfallstudie wurde zudem deutlich, wie stark der Aufbau von Nominalgruppen durch bestimmte Verwendungsweisen gekennzeichnet ist und man es bei der Untersuchung bestimmter Textvorkommen – bei aller sprachlichen Vielfalt – auch im Hinblick auf das syntaktische Verknüpfungspotential mit einem begrenzten bzw. näher eingrenzbaren Ausschnitt des Sprachgebrauchs zu tun hat. In diesem Zusammenhang ist schließlich für die Fortentwicklung korpuslinguistischer Suchstrategien mit Hilfe von Concgrams von Bedeutung, dass das gemeinsame Auftreten einzelner Lexeme innerhalb einer bestimmten Kollokationsspanne und/oder der sie miteinander verbindenden syntaktischen Beziehungen zahlenmäßig stark eingegrenzt zu sein scheint, und daher vor allem Überlagerungen bestimmter lexikalischer Konfigurationen und die dabei auftretenden Sprachmuster untersuchungsrelevant sein dürften.

---

30 In einer am 8. November 2012 vorgenommenen Google-Recherche wurden ca. 195000 Treffer für *Umweltrisiko* erzielt, während die geschätzte Anzahl der Suchergebnisse für *ripercussioni* (das Lexem wird zumeist im Plural gebraucht) in Verbindung mit dem weiteren Suchwort *ambiente* bei über 1,3 Mio Treffern lag.

Literatur (Auswahl)

Cosmai, Domenico 2007. *Tradurre per l'Unione Europea. Prassi, problemi e prospettive del multilinguismo comunitario dopo l'ampliamento a est.* Milano: Hoepli.

Elsen, Hilke 2007. Wortgruppenlexeme – Beispiele aus Enzyklopädie, Zeitung, Baurecht und Wasserbau. *Fachsprache* 1–2/2007, 44-55.

Elsen, Hilke / Michel, Sascha 2007. Wortbildung im Sprachgebrauch: Desiderate und Perspektiven einer etablierten Forschungsrichtung. *Muttersprache*, 1/2007, 1-16.

Engberg, Jan 2009. Durchschaubarkeit durch Vielfalt – Vorteile eines mehr-sprachigen Rechtssystems und ihre linguistische Beschreibung. *Muttersprache* 3/2009, 181-191.

Greaves, Chris 2009. *ConcGram 1.0: a phraseological search engine*. Amsterdam: Benjamins.

Grossmann, Maria / Rainer, Franz (Hgg.) 2004. *La formazione delle parole in italiano*. Tübingen, Niemeyer Verlag.

v. Hahn, Walther / Vertan, Christina (Hgg.) 2010, *Fachsprachen in der weltweiten Kommunikation*. Frankfurt/M. u.a.: Peter Lang.

Heller, Dorothee / Engberg, Jan 2002. Verwendungskonventionen deontischer Modalmarker im deutschen Schiedsverfahrensrecht. In Gotti, Maurizio / Heller, Dorothee / Dossena, Marina (Hgg.) 2002. *Conflict and Negotiation in Specialized Texts. Selected Papers of the 2nd CERLIS Conference*. Bern u.a.: Peter Lang, 165-188.

Höhmann, Doris 2005. Die Bedeutung der fachlexikalischen Variation für die Textualität. Eine rechtslinguistische Studie. *MediAzioni. Rivista online di studi interdisciplinari su lingue e culture* 1/2005<http://mediazioni.sitlec.unibo.it/images/stories/PDF_folder/document-pdf/2005/workinprogress2005/2hhmann.pdf> (letzter Abruf September 2013).

Höhmann, Doris 2007. Überlegungen zur qualitativen und quantitativen Untersuchung von Modalverben und ihren Konkurrenzformen in Rechtstexten. In Heller, Dorothee / Taino,

Piergiulio (Hgg.) *Italienisch-deutsche Studien zur fachlichen Kommunikation*. Frankfurt/M.: Peter Lang, 25-39.

Höhmann, Doris 2009. Zur Realisierung und Nichtrealisierung von Äußerungskomponenten in authentischen Texten. Ein Beitrag zur Valenzforschung im juristischen Sprachgebrauch. In Heinrich, Wilma / Heiss, Christine (Hgg.): *Fachsprache, elektronische Wörterbücher, multimediale Datenbanken. Empirische Forschungsansätze der Sprach- und Übersetzungswissenschaft*. Festschrift für Marcello Soffritti zum 60. Geburtstag. München: Iudicium, 38-59.

Höhmann, Doris 2011. *Lexikalische Konfigurationen. Korpusgestützte Mikrostudien zur Sprachlichkeit im deutschen und italienischen Verwaltungsrecht.* Frankfurt/M. u.a.: Peter Lang.

Höhmann, Doris 2013a. Lo studio della variazione linguistica in ambito specialistico tramite l'analisi di concgrams. L'esempio del diritto ambientale. In Bosisio, Cristina / Cavagnoli, Stefania (Hgg.) *Atti del 12° Congresso dell'Associazioni Italiana di Linguistica Applicata* (Macerata, 23-24 febbraio 2012). *Comunicare le discipline attraverso le lingue: prospettiva traduttiva, didattica, socioculturale.* Perugia: Guerra Edizioni, 285-301.

Höhmann, Doris 2013b. Korpusbasierte Wortfeldanalysen. Zum Auf- und Ausbau verwaltungsrechtlicher Sprachbestände. In Sander, Gerald G. / Wagnerová, Marina (Hgg.) *Die Rechtssprache in der internationalen Diskussion*. Hamburg: Kovač, 43-57.

Kjær, Anne Lise 1999. Überlegungen zum Verhältnis von Sprache und Recht bei der Übersetzung von Rechtstexten der Europäischen Union. In Sandrini, Peter (Hg.) *Übersetzen von Rechtstexten. Fachkommunikation im Spannungsfeld zwischen Rechtsordnung und Sprache*. Tübingen: Narr, 63-79.

Muhr, Rudolf 2009. Die Unterschiede in der Rechtsterminologie Österreichs und Deutschlands und die Folgen für die Rechtssprache Deutsch im Rahmen der Europäischen Union. *Muttersprache* 3/2009, 199-216.

Lombardi, Alessandra 2007. Rechtswidrige Taten fördern. Usuelle Wortverbindungen in strafrechtlichen Texten (deutsch vs. italienisch). In Heller, Dorothee / Ehlich Konrad (Hgg.) *Studien zur Rechtskommunikation*, Frankfurt/M. u.a.: P. Lang, 115-151.

Lüdeling, Anke / Kyto, Merja (Hgg.) 2008-2009. *Handbook of Corpus Linguistics*. 2 voll. Berlin u.a.: de Gruyter.

Polenz, Peter von 2008. *Deutsche Satzsemantik. Grundbegriffe des Zwischen-den-Zeilen-Lesens*. Berlin u.a.: de Gruyter.

Prandi, Michele 2006. *Le regole e le scelte. Introduzione alla grammatica italiana*. Torino: UTET.

Rovere, Giovanni 1998. Polisemie incrociate nella comunicazione tecnica. In Cordin, Patrizia / Iliescu, Maria / Siller-Runggaldier, Heidi (Hgg.) *Parallela 6. Italiano e tedesco in contatto e a confronto. Italienisch und Deutsch im Kontakt und im Vergleich*. Trento: Dipartimento di scienze filologiche e storiche, 371-397.

Rovere, Giovanni 2005. *Capitoli di linguistica giuridica*. *Ricerche su corpora elettronici.* Alessandria: Edizioni dell'Orso.

Schade, Elke 2009. Vertragen sich europäische und deutsche Rechtssprache? – Leidensdruck bei der Umsetzung von EU-Recht. *Muttersprache* 3/2009, 192-198.

Schmidt, Jürgen Erich 2006. Serialisierung in der Nominalphrase. In: Agel, Vilmos u.a. (Hg.) *Dependenz und Valenz*. 2. Halbband. Berlin u.a.: de Gruyter, 1036–1049.

Schmidt, Reinhard 1999. *Untersuchungen zur Wissenschaftssprache*. Modena: Il Fiorino.

Soffritti, Marcello 1999. Textmerkmale deutscher und italienischer Gesetzesbücher: Übersetzung und kontrastive Analyse. In Sandrini, Peter (Hg.) *Übersetzen von Rechtstexten. Fachkommunikation im Spannungsfeld zwischen Rechtsordnung und Sprache*. Tübingen: Narr, 119-135.

Soffritti, Marcello 2000. Categorie pragmatestuali, precisione e polisemia in codici tedeschi e italiani: i verbi modali. In Schena, Leandro / Snel Trampus, Rita D. (Hgg.) *Traduttori e giuristi a confronto*. Bologna: CLUEB, 37-53.

Wichmann, Anne / Nielsen, Jane 2000. Rights and obligations in legal contracts: corpus evidence. In Dodd, Bill (Hg.) *Working with German corpora*. University of Birmingham Press: Birmingham, 245-266.

Wimmer, Rainer 2009. Zur Verflechtung von Spracharbeit und Rechtsarbeit in der EU. *Muttersprache* 3/2009, 234-239.

CHIARA PREITE / SILVIA CACCHIANI

11. Traduire la normativité dans les arrêts de la Cour de Justice de l'Union européenne : le cas des dispositifs en français et anglais

1. Introduction

L'arrêt de la Cour de Justice de l'Union européenne (désormais CJUE), comme et encore plus que toute autre décision de justice, assume des fonctions très semblables à celles des textes de loi, car elle s'occupe de régler des comportements de manière obligatoire : en effet, son destinataire est obligé de se soumettre à ses prescriptions par un système de lois légitime. Par conséquent, l'arrêt est investi non seulement d'un pouvoir décisionnel, mais aussi d'un pouvoir normatif. Comme l'explique Cornu, du point de vue linguistique cette impérativité s'exprime en particulier à travers les *actes de langage*, notamment à travers les expressions *déontiques* et *performatives* : "Le droit […] dote les actes de langage de conséquences juridiques. Le prononcé d'une parole devient, en vertu du droit, générateur de droit […]. La décision de justice existe du seul fait de son prononcé" (2000^2 : 45).

En nous plaçant dans le sillage du concept désormais partagé d'*équivalence fonctionnelle* dans le discours juridique (Pigeon 1982), nous allons analyser l'expression de la performativité et de la déonticité dans les dispositifs d'un corpus d'arrêts dont la langue de procédure est le français, pour la comparer à ses traductions officielles en anglais. Nous partageons les affirmations de Flückiger, selon qui "la traduction doit tout d'abord respecter un style normatif spécifique qui ne donne pas l'impression d'être une traduction mais un texte qui s'approprie les traits propres à la langue de destination […]" (2005 : 356), tout en admettant que, dans le cas du droit communautaire "une

certaine standardisation des concepts est nécessaire afin d'éviter des expressions nationales trop étroitement liées à la langue ou au système juridique du rédacteur" (2005 : 356).

Partant, nous présenterons une étude contrastive qui ne poursuit pas l'objectif de suggérer des démarches traductives (qui, le cas échéant, sont souvent le résultat de processus de traduction assistée), mais de comparer les éléments linguistiques à travers lesquels les deux langues-cultures juridiques expriment l'impérativité, la normativité.

En particulier, nous nous proposons d'encadrer les notions de performativité et de modalité déontique d'un point de vue théorique (§2), avant d'illustrer brièvement la composition du corpus (§3) et de passer au repérage et à l'analyse des éléments linguistiques véhiculant cette normativité dans le corpus (§4), suivant l'hypothèse que les choix des formes employées pour l'expression des actes de langage sont déterminés par les pratiques de rédaction des documents juridiques tendant à une certaine standardisation, plutôt que par les règles morphosyntaxiques propres à chaque langue.

2. La performativité et la modalité déontique

Le cadre linguistique de la théorie des actes de langage[1], selon laquelle tous les énoncés exercent un certain type d'action sur leur destinataire, tire son origine de la philosophie analytique de Austin (1962) et de Searle (1969).

Selon Austin, certains énoncés permettent au locuteur d'accomplir un acte, à savoir d'agir sur la réalité du monde ou sur autrui. Dans ce cadre, produire un énoncé signifie performer une action. D'abord, le philosophe propose une subdivision des énoncés entre *constatifs* – qui décrivent le monde en termes de vériconditionnalité – et *performatifs* – qui exécutent par le seul fait de leur énonciation, une action pouvant réussir ou échouer. Ensuite, il

1 Voir entre autres Kerbrat-Orecchioni (2001).

reformule son hypothèse et jette les bases de la théorie des actes de langage, selon laquelle pour chaque énoncé il faut distinguer l'*acte locutoire* (la forme de l'énoncé), l'*acte illocutoire* (la valeur de l'énoncé) et l'*acte perlocutoire* (assimilé à l'effet obtenu par son énonciation). Les performatifs se trouvent donc dispersés dans la classification austinienne des *valeurs illocutoires* (*verdictifs*, *exercitifs*, *promissifs*, *comportatifs* et *expositifs*), que Searle (1976) reformule dans ses études sur les actes de langage. Searle reconnaît ainsi les actes *assertifs*, *expressifs*, *promissifs*, *directifs* et *déclaratifs.* Si ces derniers correspondent aux *performatifs* d'Austin (ils réalisent l'action qu'ils énoncent si leurs conditions de réussite[2] sont réunies), les *directifs* contiennent les *déontiques* (modalité fondamentale pour l'expression du droit), à travers lesquels le locuteur essaie de faire exécuter quelque chose au destinataire, à travers une contrainte plus ou moins obligatoire ou facultative.

Les deux catégories d'actes de langage des *déclaratifs* / *performatifs* et des *directifs / déontiques* vont retenir notre attention en ce qu'ils marquent la force des dispositifs des arrêts de la CJUE que nous allons analyser.

2.1. Les performatifs et les déontiques dans les arrêts de la CJUE

Le philosophe du droit Kalinowski a affirmé la similarité existant entre les fonctions du langage du législateur et les fonctions du langage du juge "dans la mesure où il sert lui aussi à une spécifique promulgation des normes édictées par le juge, normes singulières, appelées à régler le comportement des parties" (1974 : 65) ; et Cornu (2000^2 : 352-354) a précisé plus récemment que la performativité des arrêts de la CJUE apparaît de manière privilégiée dans les dispositifs, dont les verbes sont les véritables prototypes :

---

2 Pour un résumé critique concernant les conditions de réussite des performatifs voir Grzegorczyk (1974).

> Comme le législateur, le juge dispose […]. Dans le jugement, comme dans la loi, la disposition se fait reconnaître comme un même type d'exposé […]. L'essentiel est que le dispositif énonce ce qui est obligatoire, au moins pour les parties, ce que le juge leur impose. Le dispositif porte des marques d'autorité comparables aux marques normatives de la règle de droit […], mais dans une décision individuelle. […] Les verbes du dispositif […] sont des prototypes de performatif […]. Cette vertu […] signifie que l'énoncé des verbes du dispositif suffit à réaliser la modification de l'ordonnancement juridique qui est attaché à la décision.

En ce qui concerne les actes de langage directifs, notre intérêt se limite à la modalité déontique[3] qui, comme le dit Kalinowski (1974 : 73-74), caractérise la normativité du langage juridique :

> Les langages du législateur et du juge sont performatifs et métalinguistiques. Ils sont aussi normatifs (déontiques) […], parce que tous les deux servent, le premier exclusivement, le second partiellement, à promulguer des normes : le langage du juge a en outre une portée normative (déontique) dans la mesure où il est employé pour la formation des énoncés performatifs maintenant ou modifiant le statut juridique des parties.

La déonticité a pour fonction de faire en sorte que des actes s'accomplissent, que quelque chose ait lieu, et indiquent la manière d'agir pour parvenir à un certain résultat ; elle ne réalise donc aucun acte par sa seule énonciation, à la différence de la performativité.

Toutefois, il est des cas où les verbes déontiques peuvent revêtir également une valeur performative, notamment lorsqu'ils apparaissent dans le dispositif, car ils *agissent sur* leur destinataire :

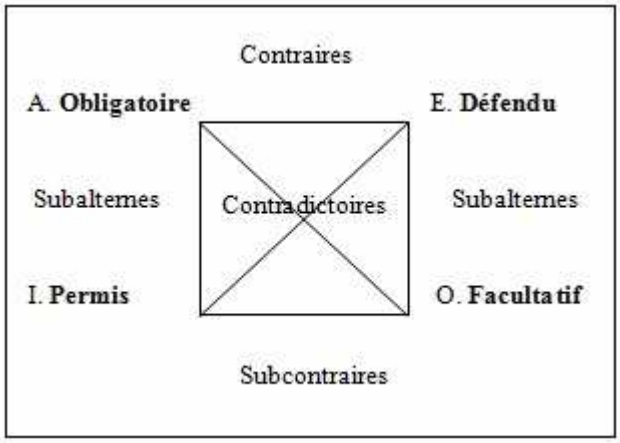
> La valeur illocutionnaire des auxiliaires modaux est particulièrement évidente dans leur interprétation déontique : en exprimant une obligation, une permission, une interdiction émanant de lui, prise en charge par lui, le locuteur *agit*, et de plus *agit sur* (sur l'allocutaire) (Cervoni 1987 : 97).

L'action prescriptive accomplie par le sujet juridique ne se limite pas à la modalité déontique de l'*Obligatoire*, mais inclut également le

---

3 Nous négligeons en revanche les modalités *aléthiques* et *épistémiques*. Voir entre autres Le Querler (1996) .

Défendu, le *Permis* et le *Facultatif* [4], qui représentent les trois autres pôles du "carré logique"[5] élaboré par Jean Ray en 1926 (1994 : 285) :



3. Le corpus

Notre corpus est formé des dispositifs de18 arrêts de la CJUE, dont la langue de procédure est le français, et de leurs traductions en anglais. Cette étude poursuit et élargit une recherche menée sur la comparaison entre les mêmes arrêts en français et leur traduction en italien (Preite 2012). Les arrêts sélectionnés remontent à l'année 2008 parce qu'au moment de la consultation du site de la Cour pour la première tranche du travail (français-italien : en février 2010) ils étaient les seuls à concerner des affaires clôturées, ayant par conséquent une forme définitive, et présentant une traduction en italien, aussi bien qu'en anglais.

4 Autrement dit *Autorisation* ("Authorization", Šarčević 2000 : 145).

5 L'élaboration du carré logique par Ray naît dans le but de classer les énonces à travers lesquels la loi exprime son impérativité, notamment dans le code civil.

CJUE_FR :	**CJUE_ANG :**
1. Affaire C-3 / 08: Décision préjudicielle	1. Case C-3 / 08: Preliminary ruling
2. Affaire C-12 / 08: Décision préjudicielle	2. Case C-12 / 08: Preliminary ruling
3. Affaire C-18 / 08: Décision préjudicielle	3. Case C-18 / 08: Preliminary ruling
4. Affaire C-59 / 08: Décision préjudicielle	4. Case C-59 / 08: Preliminary ruling
5. Affaire C-101 / 08: Décision préjudicielle	5. Case C-101 / 08: Preliminary ruling
6. Affaire C-128 / 08: Décision préjudicielle	6. Case C-128 / 08: Preliminary ruling
7. Affaire C-141 / 08_ P: Pourvoi	7. Case C-141 / 08_P: Appeal
8. Affaire C-168 / 08: Décision préjudicielle	8. Case C-168 / 08: Preliminary ruling
9. Affaires jointes C-202 / 08_P et C-208 / 08_P: Pourvoi	9. Joined cases C-202 / 08 P and C-208 / 08 P: Appeal
10. Affaire C-219 / 08: Manquement d'Etat	10. Case C-219 / 08: Failure of a Member State to fulfil obligations
11. Affaire C-285 / 08: Décision préjudicielle	11. Case C-285 / 08: Preliminary ruling
12. Affaire C-296 / 08 PPU: Décision préjudicielle	12. Case C-296 / 08 PPU: Preliminary ruling
13. Affaire C-299 / 08: Manquement d'Etat	13. Case C-299 / 08: Failure of a Member State to fulfil obligations
14. Affaire C-301 / 08: Décision préjudicielle	14. Case C-301 / 08: Preliminary ruling
15. Affaire C-311 / 08: Décision préjudicielle	15. Case C-311 / 08: Preliminary ruling
16. Affaire C-333 / 08: Manquement d'Etat	16. Case C-333 / 08: Failure of a Member State to fulfil obligations
17. Affaire C-425 / 08: Décision préjudicielle	17. Case C-425 / 08: Preliminary ruling
18. Affaire C-475 / 08: Manquement d'Etat	18. Case C-475 / 08: Failure of a Member State to fulfil obligations

Tableau 1. Sigles et typologie des arrêts du corpus : CJUE_FR ; CJUE_ANG

La CJUE représente un cas unique dans le panorama des Cours de Justice internationales et supranationales, car elle produit, pour un maximum de 23 langues, une législation jurisprudentielle[6] multilingue contraignante pour ses Etats membres et leurs citoyens. Cette législation est le résultat d'une série de "permutations de traductions" (McAuliffe 2011 : 98) à partir du français et vers le français, de la part des juges référendaires impliqués dans la rédaction des arrêts, qui ne sont pas nécessairement de langue maternelle française. De plus, le recours au logiciel GTI – système éditorial employé pour les documents préparés par la CJUE, qui aide et accélère le processus traductif à son intérieur – a conduit, du point de vue linguistique, à la naissance d'une sorte de "Court French" (McAuliffe 2011 : 98) très spécifique. Ce français de la Cour est caractérisé par la tendance à la répétition des mêmes expressions, le "couper et coller" (McAuliffe 2011 : 98) de morceaux de textes à partir de documents précédents, et l'adoption d'une langue et d'un style rigides et formulaires. Les arrêts ainsi produits constituent des documents hybrides, traduits en versions multilingues qui s'avèrent elles-mêmes hybrides, rigides et formulaires (Schäffner / Beverly 2001 ; McAuliffe 2011).

Rappelons que l'arrêt rédigé en français sert de texte source pour les traductions et que seul l'arrêt écrit dans la langue de procédure est authentique. Par conséquent, il peut arriver que l'arrêt authentique ne soit qu'une traduction, dans les cas où la langue de procédure n'est pas le français (Lashöfer 1992 : 140 ; Šarčević 2000 : 126). Par conséquent, le texte parallèle en anglais garde une structure standardisée, avec une subdivision en sections et sous-sections, afin de garantir la citation uniforme et de simplifier la comparaison avec le texte authentique, dans l'intérêt de la communication juridique plurilingue.

---

6 La CJUE rend des décisions de justice, mais dans le cadre particulier de l'Union européenne elle revêt également une fonction législative : "Par ses fonctions, la Cour cumule, en grande partie, le rôle du juge et du législateur en élaborant, à l'occasion des cas d'espèce, des règles qui favorisent […] l'intégration progressive des économies européennes" (Perelman, 1978 : 423).

La comparaison entre textes parallèles appartenant à une institution spécifique permet de relever la variété de moyens linguistiques employés pour formuler des actes linguistique exprimant le même contenu normatif, partant de la présupposition que la rédaction et la traduction de textes à l'intérieur de la CJUE suit une série de règles écrites et de pratiques non écrites, qui devraient garantir la mise en place de conditions d'exécution uniformes de l'instrument législatif, à savoir l'interprétation et l'application uniformes du contenu propositionnel de la règle juridique, indépendamment de la langue employée[7].

4. Analyse du corpus : l'expression de la valeur performative et de la déonticité dans les dispositifs

Il s'avère donc possible, pour ce qui est des dispositifs retenus, de comparer le rendement linguistique de la normativité dans les textes en français et dans leurs traductions en anglais, à travers une analyse qualitative. Après avoir composé les deux sous-corpora formés des dispositifs dans les deux langues, nous avons créé les *wordlists* respectives à l'aide de WordSmith Tools (Scott 1997). Ensuite, nous avons sélectionné à la main quelques exemples que nous avons enfin comparés à l'aide de ParaConc (Barlow 2007)[8].

---

7 Pour ce qui est desdites règles et pratiques, voir notamment les *Règles de technique législative à l'usage des services de la Commission*, de la Commission de l'Union Européenne, qui avertit : "L'usage des Règles peut être combiné utilement tant avec le Vadémécum Institutionnel et le Manuel des procédures opérationnelles qu'avec le Formulaire des actes établis par le service des juristes-linguistes du Secrétariat général du Conseil et le Vadémécum des publications communautaires établi par l'Office des publications" (1997, III éd. : p. i).

8 Vu la dimension réduite du corpus analysé, il convient de renvoyer tout calcul statistique et des fréquences à une autre recherche, à conduire sur un échantillon plus ample.

L'analyse se subdivise en trois volets : le premier prend en considération les formules qui introduisent les dispositifs, le deuxième l'expression des actes de langage *déclaratifs-performatifs* dans les "recours en manquement" et dans les "pourvois", et le troisième l'expression de la valeur *performative* et de la modalité *directive-déontique* dans les "renvois préjudiciels".

4.1. Les formules introductives des dispositifs

Les dispositifs des décisions de justice sont introduits par une formule que nous pouvons appeler, du point de vue pragmatique, "préfixe performatif" (Levinson 1983 ; Garzone 1996) ou "super-énoncé performatif enchâssant" (Kerbrat-Orecchioni 2003). Ces formules performatives dominent les énoncés dont se compose le dispositif. Et les énoncés ainsi enchâssés, dont la fonction primaire est de modifier la réalité en conformité avec les règles juridiques, relèvent au même titre de la catégorie des performatifs. Cela signifie que la valeur du préfixe performatif enchâssant s'étend au contenu du dispositif enchâssé qui assume force de loi et dévient opératoire dès son prononcé. Les expressions que nous allons étudier manifestent donc la volonté des juges de la CJUE, qui apprécient les cas qui leur sont soumis et formulent leur interprétation. Cette dernière acquiert alors une valeur normative, et précisément "dispositive".

Ces formules sont employées de manière presque systématique : les renvois préjudiciels sont introduits par l'expression "Par ces motifs la Cour *dit pour droit* ", qui est traduite en anglais par "On those / these grounds, the Court *hereby rules*" (ex. (6) à (14)), alors que les autres procédures sont introduites par "Par ces motifs la Cour *déclare et arrête*" (ex. (1) à (5)), dont la traduction en anglais est en revanche la variante abrégée "On those grounds, the Court *hereby*"[9] (ex. (2) à (5)), qui déplace les verbes performatifs à l'intérieur du

9 *Hereby* est l'introducteur typique des formules diapositives des décisions de justice en anglais, et il est également employé dans les tests diagnostiques de la performativité des énoncés (cf. Austin 1962).

contenu du dispositif. Ce dernier, grâce au préfixe performatif en *hereby*, réalise – comme les règles de droit – un effet juridique, tout en apparaissant en forme descriptive, caractérisée par l'emploi du présent de l'indicatif[10].

4.2. L'expression de la valeur performative dans les recours en manquement et dans les pourvois

Les formules introductives des dispositifs sont presque systématiquement les mêmes : la seule exception relevée se trouve dans l'exemple (1), où la formule complète "the Court *hereby rules*" introduit un recours en manquement.

Affaire C-333 / 08	Case C-333 / 08
Par ces motifs, la Cour (troisième chambre) ***déclare et arrête*** : 1. *En prévoyant* […] un régime d'autorisation préalable ne respectant pas le principe de proportionnalité, la République française *a manqué aux obligations* qui lui incombent en vertu de l'article 28 CE. 2. La République française *est condamnée aux dépens*.	On those grounds, the Court (Third Chamber) **hereby rules** : 1. *By laying down* […] a prior authorisation scheme not complying with the principle of proportionality, the French Republic *has failed to fulfil* its obligations under Article 28 EC. 2. The French Republic *is ordered to pay the costs*.

Exemple 1.

10 Des expressions comme *dit pour droit* et *hereby (rules)*, à la troisième personne du singulier du présent de l'indicatif, ancrent le dispositif au temps de l'énonciation et à la subjectivité de la CJUE qui s'exprime dans le texte. Généralement un énoncé performatif présente un verbe déclaratif-jussif à la première personne du singulier du présent de l'indicatif. Les arrêts, qui n'ont recours qu'à la troisième personne, ne sont pas pour autant exclus de la catégorie des performatifs, parce que, comme le rappelle Benveniste (1966 : 272), la *non-personne* peut être considérée comme un transfert de la première personne du singulier, qui peut toujours être reconvertie à la forme "normale" de la performativité.

Pareillement le choix des équivalents des verbes performatifs qui se trouvent dans les dispositions des recours en manquement et des pourvois montre un certain degré de récursivité, probablement dû aux raisons explicitées en (§ 3.). Les exceptions ne manquent pourtant pas, comme dans l'exemple (2), où la forme "est annulé" est traduite par *sets aside* aussi bien que par *annuls*.

Affaire C-299 / 08	Case C-299 / 08
Par ces motifs, la Cour (troisième chambre) ***déclare et arrête*** : 1. En adoptant et en maintenant en vigueur les articles 73 et 74-IV […], dans la mesure où ces dispositions *prévoient* une procédure de marchés de définition […], la République française *a manqué aux obligations* qui lui incombent […]. 2. Le recours *est rejeté* pour le surplus. 3. La République française *est condamnée* aux dépens.	On those grounds, the Court (Third Chamber) ***hereby*** : 1. ***Declares*** that, by adopting and keeping in force Articles 73 and 74-IV […], inasmuch as those provisions *lay down* a procedure for the award of marchés de definition […], the French Republic *has failed to fulfil* its obligations […]. 2. *Dismisses* the remainder of the action; 3. *Orders* the French Republic to pay the costs.

Exemple 2.

Affaire C-141 / 08 P	Case C-141 / 08 P
Par ces motifs, la Cour (première chambre) ***déclare et arrête*** : 1. L'arrêt du Tribunal de première instance des Communautés européennes […], ***est annulé*** […] ; 2. Le règlement (CE) n° 452 / 2007 du Conseil […] ***est annulé*** […] ; […]	On those grounds, the Court (First Chamber) **hereby** : 1. ***Sets aside*** the judgment of the Court of First Instance […] ; 2. ***Annuls*** Council Regulation (EC) No 452 / 2007 […] ; […]

Exemple 3.

Voilà d'autres exemples[11] suivis d'un commentaire :

Affaire C-219 / 08	Case C-219 / 08
Par ces motifs, la Cour (deuxième chambre) ***déclare et arrête*** :	On those grounds, the Court (Second Chamber) ***hereby*** :
1. Le recours *est rejeté.*	1. *Dismisses* the action;
2. La Commission des Communautés européennes et le Royaume de Belgique ***supportent*** chacun leurs propres dépens.	2. ***Orders*** the Commission of the European Communities and the Kingdom of Belgium each ***to bear*** its own costs.

Exemple 4.

Affaires jointes C-202 / 08 P et C-208 / 08 P	Joined cases C-202 / 08 P and C-208 / 08 P
Par ces motifs, la Cour (première chambre) ***déclare et arrête*** *:*	On those grounds, the Court (First Chamber) ***hereby*** :
1. Le pourvoi introduit par American Clothing Associates NV dans l'affaire C-202 / 08 P *est rejeté.*	1. *Dismisses* the appeal brought by American Clothing Associates NV in Case C-202 / 08 P ;
2. L'arrêt du Tribunal de première instance des Communautés européennes […] *est annulé* […].	2. *Sets aside* the judgment of the Court of First Instance of the European Communities […] ;
3. Le recours introduit par American Clothing Associates NV dans l'affaire T-215 / 06 *est rejeté.*	3. *Dismisses* the action brought by American Clothing Associates NV in Case T-215 / 06 ;
4. American Clothing Associates NV ***est condamnée*** aux dépens dans les affaires C-202 / 08 P et C-208 / 08 P.	4. ***Orders*** American Clothing Associates NV ***to pay*** the costs in Cases C-202 / 08 P and C-208 / 08 P.

Exemple 5.

11 L'italique et le gras dans les exemples sont les nôtres.

Les exemples (2) à (5) montrent que dans les recours en manquement l'anglais privilégie le préfixe performatif *hereby*, avec la dislocation des verbes performatifs à l'intérieur des dispositions (ex. (2) : "On those grounds, the Court (Third Chamber) *hereby* : 1. *Declares* that […] "), alors que dans les prononcés en français ils apparaissent dans le préfixe (ex. (2) : "Par ces motifs la Cour (troisième chambre) *déclare et arrête* : 1. […], la République française a manqué aux obligations qui lui incombent […]"). En anglais une partie du préfixe performatif est le plus souvent omise (à l'exception de l'ex. (1) déjà mentionné), la relative peut disparaître et *hereby* peut donc introduire les dispositions en forme de suite de phrases principales, syntaxiquement parallèles, montrant un verbe lui-même performatif à la forme active.

Les recours en manquement s'expriment généralement en français à travers une première phrase active par laquelle la CJUE constate si l'Etat membre "*a manqué* aux obligations", et un passage au passif – partagé par les *pourvois* – qui communique si "le recours *est rejeté*", si "l'arrêt *est* […] *annulé*", et / ou si l'Etat membre "*est condamné* aux dépens". En revanche l'anglais a recours à la voix active : dès lors, la Cour "*dismisses* the action / the appeal", "*sets aside* the judgment", "*orders* [a member state / a party to the action] to pay / bear the costs". Remarquons que l'exemple (1) est une exception à l'absence du passif dans les versions anglaises : "The French Republic *is ordered to pay* the costs".

Il est possible d'observer une régularité pour ce qui est des couples *supporter / condamner aux dépens* et *order to pay / bear the costs* : lorsque le français emploie, à la forme active, "[les parties en cause] *supportent* leurs propres dépens"(si les dépens sont répartis entre les parties requérante et défenderesse), l'anglais choisit systématiquement *order to bear*, alors que lorsque le français emploie, à la forme passive, "[la partie en cause / l'Etat membre] *est condamné(e)* aux dépens"(si une seule des parties doit payer les dépens), l'anglais préfère *order to pay*.

C'est la valeur attachée à *supporter* et *condamner* qui semble varier : en effet un acte de langage comme la condamnation exprime une valeur performative, car condamner quelqu'un implique que celui-ci est reconnu coupable (et la culpabilité de présumée passe au

statut d'état de fait). En revanche, *supporter* véhicule une valeur prescriptive, qui pourrait être reformulée à l'aide d'un déontique – "*doivent* supporter les dépens"– sans que le sens ne change, transformation qui ne demeure pas valable avec *condamner*[12].

En ce qui concerne l'anglais *order*, qui traduit à la fois *condamner* et *supporter*, c'est le *dictum* (*to bear / to pay*) qui véhicule la différence. En effet, *order* entraîne une lecture performative en vertu non seulement de la valeur performative dont l'investit le préfixe, mais aussi de l'emploi du présent de l'indicatif actif – qui correspond à l'un des sens que l'*Oxford English Dictionary* (http://www.oed.com) marque comme vieilli.

4.3. L'expression de la valeur performative et de la déonticité dans les renvois préjudiciels

Bien que la variabilité traductive reste toujours assez limitée, moins de régularité semble apparaître dans la traduction des renvois préjudiciels. La valeur performative servant à interpréter et à préciser certaines règles de droit communautaire recourt souvent à la modalité déontique. Rappelons qu'à travers la procédure du renvoi préjudiciel la CJUE s'adresse aux juges nationaux afin d'homogénéiser l'interprétation de normes européennes lorsqu'elle n'est pas établie de façon claire ou univoque. Ce type de dispositif revêt donc une double fonction : *performative* (véhiculée par les préfixes *dit pour droit* / *hereby rules*), en ce qu'il contribue à l'établissement d'une règle de droit, et *déontique* (marquée par exemple par des expressions telles que fr. *doit être interprété, il appartient à*, et ang. *must be* / *is to be interpreted, x must*, etc.), car la Cour attribue aux juges nationaux le devoir d'appliquer l'interprétation admise, qui assume force de loi à

12 Selon Garzone (1996 : 75) "La differenza tra i due tipi di discorso [performatif et déontique] non è identificabile attraverso alcuna "spia" morfo-sintattica, né lessicale, ma è affidata a criteri esclusivamente pragmatici, con una sostanziale identità formale tra discorso deontico (prescrittivo) e discorso performativo (costitutivo), entrambi realizzati in italiano nel testo normativo per mezzo del presente indicativo".

l'égard des Etats membres[13]. Par cette procédure, la Cour dispose que le juge national applique son interprétation : les dispositions se réalisent de manière médiatisée, en exerçant une pression sur le comportement de quelqu'un, alors que la performativité concerne l'interprétation correcte du contenu de la règle de droit.

Nous allons passer en revue les dispositifs des renvois préjudiciels et leurs traductions afin de proposer quelques remarques.

Lorsque les appréciations des juges sont exprimées au présent de l'indicatif – sans l'emploi d'expressions déontiques – la traduction ne montre pas de particularités, comme dans l'exemple (6) et dans le premier point de l'exemple (7) (cf. l'italique).

Affaire C-425 / 08	Case C-425 / 08
Par ces motifs, la Cour (deuxième chambre) ***dit pour droit*** :	On those grounds, the Court (Second Chamber) ***hereby rules*** :
L'examen des questions préjudicielles n'a révélé aucun élément de nature à affecter la validité de la directive 2004 / 73 / CE de la Commission, […].	*Examination of the questions referred has shown no factor capable of affecting the validity of Directive* 2004 / 73 / EC […].

Exemple 6.

Affaire C-301 / 08	Case C-301 / 08
Par ces motifs, la Cour (quatrième chambre) ***dit pour droit*** :	On those grounds, the Court (Fourth Chamber) ***hereby rules*** :
1. La convention pour l'unification de certaines règles relatives au transport	1. The Convention for the Unification of Certain Rules Relating to International

13 "[Le] norme comunitarie : tutte determinano una situazione di mera doverosità a provvedere, prescrivono ma non effettuano direttamente l'adeguamento normativo voluto. È noto che, viceversa, in virtù dell'art. 10 Cost. si è realizzato un meccanismo di automatico adattamento del nostro sistema giuridico ad una parte del diritto internazionale […]. Questo meccanismo si è realizzato proprio grazie al fatto che la norma non prescrive l'adattamento ma lo attua, e perciò la formula impiegata non contiene un 'dover essere' " (Carcaterra 1994 : 222).

aérien international […] *ne fait pas partie des normes de l'ordre juridique communautaire que la Cour a compétence pour interpréter* au titre de l'article 234 CE.	Carriage by Air […] *does not form part of the rules of the Community legal order which the Court of Justice has jurisdiction to interpret* under Article 234 EC.
2. Le règlement (CE) n° 2027/97 du Conseil […] *doit être interprété* en ce sens qu'*il ne s'oppose pas* à l'application de l'article 29 de la convention pour l'unification de certaines règles relatives au transport aérien international, […].	2. Council Regulation (EC) No 2027/97 […] *must be interpreted* as *not precluding* the application of Article 29 of the Convention for the Unification of Certain Rules Relating to International Carriage by Air, […].

Exemple 7.

Nous allons donc nous intéresser plus particulièrement aux dispositifs contenant des déontiques, dont nous rapportons un échantillon illustratif, avant de présenter quelques réflexions.

Affaire C-311/08	Case C-311/08
Par ces motifs, la Cour (troisième chambre) ***dit pour droit*** :	On those grounds, the Court (Third Chamber) ***hereby rules*** :
L'article 43 CE, […], *doit être interprété* en ce sens qu'*il ne s'oppose pas* en principe à une réglementation d'un État membre, telle que celle en cause au principal, […]. *Il appartient* cependant *à* la juridiction de renvoi de vérifier que la réglementation en cause au principal ne va pas au-delà de ce qui est nécessaire […].	Article 43 EC, […], *must be interpreted* as *not precluding*, in principle, legislation of a Member State, such as that at issue in the main proceedings, […]. However, *it is for* the referring court to verify whether the legislation at issue in the main proceedings goes beyond what is necessary […].

Exemple 8.

Affaire C-296/08_PPU	Case C-296/08_PPU
Par ces motifs, la Cour (troisième chambre) ***dit pour droit*** :	On those grounds, the Court (Third Chamber) ***hereby rules*** :
1. L'article 31 de la décision-cadre	1. Article 31 of Council Framework

2002/584/JAI du Conseil, […], *doit être interprété* en ce sens qu'il ne vise que l'hypothèse selon laquelle le régime du mandat d'arrêt européen est applicable, […].	Decision 2002/584/JHA […] *must be interpreted* as referring only to the situation in which the European arrest warrant system is applicable, […].
2. L'article 32 de la décision-cadre 2002/584 *doit être interprété* en ce sens qu'*il ne s'oppose pas* à l'application, par un État membre d'exécution, de la convention relative à l'extradition entre les États membres de l'Union européenne, […]	2. Article 32 of Framework Decision 2002/584 *must be interpreted* as *not precluding* the application by an executing Member State of the Convention relating to extradition between the Member States of the European Union […].

Exemple 9.

Affaire C-128/08	
Par ces motifs, la Cour (première chambre) ***dit pour droit*** :	On those grounds, the Court (First Chamber) ***hereby rules*** :
Dans la mesure où le droit communautaire, dans son état actuel et dans une situation telle que celle en cause au principal, *ne prescrit pas* de critères généraux pour la répartition des compétences entre les États membres s'agissant de l'élimination des doubles impositions à l'intérieur de la Communauté européenne, l'article 56 CE *ne s'oppose pas* à une convention fiscale bilatérale, […].	In so far as Community law, in its current state and in a situation such as that at issue in the main proceedings, *does not lay down* any general criteria for the attribution of areas of competence between the Member States in relation to the elimination of double taxation within the European Community, Article 56 EC *does not preclude* a bilateral tax convention, […].

Exemple 10.

Affaire C-59/08	Case C-59/08
Par ces motifs, la Cour (première chambre) ***dit pour droit*** :	On those grounds, the Court (First Chamber) **hereby rules** :
1. L'article 8, paragraphe 2, de la première directive 89/104/CEE du Conseil, […], *doit être interprété* en	1. Article 8(2) of First Council Directive 89/104/EEC of 21 December 1988 […], *is to be interpreted* as meaning

ce sens que le titulaire de la marque *peut invoquer* les droits conférés par cette dernière à l'encontre d'un licencié qui enfreint une clause du contrat de licence […].	that the proprietor of a trade mark *can invoke* the rights conferred by that trade mark against a licensee who contravenes a provision in a licence agreement […].
3. Lorsque la mise dans le commerce de produits de prestige par le licencié en violation d'une clause du contrat de licence *doit* néanmoins *être considérée* comme faite avec le consentement du titulaire de la marque, ce dernier *ne peut invoquer* une telle clause […], *que* dans le cas où il est établi, […], qu'une telle revente porte une atteinte à la renommée de la marque.	3. Where a licensee puts luxury goods on the market in contravention of a provision in a licence agreement but *must* nevertheless *be considered* to have done so with the consent of the proprietor of the trade mark, the proprietor of the trade mark *can rely on* such a provision […] *only if* it can be established that, […], such resale damages the reputation of the trade mark.

Exemple 11.

Affaire C-18/08	Case C-18/08
Par ces motifs, la Cour (deuxième chambre) ***dit pour droit*** :	On those grounds, the Court (Second Chamber) ***hereby rules*** :
La décision 2005/449/CE de la Commission, […], *ne peut être invoquée* par un particulier à l'encontre de la République française, […].	Commission Decision 2005/449/EC […] *cannot be relied on* by an individual against the French Republic, […].

Exemple 12.

Affaire C-12/08	Case C-12/08
Par ces motifs, la Cour (quatrième chambre) ***dit pour droit*** :	On those grounds, the Court (Fourth Chamber) ***hereby rules*** :
1. L'article 6 de la directive 98/59/CE du Conseil, du 20 juillet 1998, […], *doit être interprété* en ce sens qu'*il ne s'oppose pas* à une réglementation nationale qui instaure des procédures visant à *permettre* tant aux représentants des travailleurs qu'à ces derniers pris individuellement de faire	1. Article 6 of Council Directive 98/59/EC of 20 July 1998 […], *is to be interpreted* as *not precluding* national rules which introduce procedures intended *to permit* both workers' representatives and the workers themselves as individuals to ensure compliance with the obligations laid

contrôler le respect des obligations prévues par cette directive, mais qui limite le droit d'action individuel des travailleurs en ce qui concerne les griefs *pouvant être invoqués* […]. 2. La circonstance qu'une réglementation nationale, qui institue des procédures *permettant* aux représentants des travailleurs […]. 3. L'article 2 de la directive 98/59 *doit être interprété* en ce sens qu'*il s'oppose* à une réglementation nationale qui réduit les obligations de l'employeur […]. En appliquant le droit interne, la juridiction nationale *doit* […] prendre en considération l'ensemble des règles […]. *Il lui appartient* par conséquent d'assurer, dans le cadre de sa compétence, que les obligations pesant sur un tel employeur ne soient pas réduites par rapport à celles énoncées à l'article 2 de ladite directive.	down in that directive, but which limit the individual right of action of workers in regard to the complaints which *may be raised* […]. 2. The fact that national rules, establishing procedures *which permit* workers' representatives […]. 3. Article 2 of Directive 98/59 *must be interpreted* as *precluding* national rules which reduce the obligations of an employer […]. The national court *is required*, […] *to* consider all the rules […]. Consequently, *it must* ensure, within the limits of its jurisdiction, that the obligations binding such an employer are not reduced below those laid down in Article 2 of that directive.

Exemple 13.

Affaire C-168/08	Case C-168/08
Par ces motifs, la Cour (troisième chambre) ***dit pour droit*** : 1. Lorsque la juridiction de l'État membre requis *doit vérifier*, […] la reconnaissance et l'exécution des décisions en matière matrimoniale et en matière de responsabilité parentale […]. Cette juridiction *doit*, au contraire, *tenir compte* du fait que les époux possèdent également la nationalité de l'État membre d'origine […].	On those grounds, the Court (Third Chamber) ***hereby rules*** : 1. Where the court of the Member State addressed *must verify*, […] concerning jurisdiction and the recognition and enforcement of judgments in matrimonial matters and the matters of parental responsibility […]. That court *must*, on the contrary, *take into account* the fact that the spouses also hold the nationality of the Member State of origin […].

2. […] Au contraire, les juridictions des États membres dont les époux possèdent la nationalité sont compétentes en vertu de cette disposition, ces derniers pouvant saisir, selon leur choix, la juridiction de l'État membre devant laquelle le litige sera porté.	2. […] On the contrary, the courts of those Member States of which the spouses hold the nationality have jurisdiction under that provision and the spouses may seize the court of the Member State of their choice.

Exemple 14.

Les traductions tendent à montrer un degré minimal dans la variation des correspondances entre équivalents. Parfois la correspondance est respectée à chaque occurrence, comme dans le cas du fr. *permettre* traduit par l'ang. *permit* (ex. (13)) ou du fr. *(ne pas) s'opposer* qui est traduit par l'ang. *(not to) preclude* (ex. (7), (8), (9), (10) et (12)).

En particulier, pour ce qui est de l'expression de l'**Obligatoire**, il est possible de remarquer quelques variations : par exemple, la traduction du verbe modal dans le cas du fr. *doit/doivent être interprété(s)* montre tantôt l'ang. *must be interpreted* (ex. (7), (8) et (9)) tantôt *is to be interpreted* (ex. (11) et (13))[14] ; et encore, l'expression fr. *il appartient à* est traduite en anglais par exemple par *X must* (ex. (13)) et *it is for X to* (ex. (8)) : il s'agit de cas de "one-to-many equivalence"(Koller 1995). Bien qu'une analyse des données plus approfondie soit nécessaire, il nous paraît légitime d'observer que la relation inverse est également possible, comme le montre par exemple ang. *lay down* qui traduit à la fois fr. *prescrire* (ex. (10)) et prévoir (ex. (1) et (2)), illustrant ainsi un cas de "many-to-one equivalence".

Il est très intéressant de remarquer l'absence totale du modal ang. *shall* dans les dispositifs où généralement cette forme trouve place pour véhiculer une valeur à la fois déontique et performative (Garzone 1996 ; 2001). Dans notre corpus, il est supplanté par l'emploi fréquent de *must* (ex. (7), (8), (9), (11), (13) et (14)), à côté

14 La correspondance *doit* / *must* est gardée dans les expressions de l'ex. (14) fr. *doit vérifier / ang. must verify*, et fr. *doit tenir compte / ang. must take into account*.

de *to be to* (ex. (11) et (13)) et du moins direct *be required to* (ex. (13)). Pour le dire avec Šarčević (2000 : 138), *shall* exprime des "legal commands [which] are mandatory provisions ", alors que *must* exprime généralement des "requirements [which] indicate the existence of a duty that is usually procedural in nature" (Driedger 1976 : 14, cité par Šarčević 2000 : 138). Les obligations (*requirements*) exprimées par *must* sont également contraignantes (*mandatory*), car leur exécution est nécessaire sous peine de non validité d'une procédure ou d'un instrument. Reste, en tout cas, la particularité de l'absence de *shall*, vu que les occurrences de *must* concernent aussi des obligations sous forme de "command ".

L'expression du **Permis** trouve moins de réalisations dans les dispositifs analysés. Dans les exemples, (13) et (14) le verbe *pouvoir* – modal prototypique du Permis – est rendu par *may* (ex. (13) fr. *pouvant être invoqués* / ang. *which may be raised* ; ex. (14) fr. *ces derniers pouvant saisir* / ang. *the spouses may seize*). Rappelons également le recours au couple *permettre* / *permit* déjà mentionné, dans l'exemple (13). Le verbe *invoquer* – cas représentatif de variabilité traductive, car le fr. *(ne pas) (pouvoir) invoquer* peut être rendu par l'anglais de nombreuses manières : *can (not) invoke* (ex. (11), *can (not) rely / be relied on / upon* (ex. (11) et (12)), *may be raised* (ex. (13)) – n'entre pas seulement dans la construction du Permis. Il est utilisé également pour exprimer le Facultatif et, à la forme négative pour le Défendu.

Pour ce qui est du **Facultatif**, le modal ang. *can* traduit fréquemment le fr. *pouvoir* : par exemple, fr. *pouvoir invoquer* / ang. *can invoke* (ex. (11)) et fr. *ne peut invoquer […] que dans le cas où* / ang. *can rely on […] only if*) (ex. (11)).

Enfin, le **Défendu** s'exprime dans le dispositif par le biais de la négation du modal *pouvoir* – car l'interdiction (ou obligation négative) se réalise par le retrait d'une permission ou d'une autorisation facultative – traduit par l'ang. *cannot*, par exemple dans (12) fr. *ne peut être invoquée* / ang. *cannot be relied on*. Cette forme identifie une exception à l'application d'une permission ou d'une autorisation, à savoir elle implique une situation dans laquelle la permission ou l'autorisation ne peuvent pas être exercées.

5. Réflexions conclusives

Dans cette contribution nous avons fourni une première analyse comparative de l'expression de la performativité et de la modalité déontique dans les dispositifs de 18 arrêts de la CJUE, partant de l'hypothèse que les choix des formules employées pour exprimer les actes de langage juridiques dans les deux "Court Languages"– français et anglais – sont déterminés pas les pratiques partagées pour la rédaction des documents, plutôt que par les règles morphosyntaxiques des langues examinées.

La comparaison de textes parallèles appartenant à une Institution spécifique a permis de passer en revue les éléments utilisés pour formuler dans les deux langues des actes linguistiques qui véhiculent le même contenu normatif et de relever ainsi que l'expression de la performativité dans les dispositifs se caractérise par un degré minimal de variation dans les choix traductifs, ce qui pourrait est dû à l'emploi du logiciel GTI pour la traduction semi-automatique des arrêts. Dans le cas des performatifs, il est donc possible d'avancer l'hypothèse que la variabilité est minimale lorsqu'il s'agit des formules figées introduisant les dispositifs des renvois préjudiciels et des autres procédures (fr. *dit pour droit* / *déclare et arrête* ; ang. *hereby* / *hereby rules*). Toutefois, la traduction des éléments déontiques dans les mêmes dispositifs montre une variabilité et une liberté de traduction plus fortes, ce qui pourrait revenir, d'une part, au style personnel des juges qui rédigent la version originale en français et, de l'autre, aux choix traductifs privilégiés par les traducteurs qui ne demeurent nécessairement pas toujours les mêmes. Dans le cas des déontiques, il paraît donc possible que l'emploi du logiciel GTI n'intervienne que pour offrir au traducteur des recommandations ou plusieurs options. Les choix peuvent varier car il ne s'agit plus d'expressions figées encadrant le dispositif. Ainsi, malgré l'existence dans les deux "Court Languages" de couples d'équivalents récurrents, l'emploi de l'équivalent attendu en anglais n'est pas systématique de la part des traducteurs, ce qui confirme les résultats obtenus lors de la comparaison menée entre les mêmes éléments linguistiques dans le couple français-italien (Preite 2012). A ce propos, il convient de

remarquer que – selon les indications et les pratiques de l'Institution – les choix traductifs sont toujours subordonnés à l'exigence de garantir une interprétation et une application uniformes de l'instrument législatif – dans ce cas le contenu prescriptif des dispositifs – pour atteindre les mêmes objectifs indépendamment de la langue.

Références bibliographiques

Austin, John L. 1962. *How to Do Things with Words*. London : Oxford University Press.

Barlow Michael 2007. *ParaConc, Beta version* (build 269). © 1996, 2001, Huston : Athelstan.

Benveniste, Émile 1966. La philosophie analytique et le langage. In *Problèmes de linguistique générale*, 1. Paris : Gallimard, 267-276.

Carcaterra, Gaetano 1994. Norme costitutive. In Scarpelli, Ugo / Di Lucia, Paolo (eds) *Il linguaggio del diritto*. Milano : LED, 219-231.

Cervoni, Jean 1987. *L'énonciation*. Paris : PUF.

Commission Européenne 1997. *Règles de technique legislative. A l'usage des services de la Commission* : <http://ec.europa.eu/governance/better_regulation/documents/guide_de_la_comm_pour_la_redaction_fr.pdf>

Cornu, Gérard 2000[2]. *Linguistique juridique*. Paris : Montchrestien.

Driedger, Elmer A. 1976. *The Composition of Legislation*. Ottawa : Department of Justice.

Flückiger, Alexandre 2005. Le multilinguisme de l'Union européenne : un défi pour la qualité de la législation. In Gémar, Jean-Claude / Kasirer, Nicholas (eds) *Jurilinguistique*. Montréal : Thémis, 339-361.

Garzone, Giuliana 1996. *Performatività e linguaggio giuridico. Una proposta di classificazione*. Milano : Centro Linguistico Università Bocconi.

Garzone, Giuliana 2001. Deontic Modality and Performativity in English Legal Texts. In Gotti, Maurizio / Dossena, Marina (eds) *Modality in Specialized Texts*. Peter Lang : Bern, 153-174.

Grzegorczyk, Christophe 1974. Le rôle du performatif dans le langage du droit. *Archives de Philosophie du droit* 19, 229-241.

Kalinowski, Georges 1974. Sur les langages respectifs du législateur, du juge et de la loi. *Archives de Philosophie du droit* 19, 63-74.

Kerbrat-Orecchioni, Catherine 2001. *Les actes de langage dans le discours. Théorie et fonctionnement*. Paris : Nathan / VUEF.

Kerbrat-Orecchioni, Catherine 2003. Les énoncés du législateur : normativité ou performativité ?. In *Actes du séminaire SHS / ECLIPS, Normes, règles, régularités* : <http://www.univ-lyon.2fr/article.php3?id article=365>

Koller, Werner 1995. The concept of equivalence and the object of translation studies. *Target* 7 (2), 191-222.

Lashöfer, Jutta 1992. *Zum Stilwandel in richterlichen Entscheidungen*. Münster / New York : Waxmann.

Le Querler, Nicole 1996. *Typologie des modalités*. Caen : Presses Universitaires de Caen.

Levinson, Stephen C. 1983. *Pragmatics*. Cambridge : Cambridge University Press.

McAuliffe, Karen 2011. Hybrid texts and uniform law? The multilingual case law of the Court of Justice of the European Union. *International Journal for the Semiotics of Law* 24, 97-115.

Oxford English Dictionary : <http://www.oed.com>

Perelman, Chaïm 1978. Essai de synthèse. In Perelman, Chaïm / Foriers, Paul (eds) *La motivation des décisions de justice*. Bruylant : Bruxelles, 416-426.

Pigeon, Louis-Philippe 1982. La traduction juridique – L'équivalence fonctionnelle. In Gémar, Jean-Claude (ed.) *Langage du droit et traduction. Essais de jurilinguistique*. Montréal : Linguatech, 271-281.

Preite, Chiara 2012. L'expression de la normativité dans les arrêts de la Cour de Justice des Communautés européennes : une analyse contrastive français-italien. In Meunier, Mariette / Charret-Del Bove, Marion / Damette, Eliane (eds) *La traduction Juridique :*

points de vue didactiques et linguistiques. Lyon : Centre d'Etudes Linguistiques de l'Université Lyon.

Ray, Jean 1994. La struttura logica del codice civile francese [1926] . In Scarpelli, Ugo / Di Lucia, Paolo (eds) *Il linguaggio del diritto*. Milano : LED, 281-290.

Schäffner, Christina / Beverly, Adam 2001. The idea of the hybrid text in translation : Contact as conflict, *Across Languages and Cultures* 2, 167-180.

Šarčević, Susan 2000[2]. *New Approach to Legal Translation*. The Hague / London : Kluwer Law International.

Scott, Mike 1997. *Wordsmith Tools*, Oxford : Oxford University Press.

Searle, John R. 1969. *Speech Acts : An Essay in the Philosophy of Language*. Cambridge : Cambridge University Press.

Searle, John R. 1976. The classification of illocutionary acts. *Language in Society* 5, 1-24.

MARIE-PIERRE ESCOUBAS-BENVENISTE

12. Predicati giuridici e schemi argomentali nelle sentenze della Corte: Approccio bilingue francese-italiano[1]

1. Introduzione

L'insegnamento della lingua giuridica, la redazione di testi giuridici, la traduzione giuridica eseguita in maniera razionale e metodica devono oggi rappresentare un imperativo assoluto per le società che compongono l'Unione europea. E questo per almeno due ragioni: da una parte il crescente bisogno di numerosi professionisti plurilingui, ivi compreso, ma non esclusivamente, nei campi della mediazione linguistica, della traduzione, della redazione; d'altra parte, l'esigenza di precisione e di metodo nella formazione di questa doppia competenza giuridica e linguistica in un momento nuovo ma non inedito della storia europea, in cui i contatti di lingue moltiplicano le interferenze linguistiche e le occasioni di malintesi.

La prima ragione, d'ordine quantitativo, è legata alla considerevole produzione di testi risultanti dall' "inflazione normativa comunitaria"[2] (in particolare nel campo giurisprudenziale) e alla loro

---

1 Il presente contributo, ad eccezione di alcune precisazioni e della parte finale della conclusione, rappresenta la traduzione dell'articolo "Prédicats juridiques et schémas d'arguments dans les textes des arrêts de la Cour. Approche bilingue français-italien" pubblicato dall'autrice in Meunier, Mariette / Charret-Del Bove, Marion / Damette Eliane (dir.) *La traduction juridique : points de vue didactiques et linguistiques*, Publication du Centre d'Etudes Linguistiques, Nouvelle série 2012-1, Université de Lyon, 141-166.

2 L'espressione è di Cohen-Tanugi (1993: 768) che evoca anche, quasi dieci anni fa ormai, la "marée communautaire juridique" che sommerge una Francia

indispensabile pubblicazione nelle 24 lingue nazionali dei paesi-membri dell'Unione europea, affinché siano ricevuti, compresi ed integrati negli ordinamenti giuridici nazionali.

La seconda ragione, d'ordine qualitativo, è legata al contesto sociolinguistico europeo attuale che favorisce lo sviluppo esponenziale di situazioni di contatto tra lingue appartenenti allo stesso ceppo linguistico, il che fa sfumare, nell'uso, i contorni del senso di quelle parole dalle forme spesso similari.

Più che mai, quindi, appare indispensabile una descrizione del vocabolario giuridico e del comportamento delle unità lessicali nei testi giuridici. Una descrizione metodica e completa impone di tener conto dei fatti terminologici e dei fatti linguistici presenti ed operativi in tali testi. La descrizione dei concetti e delle nozioni appartiene al giurista comparatista mentre, in conformità con un approccio descrittivo del senso attraverso le forme linguistiche, al linguista spetta il compito di descrivere il comportamento delle parole del vocabolario giuridico nei testi. Le due sfaccettature di questa descrizione sono ormai ritenute complementari, anche se per lungo tempo è prevalsa una visione tradizionale della terminologia, consistente nel prospettare soltanto la funzione referenziale dei termini, ridotti a nomi di concetto o di nozione. Questa visione classica, concependo i termini soltanto in quanto nomi, ha largamente influenzato la lessicografia specialistica al punto che i verbi sono unità lessicali generalmente sotto-rappresentate nella maggior parte dei dizionari specialistici (Lerat 2002a, 2002b, Mathieu-Colas 2002). Oggi invece si riconosce unanimemente che una lessicografia terminologica sia necessaria e che debba lasciare "une place aux termes […] en tant que mots" e non esclusivamente "en tant que dénomination de connaissances" (Lerat 1994: 585) mentre la terminologia deve essere concepita "como una materia che permite un acceso múltiple a su objeto [Ndlr i termini], con lo che se ha legitimado una concepción lingüística del mismo" (Cabré 2007: 83) al fine di poter soddisfare le esigenze di formazione.

---

accusata all'epoca di "insuffisante adaptation à l'intégration juridique européenne".

Così, la necessità di riconoscere al verbo uno spazio paritario accanto ai sostantivi terminologici nel dizionario giuridico si è recentemente imposta al giurislinguista come una necessità inderogabile. Nella prefazione del prezioso *Vocabulaire juridique* G. Cornu afferma :

> L'innovation essentielle de ce nouvel ouvrage est d'y avoir introduit les verbes et des adjectifs. Non pas tous. Mais au moins les verbes forts qui énoncent les actions primordiales des principaux protagonistes du Droit (législateur, juge, contractants) et les adjectifs spécifiques les plus courants. (Cornu 2007: VIII)

Nel dizionario monolingue giuridico, accanto ai sostantivi della terminologia giuridica si sono posizionati dei verbi, parole altamente polisemiche, la cui polisemia è in parte imputabile alla variabilità della costruzione sintattica. È determinante per il traduttore ed il redattore possedere una padronanza di questa complessità ed una chiara conoscenza dei parametri in grado di risolvere i casi di polisemia. Occorre quindi anche che il dizionario giuridico bilingue francese-italiano-francese dedichi un trattamento lessicografico funzionale ai verbi giuridici.

Il metodo adottato per la descrizione linguistica dei predicati giuridici e delle relazioni di equivalenza interlinguistiche che essi mantengono tra loro è quello del lessico-grammatica ed i prolungamenti teorici che ne ha proposto il metodo delle classi di oggetti e di predicati. Vedremo che questo approccio propone un'analisi trasformazionale e riformulativa che consente una prima descrizione del senso linguistico dei predicati attraverso la nozione di uso predicativo e di proprietà (morfologiche, sintattiche, trasformazionali) di tale uso. La descrizione dei *prédicats juridiques* si effettuerà per mezzo di un corpus bilingue francese-italiano omogeneo sia dal punto di vista della tipologia testuale che del contesto del diritto, il che è a nostro parere un fattore facilitante per la descrizione del comportamento linguistico affidabile delle parole della lingua giuridica.

1. Cos'è un predicato giuridico?

Senza temere l'evidenza lapalissiana diremo che un predicato giuridico è una parola predicativa che appare in un testo giuridico, nel nostro caso le sentenze della Corte di giustizia dell'Unione europea.

Ispirandosi alla concezione logica di Frege, Gross (1981) definisce una parola predicativa come una funzione, oppure operatore logico[3], che ammette un numero determinato di variabili, i suoi argomenti (soggetto, oggetto1, oggetto2…), senza che nessuno di questi rivesta uno statuto particolare, a differenza delle analisi sintattiche dicotomiche tradizionali che conferiscono al soggetto una posizione dominante. Questo approccio teorico – il lessico-grammatica – animato da un progetto di dizionario elettronico leggibile dal computer, stabilisce che la parola isolata non è un'unità di analisi adatta alla descrizione lessicografica del senso e la sostituisce con la frase semplice, unità di analisi minimale non ambigua. La descrizione delle forme linguistiche è esplicita e si vuole esaustiva: i predicati (i.e. delle strutture di frasi semplici) sono organizzati in classi secondo tavole di proprietà sintattiche comuni (costruzione sintattica degli argomenti, trasformazioni harrisiane e riformulazioni ammesse, cancellabilità di determinati argomenti, possibilità o meno di selezionare dei predicati (parola o frase) in posizione di argomenti).

Nella prima fase, la descrizione è essenzialmente lessico-sintattica, per poi essere precisata e arricchita con l'introduzione di una componente lessico-semantica: le *classes d'objets* (cfr. Le Pesant *et al.*: 1998). Il predicato seleziona i suoi argomenti secondo criteri vincolanti rispetto al loro numero, alla loro costruzione sintattica e alla loro appartenenza a *classes d'objets* semantiche linguisticamente identificabili e caratterizzanti per il predicato (Gross 1999: 30)[4].

---

3 Per una definizione dettagliata della nozione di operatore logico e quella di predicato che gli è associata vedere Gross (1981).

4 "Les classes sémantiques d'arguments déterminent la nature sémantique du prédicat". Si tratta di classi semantiche costituite in estensione a partire dalla

L'insieme di queste restrizioni determina uno *schéma prédicatif* (Gross) o (Lerat) (ormai 'schema argomentale'), matrice lessico-sintattica comune all'insieme delle frasi semplici che costituiscono, in estensione, una classe degli usi del predicato.

Così il predicato *saisir* ammette tre usi specializzati che si possono descrivere con tre schemi argomentali distinti, di cui due sono predicati giuridici :

1) Informatica*: Numano saisir des <données> dans <logiciel>*
2) Diritto: *< Numano: officier de justice> saisir <bien> à <débiteur> pour <créancier>*
3) Diritto: *<partie> saisit une <juridiction> de <affaire> par <requête>*

Il verbo *succomber* ammette tre schemi argomentali, uno dei quali (3) caratterizza un uso specialistico appartenente al linguaggio giuridico:

1) *Numano succomber à <blessure>*
2) *Numano succomber à <émotion>*
3) Diritto*: <partie> succomber*

La descrizione del funzionamento dei predicati nell'ambito della frase semplice offre un approccio razionale e concreto alla polisemia che consiste nell'identificare nelle strutture linguistiche le soluzioni formali che consentono di superare le ambiguità di senso. Il senso considerato è inteso come risultante dalle relazioni delle forme linguistiche tra di loro e non come il significato scaturito da una convenzione di denominazione la quale associa un termine ed un concetto, come succede per le terminologie. Tale senso linguistico si sovrappone alla nozione di uso. La descrizione consiste nell'identificare e delimitare i vari usi di una forma predicativa per mezzo di criteri discriminanti.

---

loro cooccorrenza con dei *verbes appropriés* e che possono apparire nelle posizioni di argomenti dello schema predicativo. La loro funzionalità consiste nel fatto che risultano da una descrizione integrata della semantica da parte della lingua e non da una classificazione preliminare delle parole basata su una descrizione ontologica del mondo, così come succede abitualmente in TAL.

Delimitare un uso consiste allora:

- nel determinare il numero e la costruzione degli argomenti del predicato (*saisir1*, *saisir3* sono dei predicati a 3 argomenti, *saisir2* a 4 argomenti: soggetto, oggetto costruito direttamente, 1, 2 o 3 oggetti preposizionali rispettivamente *dans* Nlocativo, *de* Nastratto, *à* Numano / *pour* Numano); *succomber1*, *succomber2* sono dei predicati a 2 argomenti, *succomber3* a 1 argomento.
- nel recensire esaustivamente le *classes d'objets* lessico-semantiche eleggibili nelle varie posizioni argomentali degli schemi predicativi : *saisir1* (Numano, <*données*>, <*texte*>); *saisir2* (Nprédicativo < Numano: *officier de justice*>, <*bien mobilier*>, <*bien immobilier*>, < Numano: *débiteur*>, < Numano: *créancier*>); *saisir3* <*partie*>, <*juridiction*>, <*affaire*>, <*requête*>); *succomber1* (Nhum, <*blessure*>), *succomber2* (Numano, <*émotion*>), *succomber3* (<*partie*>). La categorizzazione semantica degli oggetti in *classes d'objets* è operata partendo dall'esistenza di uno o più verbi appropriati che ammettono essi stessi gli elementi della classe come argomento: *analyser* <*données*>*; posséder* <*bien*>; <juridiction> *statuer*; *juger* <*affaire*>; *introduire* <*requête*>; <*blessure*> *cicatriser; éprouver* <*émotion*>; *entendre* <*partie*>.
- nell'individuare le forme morfologiche del predicato che si manifestano nei testi. Infatti la predicazione può trovare nei testi svariate espressioni morfologiche di una radice comune. Le relazioni di significato linguistico esistenti tra le varie realizzazioni morfologiche del predicato: il verbo, il sostantivo predicativo, l'aggettivo o il participio si esprimono con regole di riformulazione e di trasformazione operanti tra le frasi semplici che essi generano. Tali frasi semplici semanticamente equivalenti sono l'espressione delle proprietà del predicato e delle sue varianti funzionali all'interno di un determinato uso.

Così la frase semplice *Pierre saisit1 les données dans Open Office* dà luogo alle seguenti riformulazioni:

Les données (sont) saisies dans Open Office par Pierre.

Pierre EFFECTUE LA SAISIE des données dans Open Office.

La frase *L'huissier saisit2 les biens à la demande du fisc* può essere riformulata nella maniera seguente:

Les biens sont saisis par l'huissier à la demande du fisc.
L'huissier PROCÈDE À LA SAISIE des biens à la demande du fisc.

La frase *La partie civile saisit3 le juge d'instruction d'une demande d'acte* può essere riformulata come segue:

Le juge d'instruction est saisi d'une demande d'acte par la partie civile.
La partie civile PROCÈDE À LA SAISINE du juge d'instruction.

L'approccio logico imposto dal metodo implica che l'istanziazione del predicato sia esso verbale, nominale, aggettivale risulti completa al livello delle frasi semplici, ossia che tutti gli argomenti siano espressi in maniera esplicita con un'espressione linguistica. Infatti, il trattamento automatico, a differenza del cervello umano, opera esclusivamente su forme attestate e non su forme sottintese. Ora, nella realtà dei testi e dei discorsi, succede spesso che alcuni degli argomenti del predicato non siano esplicitati perché, per esempio, sono già stati nominati in precedenza. In virtù di un'economia naturale dei discorsi che porta ad evitare la ridondanza, la forma predicativa è quindi attuata in una realizzazione parziale dello schema argomentale costitutivo del suo uso[5].

Essa può, per questo verso, essere semanticamente ambigua, dato che vengono allora a mancare i primi indizi linguistici formali che consentono di risolvere la polisemia dell'espressione. La ricostruzione dello schema argomentale e delle proprietà dell'uso a partire dagli indizi forniti dal contesto più o meno prossimale (frase, paragrafo, testo) è precisamente uno dei compiti essenziali per poter eliminare ogni ambiguità semantica legata a fenomeni propriamente

5 Lerat (2002: 158, 159) parla di *condensation* e di *condensés de schémas d'arguments*, concetto che utilizza per spiegare il funzionamento predicativo dei *dérivés spécialisés* che sono i predicati nominali e aggettivali e la loro proliferazione nella comunicazione specialistica.

linguistici e non concettuali.

Di quali indizi si tratta? Della variante morfologica specifica eventuale, del verbo supporto specifico e degli aggettivi appropriati, dei sintagmi derivati e lessicalmente cristallizzati.

Così l'esistenza o la forma del nome predicativo derivato può essere discriminante di certi usi dei verbi polisemici *saisir*[6] o *succomber*:

saisir1 / 2 <données> / <biens> → *la saisie*
saisir3 <juridiction> → *la saisine*
succomber1 / 2 → *ø*
succomber3 → *la succombance, il succombant, la partie succombante*

Quando la forma del nome deverbale è comune a più usi e di conseguenza polisemica (cfr. *saisie*) essa può a sua volta veder rimossa l'ambiguità semantica grazie a determinate proprietà caratterizzanti quali per esempio il verbo supporto appropriato, il cui ruolo grammaticale è quello di garantire al nome predicativo che siano espresse nella frase le categorie verbali che esso non può autonomamente veicolare – persona, numero, tempo, modo – e che sono indispensabili all'attuazione della predicazione nominale :

saisir 1 <données> → *effectuer la saisie (*lever la saisie)*
saisir 2 <biens> → *procéder à la saisie ; lever la saisie*
saisir 3 <juridiction> → *procéder à la saisine ; (*lever la saisine)*

Vi contribuiscono anche gli aggettivi appropriati al nome ed i nomi complessi derivati:

saisir1 → *saisie (vocale, automatique, manuelle) de* <données> ;

---

6 È qui escluso l'uso psicologico di *saisir* caratterizzato dal nome predicativo *saisissement.* A proposito della specificità del nome predicativo deverbale in usi specialistici, vedere l'esempio del verbo *passer* e dei suoi nomi deverbali predicativi: *passe, passée, passation, passage, passement* che condividono più di dieci domini distinti estesi dal diritto alla tauromachia attraverso la pesca ed il cappellificio (Lerat 1994: 584).

→ *données saisies (*insaisissables, *désaisies*)
→ *opération de saisie* ; *opérateur de saisie*

saisir2 → *saisie (exécutoire, conservatoire, *manuelle) de* <biens> ;
→ *biens saisis, saisissables, insaisissables ;*
→ *partie saisie, (débiteur, tiers) saisi (*dessaisi) ;*
→ *procédure de saisie,* (*saisie immobilière, saisie-attribution*)

saisir3 → *saisine de (juge, tribunal, cour) par requête*
→ *saisine simplifiée*
→ *juge (saisi, dessaisi, *saisissable) d'une affaire ;*
→ *droit de saisine*

Le relazioni di sinonimia intra- e interlinguistica (Gross 1998) sono anch'esse un mezzo di caratterizzazione degli usi del predicato. Non si tratta di sinonimia di parole ma di sinonimia di usi di parole, ossia di frasi semplici semanticamente equivalenti nel contesto di uso preso in considerazione. Infatti la sinonimia è valutabile soltanto nel contesto della frase semplice, dato che quest'ultima "peut toujours constituer une unité de sens autonome non ambiguë" (Gross 1981).

- relazioni sinonimiche intralinguistiche
  ordonner 1 → *Numano METTRE EN ORDRE Nc (Nc être en ordre; Nc être ordonné)*
  ordonner 2 → *Numano DONNER UN ORDRE à <Numano: subordonné>*
  ordonner 3 → *<juridiction> PRESCRIRE PAR ORDONNANCE une <mesure d'instruction>*

- relazioni sinonimiche interlinguistiche
  saisir des données → *INSERIRE dati*
  saisir un bien → *SEQUESTRARE un bene (saisie conservatoire → SEQUESTRO)*
  → *PIGNORARE un bene (saisie exécutoire → PIGNORAMENTO)*[7]
  saisir un juge → *ADIRE un giudice*

7 Cfr.: *Dictionnaire de droit privé français*, Serge Beaudot, Traduzione italiana accessibile all'indirizzo : <http://www.juripole.fr/Dictionnaire/index.php>.

Come si evince, la relazione sinonimica interlinguistica consente allora di rimuovere un'ambiguità nell'uso del predicato giuridico *saisir2*, il quale non distingue lessicalmente due concetti giuridici a differenza dell'italiano (*sequestrare* vs *pignorare*).

Oltre ai vantaggi che esso offre nel campo dell'ingegneria della lingua e delle sue applicazioni pratiche (ingegneria documentaria, redazione di testo assistite via computer), questo metodo descrittivo degli usi predicativi presenta un grande vantaggio per la didattica della lingua e della traduzione giuridica dato che propone un censimento esaustivo dei comportamenti linguistici differenziali dei predicati usati nei testi autentici. Il fatto di applicarlo ad un corpus bilingue costituisce la prima tappa per l'individuazione di queste proprietà che non possono essere sottaciute in un progetto di lessicografia giuridico bilingue.

2. Il corpus bilingue francese-italiano delle sentenze

La nostra ipotesi di lavoro è che, per operare una descrizione affidabile dei fenomeni puramente linguistici, si debba costituire un corpus bilingue che offra delle garanzie solide riguardo all'equivalenza funzionale dei testi, all'equivalenza dei concetti e dei termini giuridici.

Postuliamo quindi che le versioni bilingui delle sentenze della Corte rappresentino una risorsa linguistica privilegiata per l'osservazione, la descrizione e la comparazione dei fenomeni di predicazione che sono in atto nella lingua giuridica in francese e in italiano, dal momento che la sentenza dice il diritto comunitario.

A sostegno di questa ipotesi, proporremo l'idea che il tipo testuale delle sentenze della Corte di giustizia consente di eludere provvisoriamente la questione della specificità dei diritti nazionali ed i problemi di diritto comparato legati all'esistenza di nozioni culturalmente specifiche difficilmente traducibili da un sistema all'altro.

Ora, come afferma Cabré (2007: 83), quando si prospetta la terminologia dal punto di vista linguistico, le unità lessicali terminologiche non sono diverse dalle parole, ma è il loro uso in contesto che consente di caratterizzarle come termini. Di conseguenza i due soli fattori che distinguono le unità terminologiche dalle unità lessicali "no marcadas en cuanto a su condición de terminologicidad" sono "la activación de un sentido preciso y las condiciones pragmáticas de uso". Se questo è vero, allora la costituzione di un corpus di analisi omogeneo che fissa come invarianti e comuni le condizioni pragmatiche di uso per le due lingue giuridiche dovrebbe a nostro parere permettere di considerare che i fatti osservati nell'analisi comparativa sono esclusivamente imputabili al comportamento linguistico delle unità terminologiche e non a sfumature concettuali o nozionali legate a sistemi di riferimento distinti o usi stilistici e pragmatici distinti.

Il corpus costituito per l'analisi linguistica comparata è un corpus bilingue pragmaticamente omogeneo ed è composto di 27 sentenze[8] della Corte di giustizia, una delle tre giurisdizioni[9] della Corte di giustizia dell'Unione europea, selezionate in funzione della lingua di procedura adottata: l'italiano e la loro versione in francese. Con le sentenze pronunciate, l'istituzione contribuisce all'adempimento della sua triplica missione:

- controllare la legalità degli atti delle istituzioni dell'Unione europea;
- verificare il rispetto da parte degli stati membri, degli obblighi che derivano dai trattati;
- interpretare il diritto dell'Unione su richiesta dei giudici nazionali.

Per condizioni pragmatiche di uso intendiamo tre parametri: 1) il contesto giuridico di riferimento dei testi (il diritto comunitario detto nelle due lingue), 2) il contesto di uso sociolinguistico delle due lingue giuridiche (genere testuale unico giurisprudenziale: *arrêts de la Cour* /

8 La sentenza è una risposta ai giustiziabili da parte della Corte di giustizia, allo stesso modo delle ordinanze e conclusioni degli avvocati generali.

9 Accanto al Tribunale ed al Tribunale della funzione pubblica.

sentenze della Corte), 3) lo statuto rispettivo delle due lingue considerate nel contesto di produzione (italiano: lingua di procedura; francese: lingua di lavoro). *Le sens précis activé* in questo contesto risulta allora equivalente nelle due lingue, perciò le condizioni dell'analisi linguistica bilingue sembrano riunite affinché il linguista possa da una parte identificare gli usi terminologici delle parole di ognuna delle due lingue, le relazioni di sinonimia interlinguistica che codesti usi mantengono, e descrivere d'altra parte le proprietà linguistiche di tali usi.

L'ordinamento giuridico di riferimento dei testi, il diritto comunitario, che è il diritto interessato dalla giurisprudenza della Corte di giustizia, è distinto e differente dal diritto nazionale. Oltre le specificità linguistiche e giuridiche nazionali, le lingue giuridiche utilizzate sono modellate per dire in maniera intelligibile l'unicità dei concetti del diritto comunitario:

> Le juriste linguiste sert tout d'abord le droit et le juge communautaire. […] Dans son expression jurisprudentielle tout comme dans son expression législative, le droit communautaire doit démontrer son unicité et son application uniforme dans tous les États membres. Le parallélisme des formes et certains choix expressifs en sont un des garants. (Berteloot 1999: 8)

L'imperativo di unicità concettuale e nozionale incide quindi sull'uso che si fa delle lingue giuridiche e conviene non perdere di vista che l'analisi linguistica comparata verte in qualche maniera su due *eurolectes* giuridici che sono il risultato di un adattamento linguistico delle lingue francese e italiana dicenti il diritto nazionale. Tra i più significativi di tali adattamenti citiamo:

- la soppressione di certe espressioni nazionali giudicate ambigue (cfr. gli *attendus* francesi non tradotti);
- il tentativo costante di esprimere le nozioni nazionali in termini più comuni e traducibili con equivalenti più sicuri nelle altre lingue comunitarie;
- uno sforzo stilistico per lottare contro la complessità privilegiando le frasi corte e facendo un uso molto moderato del latino (cfr. Berteloot 1999).

L'omogeneità espressiva translinguistica dei testi normativi comunitari è quindi una scelta deliberata che mira e contribuisce all'istituzione del diritto sovranazionale comune ai paesi-membri. Sfocia in un parallelismo dei generi testuali giurisprudenziali (quali gli *arrêts de la Cour* / sentenze della Corte) la cui struttura è marcata con formule irrigidite indicanti le parti, lo svolgimento della procedura, la giurisdizione comunitaria adita, le spese processuali. L'idiosincrasia stilistica imputabile sia all'autore del testo sia allo stile giuridico nazionale trova perciò poco spazio. Infatti, considerata la posta in gioco pragmatica e giuridica che comportano gli atti di linguaggio rappresentati dalle sentenze pronunciate, l'espressione linguistica deve ubbidire a fortissimi vincoli formali. La sentenza rientra di fatto nella categoria dei testi delle decisioni che applicano le norme del diritto, se ci si riferisce alla tipologia tripartita dei testi giuridici proposta da Claude Bocquet (2008) e fondata "sur la forme du discours, ou plus précisément sur la logique de ce discours"[10]. L'omogeneità del corpus riguardo al genere testuale ed il carattere fortemente codificato del genere della sentenza accrescono quindi le possibilità metodologiche di operare un'analisi comparativa linguistica affidabile dei predicati.

Infine il contesto di traduzione che produce queste versioni bilingui è del tutto particolare a causa del regime linguistico proprio della Corte che fissa lo statuto speciale di due delle 24 lingue ufficiali nelle quali saranno pubblicate le versioni: la lingua di procedura[11] e la

---

10 Essendo le altre due categorie rappresentate dai testi normativi e dai testi che espongono il contenuto delle regole di diritto (Bocquet 2008:11, citato da Monjean-Decaudin 2010).

11 "Dans les procédures préjudicielles, il s'agit toujours de la langue utilisée par le juge national qui saisit la Cour. Dans les recours directs, le requérant (demandeur) a le choix de la langue de procédure. Il n'est lié ni par sa propre nationalité ni par celle de son avocat. Toutefois, lorsque le défendeur est un État membre ou un ressortissant d'un Etat membre, la langue de procédure est la langue ou l'une des langues de cet État. Une fois la langue de procédure établie, elle doit être utilisée tout au long de la procédure, tant dans les écrits que lors de la procédure orale. Ce choix lie non seulement les parties, mais aussi les 'tiers susceptibles d'être mis en intervention dans la procédure'". (cfr. <http://curia.europa.eu/jcms/jcms/Jo2_10739>)

lingua di lavoro e delle deliberazioni, il francese[12]; la versione della sentenza che fa fede è quella redatta nella lingua di procedura[13]. Per quanto riguarda la lingua di procedura e la lingua di lavoro, occorre precisare che i traduttori giuristi della Corte[14] non traducono i testi delle sentenze, ma intervengono in molteplici occasioni prima della redazione della versione bilingue definitiva. Traducono invece in francese la documentazione redatta in italiano del caso giudiziario da trattare, il che costituisce una pesante responsabilità poiché la traduzione francese iniziale del fascicolo completo condiziona il corretto svolgimento del compito dei giudici e degli avvocati della Corte e può eventualmente essere completata con nuovi documenti, i quali saranno a loro volta tradotti. Le conclusioni sono tratte in francese, lingua delle deliberazioni. Il testo della sentenza in lingua di procedura è l'unico autentico. Per quanto riguarda lo statuto stesso del francese, lingua di deliberazione, e dell'italiano, lingua di procedura, le due versioni definitive della sentenza pronunciata mantengono quindi tra loro un rapporto particolare che risulta da ripetuti riscontri e concertazioni e che rende difficile l'identificazione di una lingua di partenza e di una lingua arrivo, nonché di una direzione di traduzione propriamente detta. Si tratta quindi di versioni linguistiche equivalenti e omologate.

---

12 "La Cour nécessitant une langue commune pour délibérer, cette langue est traditionnellement, mais pas statutairement, le français" (cfr Rapport n° 258 (2008-2009) de M. Jacques Legendre, fait au nom de la commission des affaires culturelles, déposé le 11 mars 2009, http://www.senat.fr). Nella prassi recente, risulta che anche l'inglese è spesso utilizzato comme lingua comune di lavoro.

13 Il testo della sentenza è tradotto in tutte le altre lingue per essere notificato nell'ambito della procedura ai governi di tutti gli stati membri i quali potranno a loro volta intervenire nella propria lingua presentando osservazioni che saranno tradotte in francese per la formazione del giudizio ma anche nella lingua di procedura per essere notificate alle parti (Berteloot 1999: 4).

14 La Direzione generale della traduzione che lavora per le tre giurisdizioni della Corte di giustizia dell'Unione europea è composta di giuristi linguisti titolari di un diploma in scienze giuridiche nonché di linguistica. L'articolo 22 del regolamento di procedura della Corte prevede che il servizio linguistico debba essere "composé d'experts justifiant d'une culture juridique adéquate".

3. Descrizione comparata di alcuni predicati giuridici

3.1. Il vocabolario della sentenza

Il corpus francese delle sentenze contiene 717 tipi lessicali di verbi che danno luogo a 2738 occorrenze di forme flesse. Alcuni di questi verbi sono dei lemmi del Vocabulaire juridique (Cornu 2007) quali *corroborer, allouer, postuler, expliciter, intenter, interjeter*…, altri non sono recensiti in quanto verbi, mentre i loro derivati nominali e / o aggettivali lo sono, quali (*abroger) / abrogation; (adjuger) / adjugé, adjudication; (apurer) / apurement; (citer) / citation; (condamner) / condamnation, condamnatoire*…; altri infine non appartengono alla nomenclatura del Vocabulaire quali *expirer, instaurer, solliciter*.

Come ogni linguaggio specialistico (Gross *et al.* 2001: 3)[15], il linguaggio giuridico giustappone più tipi di vocabolari (Gémar 1991: 277)[16] che vanno dal vocabolario generale al vocabolario specifico del diritto. Il nostro corpus non deroga a questa regola e contiene delle unità lessicali di varia natura di cui si può proporre la seguente tipologia:

a) unità appartenenti alla lingua generale, non specifiche del vocabolario del Diritto e che possono essere riscontrate in testi che non appartengono al campo giuridico. Questo vocabolario di uso corrente è utilizzato per riferirsi alla molteplicità delle cose e delle attività umane che costituiscono la materia del diritto: *lac, remorque, atteler, véhicule, bière…* ma troviamo anche espressioni astratte che manifestano la formulazione di un ragionamento: "*En conséquence, il est nécessaire de…*" "*Les débats au sein du comité portaient sur deux questions*". Infatti il discorso delle sentenze consiste nell'esporre, spiegare,

15 Secondo gli autori "tout texte même le plus spécialisé comprend pour deux tiers de sa surface des mots de la langue générale".

16 La nomenclatura o insieme dei termini del campo, il vocabolario di sostegno (ossia l'insieme dei cooccorrenti dei termini), il vocabolario generale.

ragionare, argomentare e convincere per cui, tra i circa 700 verbi del corpus, la classe semantica dei predicati di parole è fortemente rappresentata: *affirmer, annoncer, critiquer, reprocher, dénoncer, dissuader, persuader, enjoindre, exiger, illustrer, expliquer…*

b) elementi del vocabolario generale cui è assegnato un particolare significato giuridico: nomi come *article, membre, acte, pilier, arrêt, moyen*; verbi come *succomber, attaquer, décider, porter, décerner* … Sono queste le unità lessicali che costituiscono indubbiamente uno dei più alti gradi di complessità per il neofita della traduzione giuridica. Presenti nel vocabolario generale sono semanticamente ambigue dato che in grado di assumere un significato specialistico che la loro forma lessicale non segnala. La competenza richiesta è quindi quella di saper decidere in contesto se il senso attivato è il significato comune o quello specialistico. L'obiettivo della descrizione comparata degli usi è di fornire indizi che aumentino la "*décidabilité"* semantica di queste forme.

c) unità lessicali specifiche del dominio giuridico: nomi di esseri umani (e di professione): *avocat, juge, greffier;* nomi di autorità giurisdizioniali e non *(Commission, Cour, Tribunal),* nomi predicativi (*abrogation, pourvoi…*); verbi (*abroger, surseoir, statuer…*); aggettivi (*précité, susvisé, non-fondé…*).

d) unità lessicali specifiche di un altro campo rispetto a quello del Diritto che sono legate all'ambito del caso giudicato propriamente detto: *rejet des eaux* (inquinamento ambiente) *par une usine* (vs *rejet d'un pourvoi par une juridiction* (giurisprudenza)).

Analizzare e descrivere gli usi predicativi anche tenendo conto del criterio sociolinguistico di ripartizione degli stessi tra lingua generale e lingua giuridica / lingua specialistica produce risultati specifici per ciascuna delle lingue giuridiche considerate. Infatti, né gli usi predicativi della lingua generale, né quelli specialistici della lingua giuridica e a maggior ragione i fenomeni di polisemia si

sovrappongono da una lingua all'altra, come attesta il confronto dei testi delle sentenze.

Se si aggiunge a questa considerazione il fatto che la comparazione bilingue riguarda due lingue quali il francese e l'italiano i cui lessici presentano un alto grado di somiglianza, i risultati della descrizione rivelano una scala di complessità che spazia dai predicati monosemici isomorfi – i quali rivestono rigorosamente lo stesso uso nelle due lingue – fino ai predicati polisemici – i cui usi non si sovrappongono da una lingua all'altra. A questo fatto occorre poi aggiungere l'osservazione che non necessariamente la frontiera tra parola della lingua comune e parola del linguaggio specialistico segue dei tracciati sovrapponibili nelle due lingue. È proprio questo dato che rappresenta una delle fonti di maggiore difficoltà per l'apprendista traduttore.

3.2. Monosemia e sovrapposizione degli usi

I predicati *exciper / eccepire*, per la loro estraneità intralinguistica, sono immediatamente reperibili in quanto specificamente giuridici. La forma predicativa nominale *exception / eccezione* appartiene invece al vocabolario della lingua generale. L'uso della parola francese, caratterizzato nel corpus con un unico schema argomentale, e l'uso del suo equivalente morfologico si ritrovano, dando luogo ad una relazione di sinonimia interlinguistica degli usi che risulta potenzialmente facilitante: somiglianze delle forme morfologiche del predicato (N, V); isomorfie del verbo supporto del nome (*soulever / sollevare*), isomorfia nel numero e nella natura semantica degli argomenti:

FR	IT
<partie> EXCIPER DE <fait>	*<parte> ECCEPIRE <fatto>*
Dans son mémoire en défense, la commission A EXCIPÉ DE l'irrecevabilité du recours en faisant valoir qu'Ente n'était pas, au sens de l'article 230, quatrième alinéa, CE,	*Nel controricorso la Commissione HA ECCEPITO l'irricevibilità del ricorso sostenendo che la decisione controversa non "riguarda[va] direttamente" l'Ente come richiesto*

"directement concerné par la décision litigieuse".	*dall'art. 230, quarto comma, CE.*
<partie> SOULEVER une / l'EXCEPTION de <fait>	*<parte> SOLLEVARE una / l'ECCEZIONE di <fatto>*
l'exception d'irrecevabilité (est) soulevée par la Commission	*l'eccezione di irricevibilità (è) sollevata dalla Commissione*

L'unico vero scoglio sintattico, non trascurabile in questo contesto di trasparenza delle forme lessicali[17], è la differenza di costruzione dell'argomento oggetto (preposizionale in francese, diretta in italiano).

Altro esempio di simmetria di uso e di forme predicative isomorfe è il seguente:

FR	IT
<juridiction>REJETER (recours, plainte) *<juridiction>CONCLURE AU REJET du recours*	*<giudice>RIGETTARE (ricorso, denuncia)* *<giudice>CONCLUDERE PER IL RIGETTO del ricorso*

I falsi amici sono anch'essi presenti naturalmente: predicati monosemici isomorfi rivestono proprietà distinte, in particolare il verbo supporto selezionato:

<juridiction>RENDRE (arrêt)	*<giudice>PRONUNCIARE (sentenza)*
<partie> INTRODUIRE (recours)	*<parte> PROPORRE (ricorso)*
<cadre réglementaire> INSTITUER (distinction)	*<quadro giuridico> INTRODURRE (distinzione)*

3.3. Polisemia e non sovrapposizione degli usi

Più usi di una stessa forma predicativa possono coesistere nel corpus. Nonostante la scelta di un corpus omogeneo, i fenomeni di polisemia

17 Le parole di una lingua ignorata o scarsamente conosciuta sono trasparenti se il loro significante non costituisce un ostacolo per accedere al loro significato. La nozione di trasparenza è stata proposta e definita da Mackey (1965).

non sono assenti. Come tutti gli scritti specialistici, le sentenze sono costituite per i loro due terzi di parole del vocabolario generale (cfr. nota 15), vale a dire due terzi di parole ad elevato grado di polisemia. D'altra parte la lingua giuridica presenta la particolarità che i vocaboli che la compongono sono anch'essi polisemici, pur essendo termini. Cornu (2007: XI) afferma in effetti che "la polysémie des termes [est] une des marques linguistiques essentielles du vocabulaire du Droit" che tale polisemia non è imputabile ai vari sotto-domini del Diritto poiché "très souvent, la distinction des concepts ne coïncide pas avec la classification des disciplines". La descrizione bilingue si prefigge, partendo dalle equivalenze proposte dai testi, di tentare di stabilire degli schemi predicativi bilingui di usi sinonimici e non ambigui tra predicati giuridici. "Le schéma d'arguments représente un niveau réaliste, contrôlable et suffisant, à tout le moins, pour lever beaucoup de polysémies" asserisce Lerat (2002: 160).

Ne diamo un esempio con il predicato *porter atteinte*. La natura semantica dell'argomento soggetto è discriminante in una soltanto delle due lingue, il che costringe allora a riconsiderare la delimitazione dell'uso effettuata in una prospettiva monolingue. Il predicato *porter atteinte à* N ammette come soggetto persone fisiche o morali, fatti. In italiano due predicati distinti sovrappongono questi usi, a seconda che l'argomento soggetto sia umano (*persona fisica, giuridica*) o non umano (*fatto*).

<personne> PORTER ATTEINTE À (intérêts, droits, principes) de N	*<persona>, RECARE / ARRECARE PREGIUDIZIO A (interessi, diritti, principî) di N*
<fait> PORTER ATTEINTE À (intérêts) de N *l'irrégularité A PORTÉ ATTEINTE À ses intérêts*	*<fatto> LEDERE (interessi) di N* *l'irregolarità HA LESO i suoi interessi*

Eppure, dato da un lato il grado di astrazione del discorso giuridico e dall'altro la complessità dei concetti del Diritto, non è sempre facile individuare le classi di oggetti selezionate come argomenti, proprio

perché gli argomenti dei predicati giuridici sono a loro volta dei predicati[18]. Il termine di *juridiction* per esempio, sia esso inteso come *organe institué pour exercer le pouvoir de juridiction* o *mission de juger* oppure *autorité en tel domaine* è in ogni caso un nome predicativo. È qui utilizzato secondo la codifica adottata per le classi di oggetti <*juridiction*>, ma si tratta al meglio di un rappresentante comodo degli argomenti semanticamente appropriati (*Corte, Tribunale, giudice…*) nell'ambito dello schema considerato.

Citiamo tre esempi di predicati nominali che costituiscono degli argomenti di predicati frequenti e che ci sembrano notevoli per il grado di polisemia che mette in luce il confronto delle versioni bilingui. Si tratta per l'italiano di due predicati nominali *motivo* e *spese,* e per il francese del predicato verbale *viser*, che sono anche tutte e tre parole del linguaggio comune.

dépens	*spese*
dépenses	
frais	

motif	*motivo*
moyen	
raison	

viser	*mirare*
	avere ad oggetto
	stabilire
(être) visé	*(essere) diretto a*
	(essere) previsto da
	di cui a
	su cui verte
	di cui trattasi

3.3.1. Spese

Un'unica parola *spese* si sovrappone agli usi di tre nomi predicativi francesi *dépens / dépenses / frais*. Il verbo attualizzatore *sostenere* non consente per l'italofono di discriminare gli usi specialistici, mentre

18 Lerat (2002) propone di parlare di *compatibilités sémantiques* per indicare il rapporto di appropriazione che lega il predicato ed i suoi argomenti.

exposer costituisce l'indizio di un uso giuridico specialistico poiché è distinto dal verbo supporto impiegato nella lingua generale (cfr. *infra* verbo supporto). È l'esistenza di verbi, di aggettivi appropriati e di contesti cristallizzati che consente di decidere in maniera non ambigua quale equivalente di *spese* scegliere in francese.

<partie> EXPOSER DÉPENS *<juridiction> STATUER SUR LES DÉPENS* *<partie> SUPPORTE ses dépens* *La Cour CONDAMNE <partie> AUX DÉPENS*	*<parte> SOSTENERE SPESE* *<giudice> STATUISCE SULLE SPESE* *<ricorrente>SOPPORTARE le spese* *Il giudice CONDANNA <parte> ALLE SPESE*
N.B. *(partie requérante) demande à la Cour de CONDAMNER N AUX DÉPENS.*	N.B. *(ricorrente) chiede alla corte di PORRE LE SPESE A CARICO DI N*
<partie> EXPOSER DÉPENSES *<partie> déclarer des dépenses* **(dépens, frais)* *dépenses (SE RAPPORTER À, ÊTRE AFFÉRENTES À) N* *dépenses ÉLIGIBLES *(dépens, frais)*	*<parte> SOSTENERE SPESE* *<parte> DICHIARARE delle spese* *spese RELATIVE A N* *spese AMMISSIBILI*
<personne> EXPOSER FRAIS *Les FRAIS EXPOSÉS pour soumettre des observations à la Cour, autres que ceux desdites parties, ne peuvent faire l'objet d'un remboursement.*	*<persona> SOSTENERE SPESE* *Le SPESE SOSTENUTE da altri soggetti per presentare osservazioni alla Corte non possono dar luogo a rifusione.*

Depens è il termine giuridico che designa le somme dovute al termine della procedura dalla parte contro la quale la sentenza è stata pronunciata. Appare esclusivamente nelle formule irrigidite della sentenza pronunciata. *Statuer sur les dépens / statuire sulle spese* è la funzione dell'ultimo paragrafo del testo della sentenza. *La Cour condamne … aux dépens* è la formula che rende conto dell'atto performativo pronunciato dal giudice. Occorre notare che, a differenza del francese, le versioni italiane riservano esclusivamente l'espressione *condannare N alle spese* all'atto performativo della Corte. Quando si tratta di una richiesta fatta alla Corte da una delle parti, si utilizza l'espressione nettamente più neutra *porre le spese a carico di N* (*mettre … à la charge de N*).

Dépenses è usato con verbi (*déclarer, se rapporter à*) ed aggettivi appropriati (*éligible, afférent*) che risultano specifici nelle sentenze e che escludono gli altri candidati.

Frais è impiegato esclusivamente in un paragrafo che si ripropone quasi in maniera identica in tutte le sentenze, in cui si designano le somme [impegnate] *da altri soggetti* / *autres que (celles) des partie*s.

3.3.2. Motivo

Un'unica parola, *motivo*, riveste gli usi di tre nomi predicativi francesi *moyen / motif / raison*. *Moyen* et *motif* sono delle parole della lingua comune che, nelle sentenze, entrano in usi specialistici.

<partie> INVOQUER un MOYEN *moyen (soulevé, exposé, produit) par <partie>* *moyen (EXAMINÉ, ACCUEILLI, ÉCARTÉ, REJETÉ) par <juridiction>* *MOYEN DE RECOURS, MOYEN D'ANNULLATION* *La requérante, par acte déposé au greffe…a soulevé DE NOUVEAUX MOYENS* *…les DEUX PREMIERS MOYENS de la requérante* *La République italienne soulève SEPT MOYENS D'ANNULATION dans sa requête*	*<parte> AVANZARE un MOTIVO* *motivo (sollevato, addotto, dedotto, introdotto, prodotto) da <parte>* *motivo (ESAMINATO, ACCOLTO, RESPINTO) da <giudice>* *MOTIVO DI RICORSO, MOTIVO DI ANNULLAMENTO* *La ricorrente con atto depositato presso la cancelleria … ha sollevato NUOVI MOTIVI* *…I PRIMI DUE MOTIVI della ricorrente* *La Repubblica italiana ha sollevato SETTE MOTIVI DI ANNULLAMENTO nel proprio ricorso*
N.B. *EXAMEN des moyens (*ACCUEIL)* *Ce moyen NE PEUT PAS PROSPÉRER*	N.B. *(DESAMINA, ACCOGLIMENTO) dei motivi* *Tale motivo NON PUÒ ESSERE ACCOLTO*
MOTIF *PAR CES MOTIFS, LA COUR dit pour droit…* *PAR CES MOTIFS, LA COUR déclare et arrête…* Par ces motifs, la Cour (quatrième	*MOTIVO* *CON QUESTI MOTIVI, LA CORTE dichiara…* *PER QUESTI MOTIVI, LA CORTE dichiara…* *CON QUESTI MOTIVI, LA CORTE*

chambre) dit pour droit	*dichiara e statuisce…* Per questi motivi, la Corte (Quarta Sezione) dichiara
AU MOTIF QUE…	*PER IL FATTO CHE …(*MOTIVO)*
RAISON *… l'existence de BONNES RAISONS économiques* *Une telle interdiction peut être justifiée par l'une des RAISONS D'INTÉRÊT GÉNÉRAL énumérées à l'article 30*	*MOTIVO* *… l'esistenza di VALIDI MOTIVI economici* *Un tale divieto può essere giustificato da uno dei MOTIVI DI INTERESSE GENERALE indicati nell'art. 30*

Moyen è utilizzato in un modo specifico come lo dimostra il fatto che i verbi e aggettivi appropriati presenti nel corpus (cfr. tabella) non sono gli stessi di quelli impiegati nella lingua generale:

> Langue générale:
> *(avoir, disposer de, se procurer) des moyens*
> *des moyens (humains, matériels, logistiques, financiers)*
> *des moyens sont (nécessaires, suffisants, insuffisants) pour …*
> *un moyen d'action, un moyen de pression…*

Motif appare in formule performative caratteristiche del genere testuale delle sentenze, in unità lessicalizzate specifiche della lingua giuridica *au motif que / per il fatto che.* È con dei *motifs* che il giudice, quando pronuncia il suo giudizio, deve rispondere ai motivi avanzati. È interessante osservare come, all'interno stesso di una formula, espressione irrigidita per eccellenza, la versione italiana presenta un'alternanza preposizionale (*con / per*) che pensiamo poter interpretare come una interferenza con il francese (*per / par*) in seno ad uno scritto considerato normativo.

Raison è utilizzato in un uso generale, come attesta l'aggettivo valutativo (*bonnes* / *validi*).

A livello dell'enunciato, il fatto che una lingua dispone di una parola unica per tradurre quello che l'altra esprime in due o più, comporta conseguenze non trascurabili quando la distinzione di uso è operata all'interno della stessa frase complessa. L'esempio che segue mostra come la distinzione *motifs / moyens* non esiste nella versione

italiana che non può esprimerla lessicalmente e quindi la riduce alla sola occorrenza di *motivi*.

La Commission conclut au rejet du pourvoi, mais sollicite UNE SUBSTITUTION DE MOTIFS à ceux par lesquels l'arrêt attaqué a REJETÉ LES MOYENS DE LA REQUÉRANTE portant sur l'activité d'assistance aux administrations nationales	*La Commissione conclude per il rigetto dell'impugnazione, ma chiede UNA SOSTITUZIONE DEI MOTIVI DELLA SENTENZA IMPUGNATA in relazione all'attività di assistenza alle amministrazioni nazionali nonché all'attività di normalizzazione tecnica.*

3.3.3. Viser

Il verbo *viser* appare in numerosi schemi di argomenti, essi stessi a loro volta predicativi. Questo fenomeno di predicazione sulla predicazione unito al forte grado di astrazione dà luogo ad alternanze sintattiche *viser à Npred, viser à Inf. / viser à ce que P*, nonché ad alternanze preposizionali *visé par l'article / à l'article* l'equivalenza semantica delle quali richiede una delucidazione. Alcune espressioni contenenti i participi *visé* e *visant* sembrano essere, per la loro frequenza, espressioni stereotipate. Non esistono in italiano predicati che concentrino sotto un'unica radice lessicale una tale molteplicità di usi. Il lavoro di descrizione consiste quindi nel rilevare i vari traducenti tentando di mettere in luce gli schemi argomentali attuati o perlomeno le strutture più ricorrenti.

<procédure> VISER <fait>, <acte> À <texte normatif> *<fait>, <acte> VISÉ PAR ce recours*	*<fatto>, <atto>DI CUI AL PRESENTE ricorso*
activités VISÉES	*attività DI CUI TRATTASI*
les motifs d'invalidité VISÉS À la première question	*i motivi di invalidità SU CUI VERTE la prima questione*
<fait> VISÉ À l'article… *obligation VISÉE À (l'article, le paragraphe)* *situation VISÉE AU paragraphe…*	*<fatto> DI CUI A, PREVISTO IN / DA <testo normativo>* *obbligo DI CUI A (l'art., il paragrafo)* *situazione PREVISTA NEL paragrafo*
<texte normatif> VISER <fait>	

<fait> VISÉ PAR l'article *un cas de figure, une situation VISÉE PAR l'article 14*	*in una fattispecie, una delle ipotesi PREVISTA DAL l'art. 14,*
<texte normatif>VISER DES OBJECTIFS *objectifs VISÉS À l'article 33*	*<testo normativo> STABILIRE OBIETTIVI* *obiettivi STABILITI DALL'art. 33*
<texte normatif>VISER UN OBJET de droit *l'article 14 vise une situation de conflit*	*<testo normativo> AVERE AD OGGETTO* *l'art. 14 ha ad oggetto una situazione di conflitto*
<procédure>, <texte normatif> VISER À + INF. *ce règlement VISE À conférer…*	*MIRARE A + INF;* *detto regolamento MIRA A conferire…*
l'enregistrement d'une IGP… VISE, parmi d'autres objectifs, À éviter…	*AVERE LO SCOPO DI + INF* *la registrazione di un'IGP, HA COME SCOPO, tra gli altri, quello DI evitare*
le règlement VISE À ENREGISTRER…	*ESSERE DIRETTO A + NPRED* *regolamento DIRETTO ALLA registrazione…*
<procédure> VISER À NPRED / VISER À CE QUE… *VISER À la modification de la décision, au rejet du recours…*	*<procedura> ESSERE DIRETTO AD OTTENERE NPRED / CHE ESSERE DIRETTO AD ottenere la modifica della decisione, il rigetto del ricorso…*
(pourvoi, exception de N, demande) VISANT À statuer… *accords bilatéraux VISANT À la protection…*	*(ricorso, eccezione de N, domanda) DIRETTO A far statuire…* *accordi bilaterali DIRETTI ALLA tutela…*
<denomination autonyme> VISER N *La notion de «pays» vise soit un État membre, soit un État tiers*	*<denominazione autoreferenziale> RIGUARDARE N* *La nozione di «pays» riguarda sia uno Stato membro, sia uno Stato terzo*

3.4. Verbo supporto appropriato e predicato nominale

Essendo la lingua giuridica molto astratta, vi abbondano le nominalizzazioni. I predicati giuridici sono spesso attualizzati sotto forma di nomi predicativi. Infatti, il discorso specialistico conterrebbe

il doppio di nomi deverbali (in *-tion*) rispetto al discorso generale (Cabré *et al*. 2007)[19]. Di conseguenza, si può ipotizzare che la lingua giuridica sia particolarmente ricca di verbi supporto di nomi. Ora la scelta del verbo supporto non essendo libera, ma invece, nella grande maggioranza dei casi, determinata arbitrariamente dall'uso specifico ad ogni lingua, da una parte e anche dall'uso specifico alla lingua del Diritto, essa rappresenta un'altra fonte di complessità.

Infatti, il verbo supporto giuridico non è necessariamente identico a quello comunemente utilizzato nel registro sostenuto della lingua generale:

FR: Langue générale	FR: Langue juridique des arrêts
INITIER, AMORCER une procédure	*<parte> OUVRIR une procédure*

D'altra parte le due lingue giuridiche possono ricorrere a verbi supporto distinti per lo stesso uso:

FR: Langue juridique des arrêts	IT: Lingua giuridica delle sentenze
<parte> OUVRIR une procédure	*<parte> AVVIARE una procedura (*aprire)*

Un verbo supporto unico in italiano (*proporre*) corrisponde a quattro verbi supporto in francese, sinonimi due a due, la cui scelta è determinata dalla natura dell'argomento preposizionale (*introduire, porter … devant / diriger, former … contre*):

19 Gli autori propongono criteri di caratterizzazione dei discorsi specialistici (senza distinzione dei domini) in termini di frequenza di occorrenza di classi grammaticali. L'analisi di un corpus multilingue (catalano, spagnolo, inglese) rappresentativi di più discorsi (diritto, economia, informatica, ambiente, medicina) di testi specialistici e di testi non specialistici dimostra come il discorso specialistico scritto si contraddistingue dal punto di vista lessicale dal discorso generale scritto par la maggior presenza (circa due volte in più) di nominalizzazioni in *ciòn / tion* (68,5% *vs* 31,5%), e di sintagmi N + aggettivo (60,1% *vs* 39, 9%).

FR: Langue juridique des arrêts	IT: Lingua giuridica delle sentenze
<parte> INTRODUIRE, PORTER un (recours, pourvoi) DEVANT une juridiction	*<parte> PROPORRE un / a (ricorso, impugnazione) DINANZI al giudice*
<parte> DIRIGER, FORMER un (recours, pourvoi) CONTRE <acte> <décision>	*<parte> PROPORRE un / a (ricorso, impugnazione) AVVERSO / CONTRO <atto> <decisione>*

Si può quindi schematicamente supporre che la scelta del verbo supporto si eserciti anch'essa su due livelli:

- intralinguistico: il verbo supporto in lingua generale può essere distinto / non distinto rispetto al verbo supporto specifico giuridico;
- interlinguistico: il verbo supporto in francese / italiano è distinto / non distinto rispetto al verbo supporto in francese / italiano.

La combinatoria di questi due livelli di parametri può dar luogo a casi diversi illustrati qui di seguito.

Gli usi della lingua generale possono essere simmetrici, i predicati isomorfi, mentre solo una delle due lingue introduce un verbo supporto specialistico specifico (*ouvrir*):

	Linguaggio comune	Linguaggio giuridico delle sentenze
FR	*<situation> DONNER DROIT À N*	*<situation> OUVRIR DROIT À Npréd* *les cas de figure OUVRANT DROIT À la déduction de la TVA*
IT	*<condizione>DARE DIRITTO A N*	*<condizione> DARE DIRITTO A Npred.* *le ipotesi che DANNO DIRITTO ALLA detrazione dell'IVA*

Oppure le due lingue introducono verbi supporto specialistici distinti (*instituer / introdurre*):

	Linguaggio comune	Linguaggio giuridico delle sentenze
FR	*distinguer X de Y* *(FAIRE, OPÉRER) une distinction entre X et Y*	*<cadre réglementaire> INSTITUER une distinction entre X et Y*
IT	*distinguere X da Y* *(FARE, OPERARE) una distinzione fra X e Y*	*<quadro giuridico> INTRODURRE una distinzione fra X e Y*

Gli usi della lingua generale possono invece non essere simmetrici, i predicati non sono isomorfi (*effectuer, engager / sostenere*) e solo una delle due lingue introduce un verbo supporto specialistico specifico rispetto alla lingua generale (*exposer*):

	Linguaggio comune	Linguaggio giuridico delle sentenze
FR	*<Numano> (EFFECTUER, ENGAGER) des dépenses*	*<partie requérante> EXPOSER des dépenses*
IT	*<Numano>SOSTENERE delle spese*	*<ricorrente> SOSTENERE delle spese*

Oppure le due lingue introducono verbi supporto specialistici distinti (*opérer / compiere*):

	Linguaggio comune	Linguaggio giuridico delle sentenze
FR	*<Numano> CONSTATER <fatto>*	*<juridiction><jugement> CONSTATER <fait > les deux jugements … ont constaté l'existence de …*
	FAIRE des CONSTATATIONS	*<juridiction><jugement> OPÉRER des CONSTATATIONS les constatations ainsi opérées par rapport à cette situation…*
IT	*<Numano> ACCERTARE <fatto>*	*<giudice>, <sentenza> ACCERTARE <fatto> le due sentenze … hanno accertato l'esistenza di …*
	FARE un ACCERTAMENTO	*(giudice, sentenza) COMPIERE un ACCERTAMENTO l'accertamento così compiuto in ordine alla situazione …*

Al termine di questo breve percorso illustrativo sembra importante sottolineare il valore informativo del verbo supporto specifico nell'individuazione di un uso specialistico giuridico quando il nome predicativo (cfr. *depenses, constatations*, ecc.) appartiene anche alla lingua generale. Questi pochi esempi bastano per misurare la complessità inerente la scelta del verbo supporto di nome predicativo

giuridico in una prospettiva bilingue francese-italiano, ossia in lessici che presentano forti somiglianze formali.

4. Conclusione

I didatti della lingua giuridica, della traduzione giuridica, della redazione di testi giuridici hanno oggi più che mai bisogno di risorse lessicografiche e terminologiche che recensiscano con precisione i comportamenti linguistici delle unità lessicali complesse: i predicati giuridici. La descrizione della predicazione per mezzo di schemi argomentali e di classi di oggetti predicativi si rivela funzionale per una descrizione contrastiva delle forme della lingua delle sentenze: offre un metodo di risoluzione razionale dei fenomeni di polisemia linguistica dei termini, rende possibile l'identificazione degli usi predicativi sinonimici nell'altra lingua e le condizioni linguistiche di tale sinonimia.

Utilizzando un corpus bilingue omogeneo, tanto dal punto di vista del contesto del diritto quanto da quello della tipologia testuale (decisioni che applicano le norme del diritto), ci sembra possibile controllare la complessità dell'analisi inerente da una parte la diversità dei concetti giuridici delle varie culture nazionali e, d'altra parte, la polisemia terminologica intralinguistica caratteristica della lingua giuridica, il che offre una prima garanzia che i fatti osservabili siano correttamente interpretabili.

L'approccio che abbiamo illustrato consente di recensire fatti linguistici la cui incidenza è spesso sottovalutata nella comprensione e soprattutto nella redazione di testi giuridici, di apportare un complemento lessicografico fondamentale alla descrizione delle unità linguistiche della lingua giuridica, nonché di integrare la descrizione dei predicati della lingua generale e della lingua specialistica.

Sono state così abbozzate alcune possibili piste di esplorazione per una categorizzazione tipologica dei fenomeni interlinguistici di equivalenza degli usi. Sono apparse due direttrici che potrebbero

fornire strumenti di oggettivazione e ulteriore esplorazione della complessità: quella del rapporto tra la sinonimia interlinguistica degli usi e il grado di somiglianza formale interlinguistica che i significanti delle due lingue mantengono; quella, infine, del confronto tra i rapporti che ogni lingua istaura in maniera idiosincratica tra forme significanti della lingua generale e usi giuridici specialistici.

Spostando l'attenzione dal versante della descrizione dei fenomeni intra- e interlinguistici a quello più applicato della didattica del linguaggio giuridico e della traduzione giuridica, ci preme sottolineare quanto l'uso dei corpora bilingui – specie se si tratta di corpora omogenei e omologati da giuristi e giurilinguisti – sia utile per svolgere un'azione formativa efficace e potente, che sia in presa diretta con il contesto professionale del diritto consentendo allo studente di esporsi a volumi consistenti di documenti autentici.

Questo tuttavia presuppone che ci si possa avvalere di strumenti nonché di competenze informatiche che consentano di effettuare il trattamento automatico dei corpora per poter creare delle risorse mirate e abbondanti per la didattica ma anche per poter mettere lo studente nelle condizioni di crearsi le proprie risorse (banca dati, glossari personali, ecc.).

Per illustrare quanto appena detto, presentiamo qui di seguito alcuni dati che abbiamo selezionati fra i segmenti ripetuti più frequenti del nostro corpus bilingue[20]. Non è qui la sede per farne l'analisi. Diremo soltanto che costituiscono a nostro parere, per le loro caratteristiche sia quantitative (il numero di occorrenze (N), la lunghezza dei segmenti) che qualitative (la loro natura e struttura linguistica), uno spunto molto fecondo, in ambito didattico, per una

20 L'analisi dei corpora è stata effettuata con TaLTaC2 ("Trattamento automatico Lessicale e Testuale per l'analisi del contenuto di un Corpus", Bolasco 2010), un software che consente di sottoporre i testi ad un approccio lessicometrico. Fra le altre funzioni statistiche, tale strumento permette di caratterizzare quantitativamente il vocabolario del corpus (in termini di forme grafiche e del loro rispettivo numero di occorrenze), di estrarne determinate forme grafiche, categorizzarle e lessicalizzarle, e di individuare i segmenti ripetuti nel corpus (stringhe di parole che occorrono più volte) corredati dal rispettivo numero di occorrenze.

riflessione sull'essenza del linguaggio giuridico, della lingua delle sentenze della Corte e dei fenomeni linguistici e pragmatici che li costituiscono.

Segmento	N		Segmento	N
ai sensi dell'art	124		au principal	131
sentenza impugnata	105		il y a lieu de	80
reti da posta derivanti	52		filets maillants dérivants	52
regolamento di procedura	51		aux obligations qui lui incombent	37
di detta	43		de la chose jugée	32
di cui trattasi nella	41		la sécurité de la navigation	32
di cui trattasi nella causa principale	38		responsabilité civile automobile	32
dazi doganali	36		manqué aux obligations qui lui incombent	31
È venuta meno agli obblighi	34		considérée comme	31
responsabilità civile auto	32		ayant élu domicile à Luxembourg	30
interessi di mora	32		la question de savoir	29
appalti pubblici	30		denrées alimentaires	28
con domicilio eletto in Lussemburgo	29		par ces motifs	27
agli obblighi ad essa incombenti	27		l' ordre juridique	26
soggetti passivi	18		la taxe régionale d' escale	18

Riferimenti bibliografici

Bocquet, Claude 2008. *La traduction juridique : fondement et méthode*. Bruxelles: De Boeck.

Bolasco, Sergio 2010. *TalTaC2.10, Sviluppi, esperienze ed elementi essenziali di analisi automatica dei testi*. Milano: LED.

Bertheloot, Pascale 1999. La traduction juridique dans l'Union européenne, en particulier à la Cour de justice. <http://www.tradulex.org/Actes2000/berteloot.pdf>

Cabré, Maria Teresa 2007. Términos y palabras en los diccionarios. In Cuartero Otal, Juan / Emsel, Martina (eds.). *Vernetzungen: Bedeutung in Wort, Satz und Text. Festschrift für Gerd Wotjak zum 65. Geburtstag*. Frankfurt am Main: Peter Lang, 71-84.

Cabré, Maria Teresa / Bach, Carme / Castellà, Josep / Marti, Jaume. 2007. La caracterización lingüística del discurso especializado in Mairal Ricardo *et al*. (eds.). Aprendizaje de lenguas, uso del lenguaje y modelación cognitiva: perspectivas aplicadas entre

disciplinas. Actas del XXIV Congreso Internacional de AESLA. Madrid: UNED-AESLA, 851-857.

Cohen-Tanugi, Laurent 1993. À propos des considérations générales du Conseil d'Etat sur le droit communautaire. *Rapport public du Conseil d'Etat, Politique étrangère*. Paris: La Documentation française, 58 / 3, 768-770.

Cornu, Gérard 1987, 2007[8]. *Le Vocabulaire juridique*. Association Henri Capitant. - 8. éd. mise à jour. Paris: PUF.

Escoubas-Benveniste, Marie-Pierre 2008. À quoi tient la synonymie du verbe? Le cas de quelques prédicats relationnels abstraits. In Ciagada, Sergio / Verna, Marisa (eds.) *La Sinonimia tra langue e parole nei codici francese e italiano*. Milano: Vita e Pensiero, 481-503.

Gémar, Jean-Claude 1991. Terminologie, langue et discours juridiques. Sens et signification du langage du droit , Meta 26-1, 275-283.

Gross, Gaston 1999. La notion d'emploi dans le traitement automatique. *La pensée et la langue*, Wydawnictowo Naukowe AP., Krakow, 24-35.

Gross, Gaston 1998. Pour une véritable fonction "synonymie" dans un traitement de texte. *Langages* 131, 103-114.

Gross, Gaston / Mathieu-Colas, Michel 2001. Description de la langue de la médecine. *Meta* 46 / 1, 68-81.

Gross, Maurice 1981. Les bases empiriques de la notion de prédicat sémantique. *Langages* 63, 7-49

Le Pesant, Denis / Mathieu-Colas Michel 1998. Introduction aux classes d'objets. *Langages* 131, 6-33.

Lerat, Pierre 1995. *Les langues spécialisées*, Paris: Presses universitaires de France.

Lerat, Pierre 2002a. Qu'est-ce qu'un verbe spécialisé ? Le cas du droit. *Cahiers de lexicologie* 80, 201-211.

Lerat, Pierre 2002b Vocabulaire juridique et schémas d'arguments juridiques. *Meta* 47 / 2, 155-162.

Lerat, Pierre 2006. Terme et micro-contexte. Les prédications spécialisées. In *Mots, termes et contextes*, Blampain, Daniel / Thoiron Philippe / Van Campenhoudt, Marc (eds), Paris: AUF, 89-98.

Mackey, William Francis 1965. *Language Teaching Analysis*. London: Longman.

Mathieu-Colas, Michel 2002. La représentation des verbes dans un dictionnaire électronique. De la langue générale aux langues spécialisées. *Cahiers de lexicologie* 81, 51-67.

Monjean-Decaudin, Sylvie 2010, Approche juridique de la traduction du droit, CEJEC-wp, 2010 / 1 http://www.cejec.eu/category/publications-en-ligne-working-papers.

I corpora in contesti professionali

MICAELA ROSSI

13. Définition de nouvelles terminologies et communautés de professionnels : analyse de corpus en ligne dans le domaine de la dégustation du vin

1. Introduction

Les études récentes en terminologie, à partir de l'approche socio-terminologique, en passant par la terminologie socio-cognitive et la terminologie variationniste, ont progressivement abandonné la conception des terminologies de spécialité comme des nomenclatures universelles et partagées, revalorisant les processus discursifs et sociolinguistiques qui sous-tendent la création et l'implantation de nouvelles dénominations au sein des communautés de pratique professionnelle. L'étude approfondie de ces processus, désormais réalisable à l'aide des technologies de l'analyse de corpus, s'avère alors fondamentale afin de retracer le parcours des termes nouveaux et les dynamiques d'interaction qui président à la formation de nomenclatures partagées.

Notre contribution sera focalisée sur l'analyse d'un corpus composé d'échanges provenant de forums spécialisés dans le domaine de la dégustation du vin ; nous analyserons en particulier les commentaires métalinguistiques et métaterminologiques des usagers. Le discours des professionnels sur la terminologie qu'ils utilisent nous permettra de vérifier la traçabilité des processus de création et d'implantation des termes ; l'analyse privilégiera notamment les commentaires axés sur la définition des termes et concepts spécialisés, dans la lignée des études sur la définition naturelle, afin de vérifier

dans le détail la phase de négociation collective du sens et des termes dans les communautés professionnelles.

2. De l'importance des corpus dans l'analyse terminologique

En terminologie, les acquis de la linguistique de corpus ne sont que très récents (voir entre autres Williams, en ligne) : la nature essentiellement prescriptive et contraignante de la théorie d'empreinte wüstérienne a influencé pendant longtemps la discipline ; ce n'est que pendant les dernières décennies que la terminologie s'est progressivement intéressée à la dimension textuelle et variationniste, s'ouvrant progressivement à l'étude de vastes corpus textuels. Il suffit de citer à ce propos les courants de la terminologie textuelle de Bourigault et Slodzian, mais également la terminologie communicative de Maria Teresa Cabré, sans oublier l'apport fondamental de la socioterminologie proposée par Gaudin ou de la socioterminologie cognitive de Rita Temmerman. Les analyses de corpus se sont progressivement intensifiées au cours des dernières années, permettant ainsi des opérations de plus en plus fines de fouille terminologique, par exemple pour la recherche de néologismes ou de variantes attestées, mais également pour des analyses approfondies dans le domaine de la linguistique spécialisée ; entre autres, le travail pionnier de Condamines et Aussenac-Gilles (2006) met en évidence les liens multiples entre terminologie et linguistique de corpus.

Désormais, l'importance de l'analyse de corpus dans la terminologie contemporaine n'est donc plus à démontrer : les nouveaux courants de la discipline terminologique ont su profiter des acquis de la linguistique computationnelle, ainsi que des progrès du traitement automatique des langues, pour affiner et élargir l'analyse des termes dans les textes et discours authentiques. La linguistique de corpus est devenue essentielle pour le travail du terminologue (voir à ce propos Condamines, Aussenac-Gilles 2006, en ligne, ainsi que

Condamines 2005 ou, plus récemment, le numéro spécial de la revue *Meta* dirigé par Van Campenhoudt et Temmerman 2011), dans ses multiples fonctions :

- extraction terminologique des listes de candidats termes pour la création de glossaires et répertoires ;
- analyse de corpus textuels amples pour la recherche linguistique de phénomènes discursifs propres à la communication spécialisée ;
- aides pour la traduction assistée par ordinateur, ainsi que pour la rédaction technique ;
- création d'ontologies et ressources documentaires (thesaurus, bases de connaissances terminologiques…).

Les études récentes se focalisent en particulier sur les conditions de constitution des corpus d'analyse textuelle, sur les critères linguistiques et extra-linguistiques présidant à la sélection des textes, ainsi qu'à leur traitement. Plus en particulier, l'objet des réflexions des terminologues est la notion de *corpus spécialisé*, pris en considération par opposition à corpus de langue générale sur la base de paramètres éminemment contextuels, tels que :

- le statut professionnel des participants à l'échange (spécialistes, amateurs, apprenants…) ;
- le degré de spécialisation du sujet traité ;
- les genres textuels et discursifs, ainsi que le canal de communication utilisé.

La constitution du corpus spécialisé devra en outre respecter des critères fondamentaux comme l'*équilibre* entre les différents paramètres considérés – afin de pouvoir offrir une description complète du domaine concerné – et la *représentativité* au sein du domaine thématique choisi (voir aussi Biber 1995). Les critères d'identification d'un corpus spécialisé peuvent alors être résumés dans une matrice d'analyse, que nous avons élaborée sur la base des paramètres identifiés par Biber (1995 : 380 et suiv.), adaptés en fonction des études terminologiques :

Paramètres	**Remarques**
Canal (écrit / oral)	les corpus terminologiques exploitent dans la plupart des cas des textes écrits, mais les analystes mettent en évidence la nécessité d'étudier également la dimension orale, dont l'importance dans la communication spécialisée est capitale
Format (public / non public)	aux fins de l'analyse terminologique, les documents internes aux entreprises (ce que l'on appelle la *littérature grise*) peuvent apporter des informations intéressantes
Cadre (institutionnel / personnel / informel)	le cadre de la communication tend à influencer la production et l'usage de termes sur l'axe diaphasique (termes strictement scientifiques / techniques, termes de vulgarisation, variantes géographiques / dialectales)
Destinataires (statut, compétences)	le statut de destinateur et destinataires de la communication a une incidence sur la densité informative de la communication, ainsi que sur la terminologie utilisée
Connaissances requises et partagées	s'agit-il de connaissances précises et spécialisées ? quel degré de compétence spécialisée est-il requis ?
Destinateur (statut, compétences)	voir *destinataires*
Fonction dominante du texte	s'agit-il d'un texte à visée informative, didactique… ?
Thématique du texte	domaine concerné

Tableau 1 : Matrice d'analyse.

Cette matrice sera exploitée dans les pages suivantes afin de fournir une description complète du corpus analysé.

3. L'Internet comme corpus : *Web as a corpus or Web for corpus* ?

L'avènement d'Internet a produit un changement de paradigme fondamental dans la linguistique computationnelle : la possibilité d'avoir accès à une quantité infinie de données linguistiques a ouvert des possibilités inexplorées pour les analystes (Fouquéré / Issac 2003, Fairon 2011), permettant aussi la création de vastes corpus disponibles en Réseau. Entre autres, le projet Webcorp[1] offre aux consultants la possibilité d'avoir accès à des recherches de corpus ayant comme base la globalité de la Toile. L'Internet peut être considéré à tous les effets comme un corpus énorme, indéfini en termes quantitatifs précis (ce qui dépend évidemment de sa nature partiellement éphémère et en constante évolution), présentant des occurrences linguistiques dans différentes typologies textuelles au sens large (textes informatifs, échanges interactifs…) ; il présente toutefois le défaut capital du manque de *représentativité*, vu la présence dominante de certains domaines ou thématiques au détriment d'autres (ex. les variétés régionales ou la langue parlée). En outre, l'impossibilité fréquente d'établir la source des informations, la présence de duplications d'informations, voire de plagiats, le nombre d'erreurs – même simplement sous forme de coquilles – en limitent fortement l'exploitabilité pour des recherches précises dans le domaine de la linguistique descriptive, et par conséquent de la terminologie.

Et pourtant, la présence de typologies textuelles extrêmement variées, l'abondance des documents ainsi que la possibilité d'analyser de nouvelles formes de communication font des corpus tirés de l'Internet une ressource précieuse pour les analystes des faits linguistiques. En particulier, l'essor et le développement des réseaux sociaux et des moyens de communication médiée par ordinateur (forums, listes de discussion, etc.) ont permis l'étude de typologies textuelles hybrides et multiformes, ainsi que l'accès à des formes de

1 <https://www.google.it/search?q=webcorp&ie=utf-8&oe=utf8&aq=t&rls=org.mozilla:it:official&client=firefox-a>

communication multimodales dont la traçabilité était autrefois impossible (voir entre autres Marcoccia 2004).

Dans le cadre de notre étude, compte tenu des considérations qui précèdent, l'Internet a été plutôt utilisé comme *ressource* pour le choix et la constitution du corpus spécialisé objet de l'analyse : nous avons donc opté pour la philosophie du *Web for corpus*.

4. Constitution et caractéristiques du corpus d'analyse

Pour l'analyse que nous avons décidé de mener dans les pages qui suivent, plusieurs paramètres ont présidé à la création du corpus, paramètres que nous résumerons brièvement ci-dessous :

1. *représentativité* dans le domaine : le but de l'analyse étant celui de mettre en évidence les stratégies et les phases de discussion autour des concepts et des termes de spécialité, nous avons opté pour un forum de spécialistes et d'amateurs dans le domaine du vin : http://lapassionduvin.com/phorum/ ;
2. *homogénéité* des textes produits, qui devaient être des énoncés spontanés, en situation de communication semi-surveillée, dans une situation de négociation discursive réelle ;
3. *authenticité* du corpus, qui devait être composé d'échanges spontanés et non forgés.

Le corpus se compose donc d'extraits du forum cité, dans trois *threads* différents qui partagent toutefois la thématique métalinguistique / métaterminologique, à savoir :

- *Qu'est-ce que le vieux vin ?*
- *Vocabulaire : acidité et fraîcheur*
- *De la complexité dans le vin...*

Les autres caractéristiques du corpus en revanche peuvent être résumées dans la matrice proposée auparavant :

Paramètres	
Canal (écrit / oral)	les énoncés du corpus relèvent des formes hybrides typiques de la communication médiée par ordinateur, ils présentent donc les traits typiques de l'oral sous forme écrite (présence d'émoticons, emploi expressif de la ponctuation et de la typographie)
Format (public / non public)	échanges publics
Cadre (institutionnel / personnel / informel)	cadre informel
Destinataires (statut, compétences)	les destinataires sont tous des spécialistes ou des amateurs – par rapport à d'autre forums, ici le critère d'identification à la communauté par le biais des compétences est évident
Connaissances requises et partagées	connaissances spécialisées dans le domaine de la dégustation du vin, y compris de la terminologie utilisée
Destinateur (statut, compétences)	spécialistes et amateurs
Fonction dominante du texte	la fonction dominante dans les échanges est la fonction informative – parfois, les échanges ont aussi une fonction persuasive ou d'identification à la communauté
Thématique du texte	les descripteurs, les concepts et le vocabulaire de la dégustation

Tableau 2 : Matrice appliquée au corpus d'analyse.

4. Des concepts et des termes : le langage du vin est-il vraiment une langue spécialisée ?

Le domaine choisi pour cette étude présente, par rapport à d'autres langues de spécialité, quelques problèmes de définition et de statut, sur lesquels nous nous concentrerons brièvement avant de passer à l'analyse du corpus présenté auparavant.

Le langage autour du vin est un langage complexe, varié, qui se trouve au carrefour de diverses disciplines et secteurs professionnels ; loin d'être strictement délimité, comme les disciplines faisant partie des sciences "dures" , il puise ses concepts et ses termes dans la botanique, dans la chimie, dans la physique, la météorologie, ce qui le rend hétérogène, mouvant et difficile à cerner (voir à ce propos entre autres Rossi 2009a et pour une histoire exhaustive les essais de Peynaud et Blouin 1983 et Dedourge 2000). Qui plus est, la terminologie de la dégustation souffre d'une ambiguïté de fond : les sensations organoleptiques sont essentiellement subjectives, difficiles à verbaliser, ineffables. Casamayor (cité par Rossi 2009) met en évidence la pauvreté lexicale des langues à ce propos, pauvreté qui rend nécessaire l'emprunt de termes par métaphore à la langue commune (on parlera alors de la *robe* d'un vin *costaud*, de ses tanins *veloutés*, de son *caractère* ou de sa *race*… pour un traitement approfondi du sujet, nous renvoyons aux études de Martine Coutier (1994, 1997) ou à d'autres domaines de la connaissance (l'architecture par exemple). Enfin, bien que la dégustation du vin soit un art ancien, dont le premier embryon remonte aux *haustores* romains (Johnson 2012), le vocabulaire de la dégustation est loin d'être un vocabulaire normalisé : pour des raisons historiques et économiques, l'harmonisation du vocabulaire de la dégustation est un objectif qui n'a jamais été atteint. Ce manque de normalisation engendre des incohérences au niveau de la segmentation conceptuelle et terminologique du domaine, ainsi que la présence de synonymies pléthoriques, de parasynonymies ambiguës, de polysémies qui ont contribué à une dévalorisation de la composante spécialisée dans la

communication autour du vin, bien souvent à tort considérée comme une forme de jonglerie verbale, de langage promotionnel.

Ce statut incertain complique les critères de sélection et d'analyse de corpus spécialisés dans le domaine œnologique, qui reste en tout cas un domaine fortement thématisé présupposant des compétences spécifiques de la part de ses usagers (voir aussi Rossi 2009b). En outre, ce manque de définition légitime est à l'origine d'une sensation commune d'incertitude de la part des locuteurs pour ce qui est de l'emploi d'une terminologie spécifique : les novices, mais aussi les amateurs, expriment souvent leur difficulté à sélectionner les termes appropriés, à trouver un accord universel sur le sens des termes et la délimitation des concepts, à s'exprimer par une terminologie correcte. Ce flou, ce manque de confiance, font finalement du langage de la dégustation un terrain privilégié pour l'observation que nous avons décidé d'effectuer, à savoir l'étude des dynamiques discursives à l'œuvre lors de la négociation d'une terminologie partagée dans un groupe de locuteurs spécialistes.

5. Le terme, fruit d'un consensus social : la définition comme forme de négociation conceptuelle

La terminologie à l'heure actuelle a considérablement changé sa conception de la création et de la stabilisation des termes dans un domaine spécialisé : loin d'être imposés par les autorités politiques ou linguistiques, les termes sont d'abord négociés par les acteurs du domaine concerné, par les spécialistes du secteur, à travers des processus dynamiques et interactifs. Cette conception du terme comme un produit du consensus social est bien mise en évidence dans les études en socioterminologie, à partir de l'ouvrage fondateur de Gaudin (2003 : 38) qui, dans le processus de formation des termes, qu'il appelle *sociogenèse*, reconnaît un rôle capital à la communauté des locuteurs, dans le sillage de Putnam : *"la signification est interactionnelle. L'environnement lui-même joue un rôle dans la*

détermination de ce que désignent les mots d'un locuteur ou les mots d'une communauté (Putnam, 1990 : 74)". Selon Gaudin, ce n'est qu'à travers la négociation au sein de la communauté des acteurs sociaux que le terme devient une entité stable, la circulation des savoirs et les pratiques langagières ayant une influence plus importante que les autorités politiques sur l'évolution des termes et des domaines.

L'influence des pratiques langagières sur la stabilisation des savoirs et des terminologies spécifiques amène également à la prise en compte des facteurs sociaux liés à la circulation des savoirs, de la segmentation des divers publics concernés, des discours variables autour des thématiques de domaine. Comme le souligne récemment Vicari (2012), la considération de ces dynamiques permet enfin de restructurer les paramètres observables, intégrant les acquis des disciplines liées à l'analyse des discours ordinaires, à la linguistique populaire, au *continuum* qui caractérise la communication et l'échange des savoirs entre locuteurs spécialistes et profanes. Les forums Internet représentent un lieu privilégié pour l'observation de ces dynamiques discursives et textuelles, en ce qu'ils permettent à l'observateur de suivre et de reconstituer le processus de discussion qui est à la base de l'accord et de l'acceptation d'un terme en tant que tel au sein d'une communauté de locuteurs.

La microanalyse des échanges révèle – ce qui nous intéresse plus spécialement dans ces pages – que c'est autour de la *définition* des concepts que le débat se développe : l'activité définitoire est ici interprétée dans son sens étymologique, à savoir comme "délimitation des frontières" conceptuelles dans le domaine concerné. Dans ce cas, l'attention de l'observateur se focalise sur les stratégies discursives liées à la formulation d'énoncés définitoires ordinaires (voir Martin 1990 ; à ce propos, voir aussi les travaux sur les *gloses spontanées*, Niklas-Salminen 2010 ou Julia 2001) qui visent à catalyser le consensus des spécialistes par rapport à une notion donnée, un nouveau concept, un cas d'ambiguïté ou une lacune dénominative face à une innovation technologique ou conceptuelle. La définition est alors perçue comme une activité majeure pour la structuration terminologique du domaine et pour que le consensus puisse s'établir au sein de la communauté prise en examen. Pour cette raison, nous

allons focaliser notre attention sur l'analyse des énoncés à visée définitoire dans les échanges du forum constituant notre corpus.

6. Analyse quantitative et qualitative du corpus : stratégies de définition à l'œuvre

Le corpus décrit au début de notre contribution, tiré du forum *lapassionduvin.com*, se compose de trois fils de discussion concernant des thématiques métaterminologiques : la définition du concept de *vin vieux*, un cas de synonymie présumée entre *acidité* et *fraîcheur* et la définition du concept de *complexité* dans les vins.

Au niveau de l'analyse quantitative, le corpus est composé de 266.545 caractères, correspondant à 44.417 mots. Le traitement d'analyse textuelle a été effectué à l'aide du logiciel concordancier *Antconc*[2], qui nous a permis d'effectuer d'abord quelques recherches au niveau purement quantitatif, concernant la présence de marqueurs métadiscursifs et métalinguistiques signalant la présence d'énoncés définitoires : des substantifs tels que *définition, mot, terme*, des verbes comme *définir, désigner, appeler, vouloir dire*, ainsi que des locutions comme *autrement dit, en d'autres termes* (voir également Vicari 2012).

Dans la grille suivante, les résultats quantitatifs obtenus par la sélection des marqueurs définitoires :

terme	**occurrences**	**% sur le nombre des mots du corpus**
définition	42	0,1%
définir	7	
appeler	2	0,02%
j'appelle	4	

2 <http://www.antlab.sci.waseda.ac.jp/software.html>

on appelle	3	
signifie	6	0,01%
désigner	2	0,004%
vouloir dire	12	0,02%
mot	22	0,04%
terme	34	0.07%
autrement dit	3	0,006%

Tableau 3 : Analyse quantitative.

Comme il ressort avec évidence de la grille ci-dessus, les marqueurs définitoires explicites ne constituent qu'un faible pourcentage d'occurrences au niveau purement quantitatif, ce qui semblerait indiquer que la négociation s'exprime à travers d'autres stratégies, du moins dans la plupart des cas – en d'autres termes, que les locuteurs ne verbalisent pas explicitement l'action de définition, même dans les cas où elle constitue le centre de la discussion et son objectif affiché dès le début.

Dans ce contexte, l'analyse qualitative fine et approfondie, effectuée manuellement, s'avère la seule ressource efficace : nous avons donc soumis le corpus à un dépouillement manuel, afin de faire ressortir les stratégies définitoires activées par les participants à l'échange.

La typologie des énoncés définitoires révèle des constantes fondamentales : en premier lieu, la nécessité de légitimer d'un point de vue épistémologique la définition formulée, par la caution d'une autorité reconnue dans le domaine – dans ce cas spécifique, on assiste à l'évocation d'autorités professionnelles (sommeliers, dégustateurs, œnologues renommés), mais également d'autorités linguistiques (dictionnaires, encyclopédies), ce qui témoigne du pouvoir des autorités de langue même dans la délimitation des concepts spécialisés :

Définitions d'autorité	Autorité évoquée	Exemples
	- autorité professionnelle (AP) - autorité	***- d'après le dictionnaire de la langue du vin de Martine Coutier*** *(c'est pas* ***l'INAO****, c'est* ***le CNRS****, mais tu peux*

	linguistique (AL)	*l'acheter quand même) ce sont certains arômes qui donnent l'impression de fraîcheur (que ce soit au nez ou en rétro), cette sensation peut être renforcée par l'acidité du vin.* (AL) *- Je lis la définition du mot "filigrane" dans* ***mon dictionnaire […]*** (AL) *- Voilà ce qu'en dit* ***Enrico Bernardo****. […] Une lecture attentive de son dernier et excellent* ***ouvrage "Savoir goûter le vin" – chez Plon,*** *permet de mieux cerner sa vision de la complexité. Je viens de le faire, ce week-end.* (AP) *-* ***Dans mon Gault Millau*** *2006 : complexe , se dit d'un vin dont les arômes sont variés et harmonieux.* (AP)

Tableau 4 : Enoncés définitoires d'autorité.

Lorsque la définition formulée n'évoque pas une autorité spécifique, dans les cas donc de formulation de définitions profanes ordinaires, les stratégies adoptées par les locuteurs diffèrent, suivant néanmoins des patrons formels constants, que nous résumerons brièvement comme suit :

- *X est… c'est* – dans ce cas, la définition paraphrastique (voir à ce propos Martin 1983, cité par Frassi 2010) vise à fournir une description du défini en langage ordinaire ; ce modèle passe le plus souvent par l'expression d'un synonyme / parasynoyme du défini ;
- *X + traits saillants* – ce modèle, qui reprend le modèle lexicographique dominant de la définition logique, est bien souvent exprimé par la simple mention des traits spécifiques,

explicités sous forme de liste ouverte, la mention du *genus* étant sous-entendue ;

- *X est comme + comparant* – ce modèle d'assimilation passe par la comparaison entre le défini et des fragments de l'expérience sensorielle des dégustateurs. Il s'agit sans doute de l'un des modèles dominants et les plus intéressants à l'intérieur du corpus : les termes de comparaison sont en effet variés et appartiennent aussi bien aux domaines techniques concernés qu'à des champs sémantiques apparemment très éloignés du domaine œnologique (par exemple, un vin complexe comparé à une éclipse totale) ;
- *X + glose métalinguistique* – ce modèle s'avère particulièrement intéressant dans le corpus analysé, vu les compétences supposées des participants, ainsi que leurs intérêts. En fait, les définitions métalinguistiques sont nombreuses et parfois particulièrement approfondies dans l'étude de l'étymologie des termes, de leur histoire (comme dans l'exemple ci-dessous) :

Définitions ordinaires profanes	Typologie - X est / c'est…	Exemples *- la complexité d'un vin, **c'est sa capacité** de jouer sur plusieurs registres à la fois.* *- la fraîcheur **c'est l'acidité** plus autre chose et l'acidité est sûrement indispensable à la fraicheur*
	- X + traits saillants	*- complexe veut dire **multiple, aux arômes variés, qui évoluent** et donnent des sensations renouvelées au fur et à mesure de la dégustation.*
	- X est comme + comparant	*- Un grand vin, donc complexe, **c'est comme une éclipse totale** et il*

		serait dommage de ne pas partager ce plaisir ! *- Le vieillissement* ***c'est comme l'affinage*** *de certains fromages.*
	- X + glose métalinguistique (*signifie, se dit de…*)	- *Précisément* cum plectus ***signifie*** *embrasser, entourer…(pour en faire une enceinte) et tant physiquement que par la pensée, et le cum (avec) renforce l'idée d'un tout.* ***Le latin nous enseigne*** *donc que la complexité est l'art de réunir ou d'embrasser divers éléments en l'occurrence disparates et de chercher à les réunir pour organiser un tout.*

Tableau 5 : Typologie des définitions ordinaires profanes.

Dans le corpus, les modèles explicites dominants s'avèrent être le modèle paraphrastique, bien souvent exprimé par synonyme (la synonymie étant d'ailleurs caractéristique de cette langue de spécialité par rapport à d'autre domaines plus proches des sciences physiques), et le modèle de la glose métalinguistique, qui témoigne de la conscience métalinguistique développée chez les locuteurs spécialistes et amateurs. Enfin, on remarque la présence fréquente d'énoncés qui pourraient être étiquetés comme des *refus de définir*, à savoir des énoncés dans lesquels les locuteurs refusent de donner une définition, à cause de la difficulté du concept et du manque d'accord partagé sur la segmentation conceptuelle du domaine ; ces énoncés mériteraient une analyse à part, afin d'expliciter les raisons qui rendent impossible l'activité – normalement universelle – de définir le sens d'un mot

(complexité du concept, évidence de la question…) ; voici quelques exemples :

> *Cela dépend de tellement de paramètres que cela me semble inutile de donner une définition!*
>
> *Devrais-je bientôt être capable de donner la définition de "bon" avant de dire ici que je trouve "bon" le pommard 1949 que j'ai bu l'été dernier?*

7. Négociation d'une terminologie commune : micro-analyse d'un segment de discussion

Les stratégies définitoires que nous venons d'esquisser contribuent *dans des modalités diverses et partielles à la négociation d'une* terminologie partagée au sein de la communauté des internautes-amateurs : pour terminer notre étude, la micro-analyse d'un segment de forum nous permettra de mettre en évidence les dynamiques discursives au niveau de l'agencement des arguments.

Le segment sélectionné dans le corpus est relatif à une discussion sur les deux termes d'*acidité* et de *fraîcheur* dans la dégustation des vins (nous avons choisi de reproduire les messages sans intervenir pour corriger d'éventuelles coquilles ou fautes d'expression).

Le message d'ouverture du fil de discussion présente explicitement la problématique métalinguistique, ainsi que la difficulté, voire l'impossibilité, de faire référence à une terminologie normalisée :

gwnrouge
Vocabulaire : acidité et fraîcheur
mardi 22 mai 2012 19:31:55 IP: Adresse IP enregistrée
Membre depuis : 1 an
Messages: 152 Bonsoir à tous,
En supposant une quelconque convention dans le vocabulaire de la dégustation (ce qui ne semble en fait pas exister), j'ai juste une bête question :

est-ce que pour vous, les mots "acidité" et "fraîcheur" sont-ils synonymes ? Autrement dit, est-il équivalent de dire "ce vin présente une belle acidité" et "ce vin présente une belle fraîcheur" ?
Amicalement,
Gwenaël

Gwnrouge pose une question de différenciation dans le domaine des descripteurs de l'analyse organoleptique, à savoir la distinction entre deux termes souvent utilisés comme synonymes : *frais* et *acide*.

Les participants au forum réagissent immédiatement à la question :

enzo d'aviolo
Re: Vocabulaire : acidité et fraîcheur
mardi 22 mai 2012 20:20:01 IP: Adresse IP enregistrée
Membre depuis : 5 ans
Messages: 14 161 **non c'est différent**. un vin peut être frais en perception et ne pas avoir d'acidité perceptible, surtout sur les rouges.
Si l'acidité peut contribuer à la fraîcheur, ce n'est pas le seul arôme ou la seule saveur qui participe à la fraîcheur.

Luc Javaux
Re: Vocabulaire : acidité et fraîcheur
mardi 22 mai 2012 20:21:33 Administrateur
IP: Adresse IP enregistrée
Membre depuis : 8 ans
Messages: 18 080 Quoi d'autre alors Laurent ? Mis à part la température de service bien entendu.
Luc

enzo d'aviolo
Re: Vocabulaire : acidité et fraîcheur
mardi 22 mai 2012 20:23:00 IP: Adresse IP enregistrée
Membre depuis : 5 ans
Messages: 14 161 **à mon sens, une fine amertume, des notes mentholées, des notes d'agrumes, peuvent contribuer à la perception de fraîcheur par exemple**.

La question est alors de savoir si l'extension des deux concepts est identique, ou bien si le concept de *fraîcheur* est plus complexe que celui d'*acidité*, qui ne serait que l'un des composants organoleptiques de la *fraîcheur*. Nous signalons que – bien souvent – l'expression de

définitions ou précisions est accompagnée de marqueurs métadiscursifs visant à légitimer par restriction la portée du jugement exprimé (*à mon sens, selon moi…*).

La précision apportée par Laurent (*enzo d'aviolo*) est acceptée par Gwnrouge :

> gwnrouge
> Re: Vocabulaire : acidité et fraîcheur
> mardi 22 mai 2012 22:55:03 IP: Adresse IP enregistrée
> Membre depuis : 1 an
> Messages: 152 Merci Laurent de ta réponse. Je saisi bien le type de sensation. On est d'accord que l'on ne parle pas de température... mais **c'est toujours un peu délicat le vocabulaire de dégustation**... On pourrai dire qu'un vin acide est souvent frais (l'acidité donne souvent une sensation de fraicheur) et que **c'est souvent l'acidité qui donne l'impression de fraicheur, mais pas toujours**. Tout ça pour dire que je m'aperçois que dans mes CR, j'ai considéré certainement à tort que "fraicheur" et "acidité" sont interchangeables...
> **Serait utile que l'ISO ou l'INAO "normalise" le langage du vin...**
> Amicalement,
> Gwenaël

La nécessité de normalisation émerge avec évidence de ces messages d'amateurs, qui demandent plus d'univocité dans la définition des termes ; le besoin d'une norme de référence est d'ailleurs l'un des fils rouges de ces discussions, comme il ressort du message qui suit, toujours en réponse à la question de Gwnrouge :

> Jean-Bernard
> Re: Vocabulaire : acidité et fraîcheur
> mardi 22 mai 2012 23:12:05 IP: Adresse IP enregistrée
> Membre depuis : 7 ans
> Messages: 3 415 Hello Gwenaël,
> **d'après le dictionnaire de la langue du vin de Martine Coutier (c'est pas l'INAO, c'est le CNRS, mais tu peux l'acheter quand même)** ce sont certains arômes qui donnent l'impression de fraîcheur (que ce soit au nez ou en rétro), cette sensation peut être renforcée par l'acidité du vin. **Bref tout pareil que Laurent! C'était juste pour dire qu'il existe au moins un ouvrage de référence**. Le plus de ce dico c'est qu'il y a plusieurs citations de CR en exemple qui illustrent chaque définition.

Toutefois, la légitimation de l'autorité n'est pas évidente ; le message du locuteur suivant nous révèle que le débat porte également sur la compétence et l'autorité des sources citées (à ce propos, et plus en général sur la légitimation des autorités, nous renvoyons aux analyses d'Origgi sur la notion de *confiance épistémique*, citée par Vicari 2012) :

> chrisdu74
> Re: Vocabulaire : acidité et fraîcheur
> mardi 22 mai 2012 23:48:47 IP: Adresse IP enregistrée
> Membre depuis : 4 ans
> Messages: 1 225 Je suis d'accord avec Enzo et Oliv, la fraîcheur c'est l'acidité plus autre chose et l'acidité est sûrement indispensable à la fraicheur.
> En termes de commentaires de dégustation, je dirais que tant que c'est légèrement acide, ça contribue à la fraicheur et on ne commente pas forcément le caractère acide; quand c'est vraiment acide, j'emploie le terme d'acidité (sans que soit péjoratif) et plus de fraîcheur .
> Sinon, j'intervenais surtout pour répondre à JB : **Le dico en question, je l'ai à la maison et depuis que j'ai lu à "oxydatif" la définition brute "qui a un caractère oxydé", je ne l'ouvre plus souvent**....
> Chris d'U

La discussion sur les termes en réalité est également une négociation sur le statut légitime des locuteurs participant à l'échange ; leur identification et reconnaissance par la communauté décident de la nature de leurs remarques, ainsi que de l'accord sur les arguments cités :

> Mané Galinha
> Re: Vocabulaire: acidité et fraîcheur
> mercredi 23 mai 2012 09:03:47 IP: Adresse IP enregistrée
> Membre depuis : 2 ans
> Messages: 979 Laurent, il n'y a pas un lien direct entre l'acidité et le pH d'un vin. Tu en parles comme deux choses distinctes. Il me semblait pourtant que quand le pH était bas, les acidités étaient hautes ? Non ?
> Manuel.
>
> enzo d'aviolo
> Vocabulaire: acidité et fraîcheur

mercredi 23 mai 2012 09:35:23 IP: Adresse IP enregistrée
Membre depuis : 5 ans
Messages: 14 161 **je ne suis pas chimiste** mais ce que je sais c'est que ce n'est pas linéaire et que c'est d'ailleurs parfois un problème (même si en effet c'est censé se suivre).
Bouland nous en parlait récemment en disant que la garantie qualitative d'un vin c'est que le pH soit suffisamment bas en parallèle de l'acidité pour apporter de la fraicheur en bouche et une garantie de bon vieillissement, l'inverse générant une acidité factice faisant que les vins risquaient ensuite de s'écrouler au vieillissement (j'espère ne pas trop mal reproduire son propos) Rachid, tu te souviens?

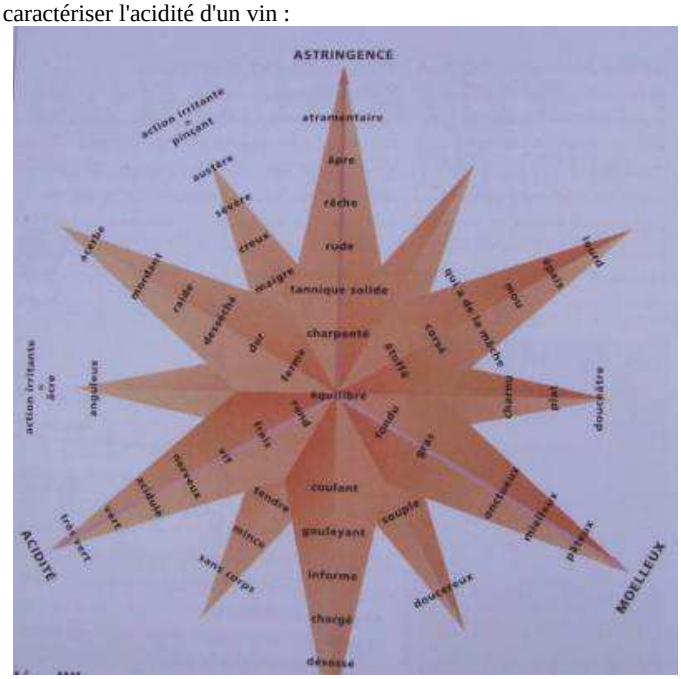
Martinez
Re: Vocabulaire : acidité et fraîcheur
mercredi 23 mai 2012 13:25:43 IP: Adresse IP enregistrée
Membre depuis : 8 ans
Messages: 3 928 Pour moi, ce sont deux mots qui n'arrivent pas au même moment dans un CR.
Je parle de l'acidité quand j'analyse le vin en le décortiquant, comme si j'essayais d'isoler ce paramètre et je parle de fraîcheur au moment de la synthèse sur le vin dans sa globalité.
L'incidence de l'acidité sur la sensation globale de fraîcheur me semble assez forte.
Jmm

Eric B
Re: Vocabulaire: acidité et fraîcheur
mercredi 23 mai 2012 13:49:21 IP: Adresse IP enregistrée
Membre depuis : 6 ans
Messages: 10 660 **De ce que je me souviens de mes cours d'oenologie**, le pH ne mesure que l'activité des ions Hydrogènes. Or, il y a une partie de l'acidité du vin qui ne dépend pas de ceux-ci. D'où l'importance de chiffrer aussi l'acidité totale.
Eric

Finalement, l'avis définitif est apporté par l'administrateur du site, par le biais de l'argument d'autorité :

Luc Javaux
Re: Vocabulaire : acidité et fraîcheur
mercredi 23 mai 2012 13:51:01 Administrateur
IP: Adresse IP enregistrée
Membre depuis : 8 ans

Messages: 18 080 Petit rappel pour confirmer que le terme "frais" est bien utilisé pour caractériser l'acidité d'un vin :



Luc

Mané Galinha
Re: Vocabulaire: acidité et fraîcheur
mercredi 23 mai 2012 13:51:55 IP: Adresse IP enregistrée
Membre depuis : 2 ans
Messages: 979 Voilà, encore une fois, **avec une explication simple mais précise, la science vient clore ce débat**.
Merci Eric.
Manuel.

Ces fragments de forum nous révèlent que les dynamiques discursives présidant à la définition partagée de concepts spécialisés sont extrêmement complexes et qu'elles dépendent de critères variés. Ce qui entre en jeu, ce n'est pas seulement la définition des concepts et des termes qui les désignent, mais également – voire surtout – la définition du statut plus ou moins légitime des locuteurs, leur appartenance à la communauté des experts. Faute d'autorités externes

et supérieures (l'INAO, l'ISO…), la définition des termes va de pair avec l'identification d'une autorité épistémique interne au groupe, qui pourra finalement apporter le dernier mot sur l'ambigüité terminologique évoquée.

8. En guise de conclusion

Les considérations présentées dans les pages précédentes nous permettent finalement de formuler quelques conclusions partielles sur les hypothèses de départ de notre analyse.

La constitution de corpus représentatifs offre désormais la possibilité d'étudier de façon plus approfondie les dynamiques discursives et textuelles qui président, dans une communauté de spécialistes, à la diffusion des termes et à l'accord sur leur délimitation sémantique ; la possibilité d'exploiter des corpus authentiques comme les forums de discussion est précieuse pour les analystes, en ce qu'elle permet de retracer de façon précise et ponctuelle les mécanismes de négociation des termes et des concepts. En même temps, la microanalyse du corpus sélectionné nous a permis de mettre en évidence les limites d'une approche purement quantitative, fondée sur les statistiques d'occurrence : la définition des termes passe par des mécanismes divers, implicites, que l'on ne peut repérer qu'à travers une analyse qualitative détaillée.

Dans notre cas spécifique, l'analyse a révélé une activité métalinguistique intense de la part des œnologues amateurs : intéressés par les questions d'usage correct de la terminologie de référence, ils s'interrogent sur la délimitation des concepts, la synonymie des termes, la définition des désignations ; leurs stratégies définitoires s'appuient souvent sur des marqueurs métalinguistiques, ainsi que sur l'avis légitime des autorités de langue – outre que des autorités professionnelles reconnues. Leurs échanges témoignent d'un besoin de normalisation au niveau de la terminologie œnologique, mais également d'une conscience métalinguistique présente et

constamment remise en question, d'un désir de *parler* du vin, de trouver les *mots justes* pour exprimer leurs sensations. S'agit-il d'un phénomène limité au domaine de la dégustation, dû à une lacune de normalisation officielle, ou bien ces tendances caractérisent-elles toute communauté de professionnels ? D'après les études conduites dans d'autres domaines (Vicari 2012), nous croyons que la nécessité de définir la terminologie sur la base d'un accord des spécialistes représente l'une des tendances constantes du discours spécialisé, indépendamment du domaine de spécialisation pris en considération : d'après ces analyses, le terme s'avère à tous les effets un *produit social*.

Références bibliographiques

Amoraritei, Loredana 2003. *La Métaphore en Œnologie*. <http://www.metaphorik.de/03/amoraritei.htm> (consulté le 25/02/2013).

Beciri, Hélène 2010. La définition en contexte dans la communication technique asymétrique : stratégies et enjeux. *Publif@rum* » 11 <http://publifarum.farum.it/ezine _articles.php?id=127> (consulté le 25 février 2013).

Biber, Douglas 1995. Representativeness in corpus design. *Linguistica Computazionale* IX-X, 377- 407.

Bourigault, Didier / Slodzian, Monique 1999. Pour une terminologie textuelle. *Terminologies nouvelles*, 19, 29–32.

Cabré, Maria Teresa 1999. *La terminologia: representación y comunicación. Elementos para une teoría de base comunicativa y otros artículos*. Barcelona : Institut Universitari de Lingüística Aplicada.

Cabré, Maria Teresa 2008. Comment constituer un corpus de textes de spécialité. *Cahier du CIEL* 2007-2008, pp. 37-56.

Condamines, Anne 2005. Linguistique de corpus et terminologie. *Langages* 157, *La terminologie : nature et enjeux*, 36-47.

Condamines, Anne / Aussenac-Gilles, Nathalie 2006. *Corpus et terminologie*, AS-34 ASSTICOT - RTP DOC STIC CNRS - Rapport d'activités, <http://hal.inria.fr/sic_00001169>

Coutier, Martine 1994. Tropes et termes : le vocabulaire de la dégustation du vin. *META* 39 / 4, 662-675.

Coutier, Martine 1997. Le champ du corps dans le vocabulaire de la dégustation du vin. *Cahiers de lexicologie* 71, 1997-II, 67-99.

Dedourge, Raphaëlle 2000. *Introduction à la Dégustation*. Iacchos, en ligne : <http://www.chateauloisel.com/vin/livre/dedourge-introduction-degustation.htm>

Diki-Kidiri, Marcel 2000. Terminologie et diversité culturelle. *Les Cahiers du Rifal* 21, en ligne : <http://www.rifal.org/cahiers/rint21/rint21.pdf> (consulté le 25 février 2013).

Dubuc, Robert 1992. *Manuel pratique de terminologie*. Montréal : Linguatech.

Enjalbert, Henri 1975. *Histoire de la vigne et du vin, l'avènement de la qualité*. Paris : Bordas.

Fairon, Cédric 2011. Constitution et étude de corpus spécialisés sur le web. *Le discours et la langue* 2.1, *L'internet, corpus sauvage : nouvelles ressources, nouveaux problèmes ?*, 73-90.

Fouqueré, Christope / Issac, Fabrice 2003. Corpus issus du Web : constitution et analyse informationnelle, *Revue québécoise de linguistique* 32 / 1, 111-134. <http://id.erudit.org/iderudit/012246ar>

Frassi, Paolo 2010. La définition lexicographique : aperçu et propositions. *Autour de la définition*. *Publifarum* 11, <http://publifarum.farum.it/ezine_articles.php?id=118> (consulté le 30 mai 2013).

Gaudin, François 2003. *Socioterminologie, une approche sociolinguistique de la terminologie*. Bruxelles : De Boeck.

Giaufret Anna / Rossi, Micaela 2010. L'année mondiale de l'astronomie : la diversité terminologique et culturelle de l'espace métaphorique, *Actes du Colloque GLAT 2010*, Lisboa, 17-19 maggio 2010, 131-144.

Giaufret, Anna / Rossi, Micaela 2013. Métaphores terminologiques, circulation des savoirs et contact entre langues. *Signes*,

Discours et société 10, *La métaphore dans les discours spécialisés*, février 2013 <http://www.revue-signes.info/sommaire.php?id=2921>.

Gilardoni, Silvia 2007. Descrivere il vino : analisi semantico-lessicale di una terminologia specifica. *L'Analisi linguistica e letteraria* 15, 25-46.

Habert, Benoît / Nazarenko, Adeline / Salem, André 1997. *Les linguistiques de corpus*. Paris : Colin.

Jacobi, Daniel 1999. *La communication scientifique : discours, figures, modèles*. Saint-Martin d'Hyères : Presses universitaires de Grenoble.

Johnson, Hugh 2012. *Il vino. Storia, tradizioni, cultura*, Feltrinelli : Milano.

Julia, Catherine 2001. *Fixer le sens. La sémantique spontanée des gloses de spécification du sens*. Presses de la Sorbonne Nouvelle : Paris.

Marcoccia, Michel 2004. L'analyse conversationnelle des forums de discussion : questionnements méthodologiques, *Les Carnets du Cediscor* [En ligne] 8 : <http://cediscor.revues.org/220>

Martin Robert 1990. La définition 'naturelle'. In Chaurand, J., Mazière, F. (dir.). *La définition*. Larousse : Paris, 86-95.

Martin, Jean-Claude 2006. *Parcours culturel dans l'identité du vin*. Agro – Montpellier version 10 mai 2006 <http://le-vin-nature.com/blog/index.php?id=84>

Niklas-Salminen, Aïno 2010. La définition dans le cadre de la glose spontanée. *Autour de la définition*. *Publifarum* 11 http://publifarum.farum.it/ezine_articles.php?id=125 (consulté le 30 mai 2013).

Peynaud, Emile / Blouin, Jacques 1983. *Le Goût du vin*. Dunod : Paris.

Rossi, Micaela 2009a. L'emploi de la métaphore comme ressource pour la néologie terminologique : le cas du langage de la dégustation du vin, Actes du Colloque *La métaphore en langues de spécialité*, Lyon, 26 septembre 2008, PUL, 199-227.

Rossi, Micaela 2009b. Langue et culture dans un verre. Pour une étude multilingue du langage du vin. In Lavric, Eva (ed.) : *Food and Language. Sprache und Essen* (Inntrans. Innsbrucker Beiträge

zu Sprache, Kultur und Translation), Frankfurt / M. : Peter Lang, 161-170.

Sierra Soriano, Ascension / Mogorron Huerta, Pedro *Quels termes pour parler de vin ? Etude contrastive français-espagnol*, <http://dialnet.unirioja.es/servlet/autentificador?redirect=/servlet/fichero_articulo?articulo=1011620>

Temmerman, Rita 2000a. Une théorie réaliste de la terminologie : le sociocognitivisme. *Terminologies Nouvelles* 21, juin 2000, 58-64.

Temmermann, Rita 2000b. *Towards new ways of terminology description : the sociocognitive-approach*. Amsterdam-Philadelphia : John Benjamins Publishing Company.

Van Campenhoudt, Marc / Temmerman, Rita 2011. Les corpus et la recherche en terminologie et en traductologie / Corpora and Research in Terminology and Translation Studies. *Meta* 56 / 2, juin 2011, 223-464.

Vicari, Stefano 2012. Del *bon usage* della terminologia delle energie rinnovabili nei forum Internet : analisi delle tipologie definitorie ordinarie. Jn Giaufret, Anna, Rossi, Micaela (eds.), *La terminologia delle energie rinnovabili tra testi e repertori: variazione, standardizzazione, armonizzazione*, Genova: GUP, 153-193.

Williams, Geoffrey. *Le corpus : constitution et contraintes* <http://geoffrey.williams.pagesperso-orange.fr/thesis/chap1.htm> (consulté le 30 mai 2013).

ERIK CASTELLO

14. Exploring Existential and Locative Constructions in a Learner and in an Expert Corpus of Promotional Tourist[1] Texts

1. Introduction

This chapter sets out to explore the use of existential and locative constructions in a learner corpus of promotional tourist texts written by university students of English and compares the findings to those obtained from the analysis of a corpus of texts written by tourism industry professionals. Existential and locative constructions are used in all registers, yet they are particularly frequent in fiction where descriptions of settings are very common (Biber *et al.* 1999: 943-956). As most tourist texts include sections in which such referents as tourist resorts, attractions and facilities are located in space and then described, it is only natural for them to contain parts expressing existence and location. In a way, this chapter pursues the research agenda put forward by Dann (1996: 2-4), for instance, who claims that "[no one] has so far systematically examined tourism as a language *per se*", while it is so extensive and pervasive that it "merits thorough investigation by tourism researchers", and also by Lam (2007: 72), who observes that "books related to tourism industry text writing […] fall considerably short of detailed discussions about the use of lexis and grammar in tourism industry texts".

1 For the purpose of this study the label "tourist" texts refers to texts produced by either tourism professionals or trainees.

As regards the methodology adopted in this study, it is believed that from the contrastive analysis of the promotional tourist materials produced by expert and novice writers respectively, specific aspects of this type of language can stand out and others can emerge which might otherwise go unnoticed. Also, language learners are likely to benefit from the comparison of their own texts to those written by expert (native) writers, in that in so doing they can not only hone their language skills (e.g. Seidlhofer 2002), but also develop the specific lexico-grammatical knowledge and textual skills needed to write effective tourist texts.

2. Research on the language of tourism

The language of tourism has been studied from various perspectives, including the literary, cultural, sociolinguistic and linguistic ones. The most specific and recurrent lexical, morphosyntactic and textual features of tourist publications have been investigated, as well as their potential for the development of linguistic and translation skills.

Dann (1996: 2-4) claims that the language of (promotional) tourism is a "discourse of its own", which "attempts to persuade, lure, woo and seduce" potential tourists, that is, to control their behaviour (Dann 1996: 246) by using a language which furnishes descriptions based on the tourist's expectations rather than on reality (Dann 1996: 85). He also affirms that "for optimal effectiveness the message [conveyed by tourist texts] should be short, clear, current, active and conclude with the identified key words" (Dann 1996: 174).

Gotti (2006) aptly distinguishes tourist communication between experts and tourist communication between specialists and non-specialists, the former representing an instance of highly specialised discourse and the latter tending towards general discourse. He then goes on to describe some lexical, syntactic and textual features which the language of tourism shares with other types of specialised discourse. The lexical features are the monoreferentiality of the terms used in such texts, the conciseness of their lexemes (e.g. *travelodge*)

and the emphasis expressed by, for instance, the use of evaluative adjectives and superlative forms. Their main syntactic features are expressive conciseness, that is, the use of extremely compact syntactic structures (e.g. *self-catering accommodation*), and pre-modification. Finally, the discourse of tourism exhibits distinctive textual characteristics which are due to the "norms governing the construction of its different text genres" (Gotti 2006: 31), the most typical of which are tourist guides, articles in specialised journals and general magazines, brochures and other advertising materials, itineraries, and professional correspondence. To these traditional and 'monologic' text types, Dann (2012) adds such recent 'dialogic' text types as blogs, virtual tourist communities and 'word-of-web overheards', and does so in an attempt to 'remodel' the language of tourism with a view to accounting for the recent changes brought about by the Internet.

As regards the morphosyntactic features of specialised discourse in general, Gotti (2011: 49) points out that:

> the specificity of morphosyntactic phenomena found in specialized languages is not a qualitative but a quantitative one. Certain features may also occur in general language but their frequency in specialized discourse makes them typical only of the latter.

Moya Guijarro and Albentosa Hernándes (2001: 365) investigate the textual phenomenon of thematic sequence (Gotti 2011: 84-87) in tourist brochures, and point out that the persuasive function of these texts:

> frequently leads the writer to move new topical entities away from the thematic slot of the sentence so that the introduction of the topic is postponed and the reader's attention is fully attracted by the area being advertised.

That is, the use of marked themes, a textual phenomenon which is common in general language as well, for 'strategic' reasons occurs more frequently in the specific language of tourism than in other text types.

The language of tourism shares some lexical and terminological components with the language produced in other professional domains, such as geography, economics, sociology and history, and

performs various rhetorical functions, including the persuasive, argumentative and informative ones (Calvi 2000). At the same time, however, it differs from these other domain-specific languages, given the tourism industry's peculiar communicative needs and the linguistic strategies that tourism professionals adopt to produce successful texts. According to Agorni (2012: 11-12), these strategies can be divided into: strategies meant to identify and address the reader, the appropriate use of specific genres, and the use of cross-cultural references. In a similar vein, Manca (2007: 115) claims that it is the advertising function in combination with the various components of the language of tourism which contributes to the development of standard lexical, morphosyntactic and textual features.

Some corpus-based studies have convincingly demonstrated that the promotional aims of the discourse of tourism account for some of its specific lexico-grammatical features. On the basis of his data, Lam (2007: 87), for instance, observes that:

> the lexical words and clusters used in tourism industry texts are very different from those found in general English in terms of their frequencies and the semantic fields to which they belong. And even if there are commonalities in the two different sources of texts, they still exhibit distinctive semantic prosodies.

As a way of exemplifying this statement, Lam (2007) looks into the collocational and colligational patterns of the words *city* and *islands* and their semantic prosodies in a 2-million-word corpus of online tourism industry texts. He also discusses the colligational patterns derived from the keyword clusters *one of the most* and *one of the more*. Other corpus-informed studies have uncovered further meaningful collocational and phraseological patters of frequently-used words in the field of tourism. Gerbig and Shek (2007), for instance, discuss the diachronic use of the keywords *tourist/s*, *travel* and *package holiday* in various corpora, including the Lancaster-Oslo-Bergen (LOB), which was compiled in the 1960s, and the British National Corpus, which dates back to the 1990s. Tognini Bonelli and Manca (2002) investigate the use of *welcome* and of its Italian functional equivalents in a comparable corpus of English and Italian

texts promoting farmhouse holidays, while Manca (2007) carries out a contrastive analysis of English and Italian nouns and adjectives belonging to the semantic fields of *beauty* and *tranquillity*.

Most of these studies also examine the implications that their findings can have for teaching and learning how to write tourism industry texts. Further research has been carried out on the features of the special language(s) of tourism and on how to best teach novice writers and trainee translators to respectively write and translate such texts (e.g. Kelly 2000; Palusci/Francesconi 2006; De Stasio/Palusci 2007; Curado Fuentes 2007; Fodde/Denti 2012). To the best of my knowledge, however, no study has so far investigated in detail either the syntactic and textual phenomena of existential and locative constructions in tourism industry texts or the problems that they might cause to novice non-native writers. The following sections will look into and attempt to shed light on these aspects.

3. Existential constructions in English

The lexico-grammatical and textual features explored in this chapter can be ascribed to the phenomena which are generally known in linguistics as existential and locative constructions. Because of their peculiarities, these constructions have attracted the attention of a large number of researchers working within various linguistic traditions and have been studied extensively (e.g. Jespersen 1924; Clark 1978; Lakoff 1987; Givón 1993; Biber *et al.* 1999; Huddleston/Pullum 2002; Martínez Insua/Palacios Martínez 2003; Palacios Martínez/Martínez Insua 2006).

In English, existential constructions take the form of expletive and unstressed *there + be + a noun phrase* (also called the Existent) and are the core of existential clauses. Besides the typical verb *be*, other verbs can be used after *there*, i.e. certain intransitive verbs expressing positional states (e.g. *stand, lie, stretch, remain, exist*) and

a few intransitive dynamic verbs of 'occurring', 'coming into view' or 'arrival on the scene' (e.g. *occur*, *follow*, *emerge*, *appear*, *loom*) (Downing/Locke 2006: 153). Existential *there* is relatively common in all types of registers: it is frequently used in casual conversation and fiction, and, to a slightly lesser extent, also in academic prose and in news (Biber *et al.* 1999: 948).

From a discourse and textual perspective, existential *there* mainly functions as a 'presentative' device in discourse, pointing as it does to the new information conveyed by the upcoming noun phrase. However, a distinction needs to be drawn between 'bare' existential clauses and 'extended' existential clauses. The former are exactly about the existence of an entity and are accompanied by "adjuncts of no syntactic significance for the existential construction" (Huddleston/Pullum 2002: 1393), while the latter contain complements or adjuncts that are less constrained or are followed by clauses that expand the existent. As a consequence, the former type cannot be reworded by means of a corresponding plain non-existential clause, whereas the latter can (Downing/Locke 2006: 257-258), as exemplified by (1), (2) and (3):

(1)	There's no milk (again).	*(Again) is no milk.
(2)	There's one copy *on the table*.	One copy is *on the table*.
(3)	There was a dog *barking outside*.	A dog is *barking outside*.

As can be noticed, the function of the 'extended' existential clauses in (2) and (3) is not primarily that of introducing a new referent into the discourse, but rather to expand on it and to establish coherence with the previous discourse (Downing/Locke 2006: 258-259). Specifically, the main information in (3) resides in the post-modifier of the noun phrase, i.e. *barking outside*. Huddleston and Pullum (2002: 1396-1397) also discuss some pragmatic constraints on the use of referents in the noun phrase. According to them, an existential clause can have an equivalent plain clause if a definite and physical entity is at issue (4). This also holds true when the abstract and indefinite referent in a noun phrase is already clear from the previous discourse (5):

(4) There was *a furniture van* in the drive. *A furniture van* was in the drive.

(5) There's *one performance* at noon. *One performance* is at noon.

If an adjunct of place or time is used in thematic position and is thus emphasised and foregrounded, existential *there* may be omitted, but is easily retrievable if a tag question is appended to the clause (Halliday 2004: 258). Lakoff (1987: 542) discusses this latter phenomenon by making reference to the cognitive processes underlying existentials and locatives and claims that:

> locatives are concerned with presence – presence in space, in our perception, in discourse, or in our imagination. The less vividly on stage an action is […] the more necessary existential *there* becomes. Abstractions such as absence, probability, and generic activities are not 'vividly on stage' and therefore require existential *there*. […] Existential *there* is necessary when there is no concrete object 'on stage'.

That is, existential *there* can be omitted in clauses such as (6), while it cannot be omitted in (7):

(6) On the wall (*there*) was a Picasso painting (Halliday 2004: 258).

(7) In the house *there* was *no* sign of life (Lakoff 1987: 542).

This phenomenon is also known as 'locative inversion', which is more common than existential *there* after an opening place adverbial. It mainly occurs in fiction when the relationships between the entity expressed in the adjunct and that in the locative clause are described in physical terms. In these cases, however, verbs other than *be* tend to be used, such as *hang, lay, run, sit,* and *stoop* (Biber *et al.* 1999: 954-955).

Another issue which concerns existential *there* constructions is the absence of number agreement which at times occurs between the noun phrase and the existential verb. As suggested by Lakoff (1987: 547-549), when the contracted form *there's* is used, some native English speakers tend to interpret existential *there* as the grammatical

subject, and do so irrespective of the number of the head noun, thus ending up using the singular form of the verb. In this respect, Martinez Insua and Palacios Martinez's (2003) investigation of the British National Corpus has revealed that non-concord occurs mainly in British spoken English, even though it is not as frequent as in other varieties of English (e.g. Canadian English). Also, its occurrence is highly correlated with the "presence of lengthy post-verbal sequences or of some adverbial or parenthetical element intervening between the verb and the notional subject" (Martinez Insua/Palacios Martinez 2003: 279), as exemplified by in the following example:

(8) And, *there's at least* eight *discs* available at the present time, and more could be made available. (Martinez Insua/Palacios Martinez 2003: 280)

Finally, Biber *et al.* (1999: 186) observe that in conversation *there's* "tends to behave as a single invariable unit for the purpose of speech processing" and that "the connection is far less close with *there was*, which is not reduced to a single syllable in speech and is not contracted in writing."

4. Methodology and data

The corpus-based investigation described in this chapter is conducted with reference to Hoey's (2005) framework of 'lexical priming', whereby words and structures are viewed as being primed for use in particular ways in specific contexts. To be precise, the focus is on existential and location patterns, and the research hypothesis is that they are primed for specific uses in tourist texts. The patterns investigated can be defined as 'colligational', in that the relationship between their components is 'colligation', a term which Hoey (2005: 43) defines as follows:

> [colligation is] the relation holding between a word and a grammatical pattern [and] going beyond traditional grammar relations and embracing such

> phenomena as the positioning of a word or word sequence within the sentence or paragraph and even its positioning within the text as a whole.

The method of analysis adopted is Contrastive Interlanguage Analysis (CIA), which aims to explore "non-native features of learner writing and speech through detailed comparisons of linguistic features in native and non-native corpora" (Granger 2002: 22). It involves the comparison of a corpus of tourist texts written by native or expert writers (NS) to a corpus of texts written by learners of English (NNS). Not only is the investigation conducted with the purpose of highlighting errors in the learner texts, but also (and mainly) to find "instances of under- and overrepresentation of words, phrases and structures" (Granger 2002: 22). The corpus analysis has involved the study of word lists and concordance output retrieved by means of the software package *Wordsmith Tools 5* (Scott 2008). The aim is to gain insights into how a given set of words establishes colligational relations with other words or grammatical patterns, thus contributing to the formation of existential and location patterns in the learner and in the expert corpus respectively.

The learner corpus compiled for this study is an electronic collection of authentic foreign language textual data (Granger 2002: 7) which was assembled in the academic year 2011/2012, with the aim of representing the promotional tourist language produced by a group of second-year university students of *Mediazione Linguistica e Culturale* (Linguistic and Cultural Mediation) at the University of Padua. It consists in 105 tourist texts (39,244 words) promoting the city of Padua and written by the students as part of a task which they completed under assessment conditions with the aid of dictionaries. The texts were written as responses to a prompt written in Italian which specified the three parts of the target text (i.e. introduction, what to see, how to get there) and listed the main points to be developed. The majority of the students had an Italian mother-tongue background or were highly proficient in Italian, and had all been studying at least one other European language. They had passed the first-year English exam, set at B1 level of the Common European Framework of Reference for Languages (Council of Europe 2001) in both receptive and productive skills, and were attending English

classes at B2 level. During the course they had been exposed to tourist texts in English, which were used as source texts for analysis and translation.

The expert tourist corpus is made up of 72 promotional tourist texts (144,291 words) collected between 1998-2008. Most of them are promotional in nature, i.e. 12 magazines, 12 brochures and 25 web pages, while the 28 extracts from tourist guides are mainly informative. All of them are texts concerning tourist attractions and resorts in the South of England and spa hotels in Bath and in the Lake District. In spite of the fact that all the texts were produced by tourism professionals with the aim of promoting those specific areas, there are some statistically significant linguistic and textual differences between the four related genres represented in the corpus, which Castello (2002) explores with reference to a sub-corpus of the expert corpus (92,500 words). While the brochures tend to be lexically dense, he found that the magazines and the tourist guides are grammatically more intricate (e.g. Halliday 2004). Also, the brochures contain the largest number of imperative and minor clauses, while the magazines are characterised by the use of a larger number of interrogative clauses and of modal verbs. Table 1 displays some figures concerning the two corpora:

	Learner c.	*Expert c.*
Texts	105	72
Tokens	39,244	144,291
Average text length (tokens)	373.75	2004
Standardised T/T ratio (basis: 300)	59.57	60.61

Table 1. Number of texts, total number of tokens, average text length and the standardised type/token ratio in the learner and in the expert corpus.

As can be seen, the texts in the expert corpus are longer than those in the learner corpus, but only slightly lexically more varied. As the students were given a specific task to complete under precise time constraints, their texts were expected to be rather repetitive. Also, the students being non-native speakers of English, they were thought to be less ‘creative’ than the expert writers. In fact, the figures for the standardised type/token ratio suggest that the expert writers do not

necessarily use a wider repertoire of word types in their texts than the learners do.

In order to compare the frequency of pattern occurrence across the two corpora, the linguistic counts of most patterns were normalised to their rate of occurrence per thousand words. For some patterns, however, the linguistic counts are given as percentages of the total number of *there* existential constructions.

5. Results and discussion

This section explores and discusses the features of the colligational patterns which describe existence and location in the learner and in the expert corpus respectively. Some of the patterns produced by the learners had been perceived as peculiar and non-native-like already on first reading them, and were, therefore, deemed worth further investigation. In the following sub-sections, the data being investigated are presented as sets of related patterns or pattern types.

5.1. Pattern type 1: there is/'s/are + *noun phrase*

The first set of patterns contains the existential constructions *there is/are/'s.* A glance at Table 2 suggests that the students use more existential clauses than the experts do (if the normalised values are considered). Both of them, however, use the pattern *there's* with about the same frequency.

	Learner c.	*‰ of tokens*	*Expert c.*	*‰ of tokens*
there is	141	3.59	127	0.88
there's	18	0.46	74	0.51
there are	138	3.52	157	1.09
Total	297	7.57	358	2.48

Table 2. Occurrences and standardised frequency counts per thousand words (out of the number of tokens) of *there is, there's* and *there are* in the learner corpus and in the expert corpus.

The learners' remarkably frequent use of existential patterns is definitely mainly due to the instructions which they had to abide by when writing their texts. To be precise, they had been explicitly asked to describe the city of Padua in general and some of its main monuments in particular: the *Cappella degli Scrovegni, Palazzo del Bo'* and *Palazzo della Ragione.* In their texts they do indeed employ *there* constructions to introduce these attractions, and to do so they either use the original Italian names in inverted commas or provide an English translation (*the Scrovegni Chapel*). Also, a look at the concordance lines for *there are* reveals that the students do at times make use of 'non-native-like' temporal adjuncts before the pattern, such as *in every moment of the year (there are), during the year (there are), during all the year (there are)*, which the expert corpus does not contain. It presents, instead, such expressions as *throughout the year* and *all year round*, which proved to be useful phraseology for the students to memorise.

A closer look at these patterns (see Table 3) reveals that the learners do quite often employ them to introduce definite referents into the text or to locate them in space, while the experts nearly always use them to present or locate indefinite referents, as exemplified by (9) and (10) respectively:

(9) Then there is the suggestive [sic] Palazzo della Ragione, built in 1218.

(10) There is also a Heritage Centre in Carlisle Road […].

	Learner c.	*% there ex.*	*Expert c.*	*% there ex.*
there's/are/is + *definite ref.*	124	41.75	10	2.79
there *extended existentials*	152	51.18	79	22.07

Table 3. Occurrences and percentages of use (out of the total number of *there* existential constructions) of *there's/are/is* + definite referent and of *there* extended existentials.

Furthermore, the majority of the clauses containing these patterns in the learner corpus are extended ones, that is they not only express the referents' existence or location, but also provide further information about them (mainly with an embedded clause, as can be seen in 11). In the expert corpus, by contrast, bare existentials are more recurrent, as illustrated by (12):

(11) There are many facilities which allow you to practice your favourite sports.

(12) There are good views from the Gun Garden - a nice picnic spot.

These findings are in line with the results of Palacios Martínez and Martínez Insua's (2006: 222-223) investigation of two Spanish learner corpora. Their learners "tend to use non-minimal, extended *there* constructions more frequently than the native speakers do," and the authors suggest that in their data *there* constructions are used "as devices for drawing long and heavy subjects towards final position […] rather than as signalling a number of other pragmatic functions." As regards the data discussed in the present study, however, this difference may well be also due to the Italian students' attempts to write concise texts, and specifically to the fact that to do so they often pack information inside the post-modifier slot of noun phrases or in subsequent clauses. On the other hand, the experts' texts seem to be informed by a different pragmatic strategy, which involves the introduction of the referents first and the gradual addition of separate information units afterwards.

5.2. Pattern type 2: there is/'s + *plural head noun*

A noteworthy aspect of the experts' existential patterns which emerged from the analysis of the concordance output is the absence of number agreement between the noun phrase and the verb in some of the clauses. As can be seen from Table 4, 3.07% of the occurrences of *there is* and 3.07% of *there's* are indeed characterised by non-agreement, as exemplified by (13) and (14):

(13) There is also much interesting graffiti and two pillars brought here from Reculver Towers.

(14) There's beach space galore, delightful gardens and parks, a traditional pier and a host of excellent pubs and restaurants.

	Learner c.	*% there ex.*	*Expert c.*	*% there ex.*
there is + *plural ref.*	0	0	11	3.07
there's + *plural ref.*	0	0	11	3.07
there are + *singular ref.*	0	0	2	0.56

Table 4. Occurrences and percentages of use (out of the total number of *there* existential constructions) of *there is/'s* + plural referent and of *there are* + singular referent.

Another two cases of non-agreement involve the use of *there are* followed by the quantifiers *a wealth of* / *a wide variety of* and a plural head noun, as can be seen in (15). Furthermore, in seven instances a pattern can be noticed which consists in the use of *there is / there's* followed by such quantifiers as *a wide / extensive range of*, *a host of*, *an abundance of* and by plural head nouns. Even though in these cases there is indeed concord between the verbs and the head nouns, the fact that the post-modifiers consist of plural nouns might at first give the impression that there is no agreement, as (16) illustrates:

(15) There are a wealth of Leisure and Relaxation activities at Lucknam Park - you don't really need to leave the estate!

(16) There is an extensive range of excellent home made bar snacks and light meals served.

The majority of the occurrences of non-agreement were retrieved from the most persuasive genres represented in the expert corpus, that is the magazines and the web pages. This might suggest that non-agreement is used as part of a promotional strategy whereby a feature which is typical of spoken and possibly also of childlike English is used in writing with the aim of "communicating between often anonymous parental senders and readily identifiable childlike receivers" (Dann 1996: 249). Yet, some occurrences were also found in the tourist guides, which partly contradicts this speculation and leaves room for further investigation.

5.3. Pattern type 3: Adjunct of space + there is/'s/are + *noun phrase*

As can be noticed from the data displayed in Table 5, another prominent feature of the learner corpus is the students' preference for the existential construction *there is/'s/are* after a thematised adjunct of space, while in similar cases the expert writers seldom choose this option and favour other patterns instead.

Adjunct of space +	*Learner c.*	*% there ex.*	*Expert c.*	*% there ex.*
there is	112	37.71	11	3.07
there's	13	4.38	5	1.40
there are	80	26.94	4	1.12
Total	205	69.02	21	5.87
where there is/'s/are	35	11.78	5	1.40
in which there is/'s/are	6	2.02	1	0.28

Table 5. Occurrences and percentages of use (out of the total number of *there* existential constructions) of adjuncts of space + *there is/there's/there are* and of two specific location patterns.

In other words, when the students introduce new referents into the discourse and locate them in space with reference to the position of other known referents, they nearly always make use of such patterns as those exemplified by the following extracts:

(17) Next to the chapel there are the Eremitani museums.

(18) Near the palace there are the two main market squares […].

(19) Downstairs there are historical Paduan shops […].

Two specific realisations of this pattern which are over-represented in the learner data are *where there is/'s/are* and *in which there is/'s/are*, as illustrated by (20) and (21):

(20) Here you can visit the Quaranta's Room, where there is the podium of Galileo Galilei […].

(21) The Scrovegni Chapel, in which there's a complete cycle of frescos painted by the Tuscan artist Giotto.

As we see, the students favour the use of the relative pronouns *where* and *which* whose antecedents are already known referents.

5.4. Pattern type 4: Adjunct of space + is/are + *noun phrase*

In contexts similar to those described in sub-section 5.3, the expert writers mainly opt for constructions with locative inversion, which entails the omission of *there* and the thematisation of an adjunct of space and aims at better focusing on and describing the position of a given attraction or facility, as illustrated by the examples below:

(22) Near the Palace Pier is the Sea-Life Centre […].

(23) Further up Dyke Road is the 'Booth Museum' […].

(24) Downstairs is a cosy bar/venue open late for music […].

The learner data contains no instance of locative inversion, which suggests that the learners are not familiar with this construction. Locative inversion is relatively common in fiction, which often includes descriptions of settings (Biber *et al.* 1999: 954-955). In a way, therefore, its frequent use in the expert corpus stresses the

descriptive function of professional tourist publications. The omission of *there* after such adjuncts is also reminiscent of the language of literary travel writing, memoirs and travelogues, as illustrated by the following extracts from *Pictures from Italy* by Charles Dickens ([1846] 1913):

(25) On the first floor also, is a whole house, and a good large residence too.

(26) In the old refectory of the dilapidated Convent of Santa Maria delle Grazie, is the work of art, perhaps, better known than any other in the world: the Last Supper, by Leonardo da Vinci.

5.5. Pattern type 5: (Adjunct of space) + intransitive verbs expressing a positional state or dynamic verbs + *noun phrase*

Other verbs which are frequently used by the expert writers are *stand, lie, stretch, hang,* and *loom*. As can be seen in Table 6, the most recurrent ones (including their infinitive, third-person singular, past, past participle and *ing*-forms) in the expert data are *stand* and *lie*, which are mainly used without locative inversion (i.e. without an adjunct of space preceding them).

Various verb forms of	*Learner c.*	*‰ tokens*	*Expert c.*	*‰ tokens*
stand	2	0.05	52	0.36
lie	1	0.03	38	0.26
stretch	0	0	11	0.08
hang	0	0	6	0.04
loom	0	0	1	0.01
Adj. of space + lie/hang/ stand/stretch	1	0.03	19	0.13

Table 6. Raw frequency counts and standardised frequency counts per thousand words of verbs expressing a positional state or dynamic verbs.

Extracts (27) to (30), taken from the expert corpus, exemplify this pattern:

(27) At the highest point of Rye stands St Mary's Church, […].

(28) Eastward lies Southsea, an Edwardian seaside resort.

(29) Beyond the green belt spreads a sizeable stretch of suburbia.

(30) Above his tomb hang copies of his funeral achievements, […].

The students, by contrast, hardly use verbs other than *be* in such contexts, which not only indicates that the deployment of such alternative verbs is a native-like feature, but also suggests that it might represent a characteristic of professional tourist writing.

5.6. Pattern type 6: (is/are) + (adverb) + situated/placed/located/set

The last set of location patterns consists in the optional use of *is/are* and an adverb followed by the past participles of either one of the transitive verbs *situate, locate, place* or *set*. These and other past participles can indeed be used to locate a given referent in space in combination with 'location elements' consisting in either complements or adjuncts (Huddleston/Pullum 2002: 680). Table 7 offers a closer look at this pattern type and shows that *is/are + situated/ placed/located* is over-represented in the learner corpus, while *is/are set* is only used by the expert writers. Also, the experts favour the use of the plain past participles, with a preference for *set*.

According to Manca (2012), dictionaries often describe these verbs as synonymous and interchangeable, while she argues that, at least in the corpus of British farmhouse holiday websites that she analysed, they tend to display different collocational profiles. That is, as they are used to locate different kinds of items in space, they are not always interchangeable. Following Manca's classification, it can be said, for instance, that in the expert corpus *situated* is used to alternatively express 'general' location (e.g. "*the town is situated on the Atlantic Heritage Coast"*), 'measured' location (e.g. "*situated 4 miles north of Littlehampton"*) or 'specific' location (e.g. "*situated underneath the Grafton multi-storey car park"*), while in the learner corpus it only expresses general and specific location (e.g. "*situated*

between Venice and Treviso", *"situated near Piazza delle Erbe e della Frutta"*).

	Learner c.	*%o tokens*	*Expert c.*	*%o tokens*
- is/are situated - is/are + adv. + situated	32 0	0.82 0	19 2	0.13 0.01
- situated	14	0.36	18	0.12
- is/are placed - is/are + adv. + placed	9 0	0.2 0	0 6	0 0.04
- placed	3	0.08	0	0
- is/are located	8	0.20	9	0.06
- located	8	0.20	11	0.08
- is/are set - is/are + adv. + set	0 0	0 0	6 2	0.04 0.01
- set	2	0.05	42	0.29

Table 7: Raw frequency counts and standardised frequency counts per thousand words of the pattern (Adj. of space) + *is/are* + *(adv.)* + *situated/located/placed/set*.

At times the learners also use the erroneous pattern 'Adjunct of space + *is/are* + *situated/placed* + noun phrase', thus producing clauses such as (31) and (32):

(31) *Near the Chapel is situated the civic museum of the Eremitani;

(32) *the 'Sala dei Quaranta' hall where is placed Galileo Galilei's podium.

Both these clauses contain an apparent syntactic mistake, which is likely due to the influence of the Italian language. That is, the learners use the adjunct of space at the beginning of the clause and postpone the noun phrase acting as subject after the verb, which the English syntax does not allow for. The students could have opted instead for locative inversion, which, as already said, in professional writing is mainly associated with the choice of verbs under-represented in the learner corpus, such as *stand* and *lie*.

The students use *placed* to express any type of location with the exception of measured location (e.g. "*placed in the north-east of Italy"*; "*placed in the Sala dei Quaranta room"*), whereas the experts

never choose it. Furthermore, while the experts employ *is/are + adverb + placed/situated/set* to indicate "convenient" location (Manca 2012: 389), such as "*Dover is ideally placed to*", "*Youth hostels are well placed along the coast for stops*", "*Hastings is ideally situated for the ports of Dover"*, the students never do so.

A look at the most recurrent content words which collocate with the patterns *is/are + situated/located/placed* in the British National Corpus[2] (see Table 8) indicates that *is/are placed* is not used as a location pattern in this large reference corpus either, while collocations with *located* and *situated* are instead often used to this end.

is/are located		*is/are situated*		*is/are placed*	
centre	21	area	36	emphasis	65
hotel	19	hotel	35	context	26
village	16	centre	31	box	10
site	11	village	30	reliance	9
cells	9	miles	26	stress	9
sites	8	town	21	foot	9
housing	8	road	21	layer	8
chromosome	7	city	19	orders	8
residential	7	south	19	magnetic	7
island	7	park	17	tank	7

Table 8. The most recurrent collocates of *is/are + situated/located/placed* in the BNC (all registers; span = 8).

As can be seen, in the British National Corpus *placed* often co-occurs with words such as *emphasis*, *context* and *reliance*, which suggests that it is mainly used metaphorically and not to describe physical spaces.

2 <http://corpus.byu.edu/bnc/x.asp>, last visited on 25th June 2013.

6. Conclusions

This chapter has explored some existential and location patterns in a corpus of promotional tourist texts written by learners and has compared them to those found in a control corpus made up of texts written by experts. The methodology has, therefore, involved the compilation of the two corpora and their contrastive analysis. Not only has the study aimed at investigating this specific type of learner language *per se*, but has also sought to take advantage of the learner data to identify some salient aspects of the expert data and to bring them into focus.

The analyses conducted have revealed that some existential and location patterns are used in one corpus and not in the other, yet the majority of them are represented in both but with different frequencies. Firstly, the learners favour extended existentials to locate definite referents, while the strategy adopted by the experts entails using bare existentials with indefinite referents. Secondly, only the experts produce clauses with non-concord between *there is/'s/are* and the head noun, and are likely to do so to give a flavour of informality to their texts. Thirdly, the learner dataset presents a large number of instances in which the pattern *there is/'s/are* is used with locative inversion, whereas the expert data reveals that in such cases the professional writers preferably opt to either omit existential *there* or to use such verbs as *stand* and *lie*. Fourthly, the pattern *is situated/placed/located* is very recurrent in the learner corpus, while in the expert corpus it is the bare forms of the verbs *set, situated, located* that are preferred. Also, the patterns *placed* and *is placed* are not present in this latter corpus, which suggests that these verb forms should not be used to describe physical location in space. Fifthly, the experts employ *is/are* + *adverb* + *placed/situated/set* to express "convenient" location, while the learners never do so. Finally, the analysis of the learner corpus has revealed the presence of some non-native-like and some erroneous patterns, such as the use of unusual temporal adjuncts or the wrong pattern 'Adjunct of space + *is/are* + *situated/placed* + noun phrase'.

We believe that this study has some implications for language teaching. The "learning-driven data" (Seidlhofer 2002: 231) and the findings discussed here could indeed be used to prepare 'data-driven' and 'quick and dirty' activities (e.g. Tribble/Jones 1990; Seidlhofer 2002), which would make it easy for the students to notice and critically look at the differences between their own texts and the professional ones. It would also give them the opportunity to improve their skills in writing promotional tourist texts. All this, however, would require caution and appropriate teacher guidance, as students' direct access to learner data might confuse learners "by showing incorrect and inappropriate usage and encouraging wrong generalizations" (Zagrebelsky 2004: 58). This study does also have methodological implications for further investigations into the specific lexico-grammatical and textual features of professional tourist publications and possibly also into those of other specialised languages. To be precise, this study supports the claim made in section 1: if compared to an expert corpus, even a small yet well designed learner corpus has the potential to reveal not only learner deficiencies but also some of the peculiar features of the text type it represents. This holds especially true for such blended text types as promotional tourist texts, whose linguistic features are difficult to pin down.

References

Agorni, Mirella (ed.) 2012. *Prospettive Linguistiche e Traduttologiche negli Studi sul Turismo*. Milano: Franco Angeli.

Biber, Douglas / Johansson, Stig / Leech, Geoffrey / Conrad, Susan / Finnegan, Edward 1999. *The Longman Grammar of Spoken and Written English*. London: Longman.

Calvi, Maria Vittoria 2000. *Il linguaggio spagnolo del turismo*. Viareggio: Baroni.

Clark, Eve V. 1978. “Locationals: 'existential', 'locative', and 'possessive' constructions.” In Harold, Joseph / Charles, Greenberg / Ferguson, Albert / Moravcsik, Edith A. (eds.) *Universals of human language*: *Vol. 4, Syntax*, 85-126.

Castello, Erik 2002. *Tourist Information Texts: A Corpus-based Study of Four Related Genres.* Padova: Unipress.

Council of Europe 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press

Curado Fuentes, Alejandro 2007. A Corpus-based Assessment of Reading Comprehension in English for Tourism Studies. In Hidalgo, Encarnación / Quereda, Luis / Santana, Juan (eds) 2007. *Foreign Language Classroom: Selected Papers from the Sixth International Conference on Teaching and Language Corpora (TaLC 6)*: Amsterdam: Rodopi, 309-326.

Dann, Graham 1996. *The Language of Tourism: A Sociolinguistic Perspective*. London: CAB International.

Dann, Graham 2012. Remodelling a Changing Language of Tourism: from Monologue to Dialogue and Trialogue. *Pasos* 10/4, 59-70.

De Stasio, Clotilde / Palusci, Oriana (eds) 2007. *The Languages of Tourism: Turismo e Mediazione*. Milano: Unicopli.

Dickens, Charles [1846] 1913. *Pictures from Italy*, Price, David (ed.). London: Chapman & Hall. <http://www.gutenberg.org/ebooks/650>, last visited on 31st October 2012.

Downing, Angela / Locke, Philip 2006. *English Grammar: A University Course*. London: Routledge.

Fodde, Luisanna / Denti, Olga 2012. Il discorso turistico: peculiarità linguistico-comunicative nella didattica dell’inglese specialistico. In Agorni, Mirella (ed.), 23-46.

Gerbig, Andrea / Shek, Angela 2007. The Phraseology of Tourism: A Central Lexical Field and its Cultural Construction. In Skandera, Paul (ed.) *Phraseology and Culture in English.* Berlin: Mouton de Gruyter, 303-322.

Givón, Talmy 1993. *English Grammar: A Functional-Based Introduction* (Vols. 1 and 2). Amsterdam: John Benjamins.

Gotti, Maurizio 2006. The Language of Tourism as Specialized Discourse. In Palusci, Oriana / Francesconi, Sabrina (eds) 2006.

Translating Tourism: Linguistic / Cultural Representations. Trento: Editrice Università degli Studi di Trento, 15-34.

Gotti, Maurizio 2011. *Investigating Specialized Discourse.* Bern: Peter Lang.

Granger, Sylviane 2002. A Bird's-eye View of Learner Corpus Research. In Granger, Sylviane / Hung, Joseph / Petch-Tyson, Stephany (eds) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 3-33.

Halliday, Michael 2004. *An Introduction to Functional Grammar*, revised by Matthiessen Christian. London: Arnold.

Hidalgo, Encarnación / Quereda, Luis / Santana, Juan (eds) 2007. *Foreign Language Classroom: Selected Papers from the Sixth International Conference on Teaching and Language Corpora (TaLC 6)*: Amsterdam: Rodopi.

Hoey, Michael 2005. *Lexical Priming: A New Theory of Words and Language*, London, Routledge.

Huddleston, Rodney / Pullum, Geoffrey 2002. *The Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.

Jespersen, Jens Otto Harry 1924. *The Philosophy of Grammar.* London: The University of Chicago Press.

Kelly, Dorothy 2000. Text Selection for Developing Translator Competence: Why Texts from the Tourist Sector Constitute Suitable Material. In Schäffner, Christina / Beverley, Adab (eds) *Developing Translation Competence*. Amsterdam: John Benjamins, 157-167.

Lakoff, George 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* Chicago: The University of Chicago Press.

Lam, Peter 2007. A Corpus-driven Lexico-grammatical Analysis of English Tourism Industry Texts and the Study of its Pedagogic Implications in English for Specific Purposes. In Hidalgo, Encarnación / Quereda, Luis / Santana, Juan (eds) *Foreign Language Classroom: Selected Papers from the Sixth International Conference on Teaching and Language Corpora (TaLC 6)*: Amsterdam: Rodopi, 71-89.

Manca, Elena 2007. *Beauty* and *Tranquillity* in the Language of Tourism: Linguistic and Cultural Reasons. In De Stasio, Clotilde / Palusci, Oriana (eds) 2007. *The Languages of Tourism: Turismo e Mediazione*. Milano: Unicopli, 113-128.

Manca, Elena 2012. Synonymy and Language Teaching. In Dalziel, Fiona / Gesuato, Sara / Musacchio, Maria Teresa (eds) *A Lifetime of English Studies. Essays in Honour of Carol Taylor Torsello*. Padova: Il Poligrafo, 381-391.

Martínez Insua, Ana / Palacios Martínez, Ignacio 2003. A Corpus-Based Approach to Non-Concord in Present Day English Existential There Constructions. *English Studies* 84/3, 262-283.

Moya Guijiarro, Arsenio Jesús / Albentosa Hernández, José Ignacio 2001. Points of Departure in News Items and Tourist Brochures: Choices of Theme and Topic. *Text* 21/3, 347-371.

Palacios Martínez, Ignacio / Martínez Insua, Ana 2006. Connecting linguistic description and language teaching: native and learner use of existential *there*. *International Journal of Applied Linguistics* 16/2, 213-231.

Palusci, Oriana / Francesconi, Sabrina (eds) 2006. *Translating Tourism: Linguistic / Cultural Representations*. Trento: Editrice Università degli Studi di Trento.

Scott, Mike 2008. *WordSmith Tools 5*. Liverpool: Lexical Analysis Software.

Seidlhofer, Barbara 2002. Pedagogy and Local Learner Corpora: Working with Learning-driven Data. In Granger, Sylviane / Hung, Joseph / Petch-Tyson, Stephany (eds) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 213-234.

Tognini Bonelli, Elena / Manca, Elena 2002. Welcoming Children, Pets and Guests: Towards a Functional Equivalence in the Languages of 'Agriturismo' and 'Farmhouse Holidays'. In Aijmer, Karin / Altenberg, Bengt (eds) *Advances in Corpus Lingustics*. Amsterdam: Rodopi, 371-385.

Tribble, Christopher / Jones, Glyn 1990. *Concordances in the Classroom: A Resource Book for Teachers*. London: Longman.

Zagrebelsky, Maria Teresa 2004. *Computer Learner Corpora. Theoretical Issues and Empirical Case Studies of Italian Advanced EFL Learners' Interlanguage*. Alessandria: Edizioni dell'Orso.

EUGENIA DAL FOVO

15. The Language of Interpreters on Television: Characteristics, Tendencies and Idiosyncrasies

> As you are reading these words, you are taking part in one of the wonders of the natural world. For you and I belong to a species with a remarkable ability; we can shape events in each other's brains with exquisite precision. I am not referring to telepathy or mind control or the obsessions of fringe science; even in the depictions of believers these are blunt instruments compared to ability that is uncontroversially present in every one of us. That ability is language. (Pinker 1994: 15)

This quote clearly underlines the connection between language and its general purpose, i.e. communication, as opposed to languages for special purposes (LSPs), which serve a specific communicative purpose, determined mainly by the context in which communication takes place and biased by such factors as situational specificities and the interlocutors involved. On the basis of this claim, and examining the specific contextual and situational aspects of interpreter-mediated communication on television, is it possible to identify an LSP of interpreting, a "language of interpreters working on television"? This is the question that will be addressed in this chapter by presenting a few attempts to answer it on the part of Interpreting Studies (IS) scholars, who applied the corpus-based approach to research in the field of television interpreting.

1. Setting the framework: language for special purposes

Studies on LSPs account for a long-standing and particularly rich field of research. The corpus-based approach has enabled analysts to

provide many definitions of LSPs, according to different perspectives and parameters. LSPs have been described mainly from a lexical point of view: Halliday (1978: 138-139), for instance, identifies LSPs through a very specific phenomenon, i.e. "foregrounding", namely "the phenomenon of linguistic highlighting, whereby some features of the language of a text stand out in some way". Cortelazzo (cf. 1990: 6) defines LSPs as languages that follow their own specific rules and symbols; more specifically, according to Gotti (1991), LSPs are languages that share communication conventions with a given general language and, at the same time, have conventions of their own that are not included in the set of linguistic features of that particular language[1]. The very denomination of LSPs, however, suggests that the difference between general language and LSPs may also be found in the use speakers make of them: the special purpose of LSPs might be ultimately defined as the use of language by professionals/experts when referring to typical aspects of their own professional field or domain of expertise (Gotti 2003).

In this respect the language used on television could in a way – and has been to some extent – be defined as an LSP, as it has its own rules and limitations, responding mainly to the imperatives and ethics of entertainment (cf. Katan/Straniero Sergio 2001, 2003; Straniero Sergio 2007). In this case, television context and setting are a pivotal defining aspect of a language whose specific purpose is ultimately being *telegenic*: according to Katan/Straniero Sergio's (2003) view, television making is almost invariably driven by a set of ideologies, mainly consumer capitalism and popular culture. Broadcasters' actions and decisions are heavily dependent on the viewer's opinion, whose "deadly click of the thumb" (Hartley 1992: 97), namely his/her switching to another channel, must be avoided in order to maintain a high level of viewing share. Viewers' expectations and subsequent appreciation of broadcast content are strongly related to the degree to which it satisfies their "'comfort factor': the degree to which the TV

1 "[Q]uei linguaggi che, pur condividendo in generale le convenzioni comunicative di una data lingua, ne posseggono anche alcune che non sono comprese nel patrimonio linguistico comune" (Gotti 1991: 7).

audience is entertained" (Katan/Straniero Sergio 2003: 131). In the case of interpreter-mediated communication on television, viewers' entertainment requires the interpreter to use a language which is conversational and decent, faithful to the original speaker's register and yet marked by the interpreter's own distinctive informal and witty style. Furthermore,

> the viewer[s] will expect to stay within the safety of their own bubble, and will expect the interpreter to […] localize and appropriate the message, so that the interpreted event is transferred to within the safety and comfort of the viewer's cognitive bubble. (Katan/Straniero Sergio 2003: 134).

On the basis of the above-mentioned assumptions, the research question can now be reformulated as follows: given the specific linguistic constraints within which an interpreter on television has to perform his/her task, is it possible to identify recurring features *foregrounding* (Halliday 1978: 138-139) the language of interpreters on television? If this is the case, how can such features be identified?

2. Looking for a television interpreters' language: the corpus-based approach

The focus of the present chapter is communication mediated by an interpreter and, in particular, those lexical choices interpreters make to mediate communication when translating messages from one language to another in a specific context, i.e. television.

Interpreter-mediated communication, in very simple terms, means verbal re-expression of a message originally uttered in a given source language (SL) − be it general language or LSP − in a given target language (TL). Interpreters are producers of texts of which they are not the owners in semantic terms: their speaking turns and the message they express do not involve any sort of (or very little) "semantic autonomy" (Straniero Sergio, in press), and yet they are entitled to speaking turns within the communication structure as much

as any other participant. In taking part in communication, interpreters practice their profession, making the message intelligible for the other interlocutor(s). The use they make of language has the specific purpose of serving communication and has to enable them to do so within specific constraints and limitations, some of them very practical and common to all interpreting settings, such as time constraints for simultaneous interpreting, and others which are more context-specific, such as medium-related norms and requirements (e.g. entertainment in television settings, as already mentioned above, cf. Katan/Straniero Sergio 2003). In order to be able to make statements about what this language actually does consist of while maintaining an "evaluation-free and description-oriented" (Straniero Sergio 2012: 211) attitude, it is useful to resort to a corpus-based approach. As a matter of fact, being able to analyse all variables on a large scale may indeed help interpretration scholars identify "a number of regularities or recurring patterns in interpreters' translational behaviour" (Straniero Sergio, in press) − some sort of set of indicators of their recurring lexical choices that remain unvaried across television genres and over time − and ultimately identify the special language interpreters use on television, if any.

In the field of translation studies (TS), corpus-based TS (CTS) have so far produced a considerable amount of research work (among others, Baker 1996, 2004; Kenny 2001; Laviosa 2002; Olohan 2004; Anderman/Rogers 2008) with the aim of investigating "the nature of translated texts as mediated communicative events" (Baker 1993: 263). It is worth noting that, by identifying recurring elements that help define a text as "translated" on the basis of various factors − the most prominent being lexical aspects − researchers are able to study a translator's language behaviour and identify his/her own translational "thumbprint" (Baker 2000: 244).

Corpus-based Interpretation Studies (CIS) were born in order to pursue a very similar goal: "to search for regularities in interpreters' output and provide explanations for them" (Straniero Sergio/Falbo 2012: 15). More than a decade ago Shlesinger (1998: 487) wrote:

> many of the observations encountered in the literature on interpreting are based on sparse, often anecdotal data […]. The compilation of bilingual and

> parallel corpora is indeed overdue, given the potential to use large, machine-readable corpora to arrive at global inferences about the interpreted text.

Nonetheless, since then, CIS "have remained in their infant stage" (Straniero Sergio/Falbo 2012: 9). This delay is, however, not surprising: making interpretation corpora electronically available to the scientific community requires going through a number of stages, some of which are common to CTS, such as corpus design and classification, whereas others are specific of oral texts and particularly onerous and time-consuming, such as transcription of spoken data. As Cencini/Aston (2002: 47) wrote, "Like all speech, interpreting dies on the air. In order to study it, we need to resurrect the corpse by recording and transcribing it, thereby transforming the corpse into a corpus". Some of the latest developments in CIS are illustrated in the works of Setton (2011) and Straniero Sergio/Falbo (2012). Both studies clearly illustrate the difficulties of applying the corpus-based approach to Interpretation Studies (IS), and yet underline as clearly the necessity of resorting to corpora in order to achieve "a higher degree of descriptive adequacy" (Granger 2003: 19), as opposed to the long-standing tradition of prescriptive and evaluation-oriented studies that has been a distinctive feature of research on interpreting in its early stage. In 3.1 some of the studies on interpreters' translational behaviour will be presented, which highlight the need for a description-oriented approach to the analysis. In 3.2 four studies conducted on the Television Interpreting Corpus (CorIT) will be illustrated, with the aim of addressing the research question through corpus-based research combined with a strictly description-oriented approach to the analysis.

3. Looking for a television interpreters' language through CIS: early studies vs. CorIT

CIS have been rapidly gaining momentum in the past few years, providing for a significant volume of investigations, focusing on various aspects of interpreters' output.

3.1. Corpus-based studies on the language of interpreters: state of the art

When studying language aspects, one of the major difficulties IS analysts have to face is limiting their observations to the interpreter's output as such, namely the interpreted text (IT). Indeed, the interpreting *product* (IT) can never be entirely isolated from the interpreting *process*, as recurring choices are frequently driven not only by external constraints and conditions, but also internal (conscious, deliberate) habits and decisional patterns. A few examples may clarify this point: when observing the occurrence of explicitation of place names in ITs, Baumgarten *et al.* (2008: 198) speak of a strategic dimension of interpreting and come to the conclusion that "explicitation is not a universal feature of interpreting but rather an option that may be more frequent in certain settings or interpreting modes". As Marzocchi (2005: 29) very insightfully puts it, "it may be the case that norms start their lifecycle as a strategy to cope with cognitive constraints in a given situation and are then internalized and generalized". Indeed, identifying recurring (universal) features frequently lead the researcher to define them as the result of on-line planning and self-monitoring activities performed by interpreters (see Tissi 2000, Petite 2005, Papa 2010). Another case in point are corpus-based studies on lexical density, lexical patterns and disfluencies conducted on EPIC (European Parliament Interpreting Corpus) by Russo *et al.* (cf. Bendazzoli/Sandrelli 2005, 2009; Sandrelli/ Bendazzoli 2005; Russo *et al.* 2006; Bendazzoli *et al.* 2011). Despite efforts to keep their analysis to a strictly observational level, on the basis of Laviosa's (1998) study on lexical density in the Translational English Corpus (TEC), results were inevitably classified using a mixture of descriptive (e.g. repetition) and evaluative (e.g. self-correction, explanation) definitions (cf. Russo *et al* 2006). Indeed, by describing an item or recurring feature of the IT as self-correction, for instance, the analyst ascribes to the interpreter a specific intention, thereby expressing an evaluation of the interpreter's choice, instead of simply defining the linguistic phenomenon as it appears in the IT. Nonetheless, the above-mentioned studies proved extremely useful in

order to assess the validity of the corpus-based approach to IS. Indeed, analyses of lexical patterns conducted on EPIC represent a successful endeavour to avoid a mere quality-assessment approach to IS, based on pre-established standards and shared expectations, in favour of a more neutral, evaluation-free and description-oriented perspective. This is also the main principle driving the case studies conducted on the Television Intepreting Corpus (CorIT, cf. Straniero Sergio/Falbo 2012).

3.2. Corpus-based studies on the language of interpreters: CorIT and the studies on interpreters' style.

CorIT is a collection of more than 2,700 recordings of interpretations delivered on Italian television in the last 50 years, from various foreign languages into Italian, in different interpreting modes[2] and within various communication settings and interaction types. It is an "open, multimedia, partially parallel spoken corpus" (Falbo 2009: 107). CorIT is currently being transcribed with the software WinPitch (cf. Martin 2005, 2009), in a way that provides access to its content as a whole, i.e. "making audio and video tracks available simultaneously, as a constant reminder of the multimedia dimension of the text, as opposed to a simple transcript" (Falbo 2012: 175).

CorIT represents an unprecedented opportunity to look for recurring features *foregrounding* the language television interpreters use, since, as Straniero Sergio (2012: 211) points out,

> What makes CorIT [...] significantly different from other interpreting-based corpora is the availability of a large number of simultaneous (SI) and/or consecutive interpretations (CI) delivered by the same interpreter over a period of 15-20 years.

This has led to a series of case studies conducted on CorIT items focusing on the translational behaviour of individual interpreters

2 Simultaneous, consecutive, liaison, etc. (cf. Falbo 2012: 161).

working for Italian television, with the aim of identifying their individual characteristics, tendencies and idiosyncrasies through the language they use and aspects that define their own distinctive manner of interpreting on television, in other words, their own interpreting style. As already mentioned, so far IS have been concentrating mainly on the quality of interpreters' performances, only rarely focusing on interpreters' style. Unlike quality, style does not require any sort of evaluation of the interpreter's output. It is rather the sum of a series of indicators distinguishing one particular *modus interpretandi* from the others, indicators that are idiosyncratic and constantly present in time. Admittedly, even in those very few studies on style, such as Yagi (2000) and Van Besien/Meuleman (2008), quality and style are considered equivalent (Yagi 2000) and the definition of style still includes strategies alongside idiosyncratic traits (Van Besien/ Meuleman 2008). The analysis template applied to the following case studies on style conducted on CorIT, conversely, avoids any reference to quality, limiting any observation to those aspects of interpreters' output that are irrespective of their situational and contextual behaviour.

3.2.1. CorIT case study no.1: Using corpus evidence to discover style in interpreters' performances

Straniero Sergio's (2012) study focuses on the translational behaviour of three different interpreters, whose interpreting performances on television cover a period of 20 years of activity. Straniero Sergio identifies a series of distinctive features of each interpreter's output accounting for his or her *modus interpretandi*. Such indicators are defined "style indicators", namely language aspects "not subject to variation over time in individual interpreting performance" (Straniero Sergio 2012: 213). Selected style indicators are lexical choices and language use, interpreter-generated discourse markers (not present in the source texts), and décalage (segmentation of the input/output flow). The study includes the case of "the 'extraordinary' interpreter" (see Examples 1, 2, 3, 4), who uses the term *straordinario* (extraordinary) to translate 23 different terms of the English language over a period of 8 years, in various communicative situations (press

conferences, talk shows, presidential debates). Such lexical choice is considered a *passepartout*, namely a translational solution applied not just to specific words, but an entire semantic field.

(1) **amazing** award
[premio s**traordinario** davvero]
(Academy Awards, 23.3.2003)

(2) a **brilliant** script
[un copione veramente **straordinario]**
(Domenica in, Rai Uno, 15.12.1996)

(3) You have done it in front of your home fans and you must be **delighted**
[stavolta l'hai fatto di fronte ai tuoi tifosi: ed è stata forse una cosa **straordinaria]**
(GP F1 Italia, Rai Uno, 10.9.2000)

(4) they have done a **difficul**t job
[hanno fatto un lavoro **straordinario]**
(Serata TG1, 19.12.1998)

This and similar results obtained confirm the initial hypothesis, namely the presence of some distinctive features of each interpreter's output accounting for his or her style, thus signalling his or her recognisability. The very aim of Straniero Sergio's (2012) study might appear to be in contrast with that of the present chapter, as it focuses on "characteristics, habits and idiosyncrasies" and ultimately personal ways of using language "of an individual interpreter" (Straniero Sergio 2012: 211) instead of looking for recurrent features common to all interpreters' use of language. It is not, however, an entirely divergent course of analysis. In fact, aspects identified as style indicators in this study are selected on the basis on their diachronic frequency, extrapolated from interpretations carried out in different TV genres. This approach ensures that the resulting features are not dependent on any situational or contextual constraints and consequent application of *ad hoc* strategies, that is, they ultimately display that general character required to properly identify a "language of interpreters".

3.2.2. CorIT case study no.2: Television interpreter' style

Straniero Sergio's (2012) study was replicated by Trisciuzzi (2010)[3] in his unpublished master dissertation. Trisciuzzi applies Straniero Sergio's analysis template to two ITs delivered by two different interpreters working on television, namely the interpretation of two press conferences broadcast on Italian television. The first one is a press conference between former Italian Prime Minister Silvio Berlusconi and former US President Bill Clinton, that took place in Rome on June 2, 1995 for the 50th anniversary of the liberation of the Italian capital city. Interpreters here work *in praesentia* (Falbo 2012: 163), sharing the *hic* and *nunc* with the primary interlocutors. The second one is the news-conference delivered by former US President George W. Bush at the White House on April 13, 2004, in which Bush disclosed a CIA report he received in 2001, five weeks before the 9/11 attack. Interpreters here work *in absentia* (Falbo 2012: 164), as they do not "participate directly in the event", not sharing the *hic* with the primary interlocutors, "who do not require any translation to communicate with each other" (Falbo 2012: 164). By analysing the two interpreters' lexical choices and use of discourse markers, Trisciuzzi identifies the personal style of each interpreter. Trisciuzzi's results confirm what emerged from Straniero Sergio's (2012) study: both interpreters' outputs display recurring features, irrespective of the different interpreting modes (*in praesentia* and *in absentia*) characterising the two press conferences.

3.2.3. CorIT case study no.3: Double Renditions in simultaneous interpreters' output

Straniero Sergio's (in press) study aims at describing structural, semantic and sequential features of 1,544 double renditions (DRs) found in the simultaneous interpretations (SIs) from English, French

---

3 Trisciuzzi had access to Straniero Sergio's study before its publication in 2012, i.e. after its completion in 2010. Trisciuzzi's master dissertation is therefore subsequent to Straniero Sergio's study, despite the year of publication.

and Russian into Italian, delivered by 90 interpreters over a period of 25 years, for a total analysed output of 96 hours, 11 minutes and 40 seconds. The term "double renditions" (DRs) was introduced by Straniero Sergio in order to describe the phenomenon of interpreters producing more than one translated text (TT) equivalent for the same source text (ST) lexical unit. More specifically, DRs are identified as two equivalent semantic units produced one after another, with the same intonation contour and no pause longer than 3 seconds between the first and the second DR item(s), as in the example below[4]:

(5) I saw the tears of the families whose lives were lost…
[ho visto le lacrime **delle famiglie dei familiari** che hanno perso i loro cari…]
*I saw the tears **of the families of the relatives** who lost their beloved*
(Tg1, 12.10.2001)

The starting hypothesis is that "DRs are not only cognitive (i.e. peculiar to the nature of simultaneous interpreting), but also idiosyncratic (i.e. peculiar to an individual interpreter's behaviour)" (Straniero Sergio, in press). This phenomenon has already been observed in IS, and has been defined in several ways, such as expansion, repetition, explicitation, and self-repair. Each one of these definitions does not, however, suit the aim of Straniero Sergio's study, since, as in the case of previously mentioned studies on EPIC, they all describe different operations by ascribing different intentions to the interpreter. The term "double rendition", on the other hand, simply indicates the presence of two equivalents in the target text (TT), i.e. ST units rendered twice. A summary of the analysis and preliminary results is provided in the following paragraphs[5]. Possible combinations of DRs are: noun (6), verb (7), adj+noun (8), clause (9).

(6) and we have taken some action so far

4 Examples from CorIT transcripts consist of the ST excerpt, followed by the relevant TT portion and its glossed translation into English in *italics*. They do not display any punctuation, except those elements signaling specific aspects of spoken discourse, in line with the relevant transcription conventions adopted for CorIT (Falbo 2012).

5 For further details, reference should be made to Straniero Sergio's (in press).

[finora abbiamo adottato **delle misure degli interventi]**
so far we have adopted measures actions
(Tg1, 7.11.2008)

(7) …to restore peace and justice
[...per **restaurare ripristinare** la pace e la giustizia]
…to reinstate to restore peace and justice
(Tg1, 17.1.1991)

(8) …these people are having some sort of sexual affair
[…che queste persone ha- hanno **un rapporto sessuale una storia d'amore]**
…that those people have a sexual relationship a love story
(Tmc, 21.9.1998)

(9) on ne peut pas dire des choses comme cela
[non si possono dire queste cose non si possono fare queste affermazioni]
you cannot say these things you cannot make these statements
(Tg3, 28.4.1988)

In addition to this, Straniero Sergio identifies 6 modalities or patterns of equivalents co-selected in the TT, on the basis of the semantic bond existing between them and in relation to the ST. These are not included in the present chapter. After presenting quantitative aspects of his analysis, Straniero Sergio focuses on the qualitative perspective, discussing results on the basis of the possible strategies they may correspond to. What is relevant to the present chapter is the following: DR variability does not appear to be related to ST textual features so much as to a single interpreter's style, "that is to say the habit of glossing and paraphrasing his or her target text" (Straniero Sergio, in press). In other words, DRs do not appear to serve the purpose of facilitating ST processing, nor of making utterances more intelligible for the audience, thereby confirming the initial "methodological assumption that the interpreter, through DR, reformulates not only the original speech (interlinguistic level), but also his or her speech (intralinguistic level), acting as a producer of synonymic and paraphrastic strings." Bearing in mind the observations previously made on the nature of ITs, namely that they are texts interpreters produce with very little semantic autonomy, the author concludes that:

> the practice of retroactive elaboration, whereby alternatives are continuously provided, may even reveal a sort of self-complacency on the part of the interpreter, in showing off his or her linguistic skills. 'Playing with synonyms', then, would compensate the interpreter's lack of semantic autonomy, foregrounding his or her role as a producer of a text of which s/he is not the owner. (Straniero Sergio, in press)

3.2.4. CorIT case study no.4: The more the merrier? Teamwork and lexical variation in simultaneous interpretation on television

Dal Fovo's (2012) study is a corpus-based analysis conducted on interpretations of American presidential debates broadcast on Italian television. It was conducted on the CorIT sub-corpus 'Presidential Debates', and more specifically on the second Bush/Kerry debate held on October 9, 2004, and its two interpreted versions broadcast within two Italian *quoting broadcasts* (Falbo 2012: 174): SkyTG24 and TG5. Lexical variation is here referred to occurrences of one single term or lexical unit in the ST that are translated with more than one equivalent in the IT(s), either within the turn of one single interpreter or between two turns. Equivalents in the IT are analysed on the basis of their co-referential value as anaphoric referents (cf. Vanelli 1981, Simone 2001, Beccaria 2004) of their co-referring expression. The first occurrence of an equivalent in the IT is identified as the head of the anaphoric chain (HC). Such lexical units are divided into three main categories: single terms; noun groups; and culture-bound terms or noun groups. Equivalents appear in the forms of repetition/loanword (Tanskanen 2004); contextual synonym; hyponym; proform (Conte 1999); paraphrastic reformulation; substitution/sense reconstruction (Gile 1995; Wadensjö 1998; Straniero Sergio 2003); and double rendition (Straniero Sergio, in press).

The following examples illustrate lexical variation in the occurrences of the term "health care" and its IT equivalents. In the first passage (10) the same interpreter uses two different equivalents for the term "health care" within the same speaking turn (i.e. John Kerry's answer turn). The first occurrence of "health care" is translated by Kerry's interpreter (INK) as *assistenza sanitaria*, which is the formally closest alternative to the ST term and becomes the HC.

Later on, however, INK does not reiterate his first choice, selecting a contextual synonym instead: *copertura sanitaria*.

(10) He can't come here and tell you that he's created **health care** for Americans because, what, we've got 5 million Americans who have lost their **health care**, 96,000 of them right here in Missouri.
[Non può venire qui e dirvi che ha creato nuovi lavori qui per voi americani son st(ati) persi non ha aiutato con **l'assistenza sanitaria** (.) ci sono cinque milioni di americani che hanno perso la **copertura sanitaria** (.) er molte di queste proprio qui nel Missouri]
*He cannot come here and tell you that he has created new jobs for you, Americans, they were lost he did not help with **health care** there are 5 million Americans who have lost their **health insurance**, many of them right here in Missouri.*
(SkyTG24, 9.10.2004)

The following question (11) is addressed to Kerry. The interpreter translating both mediator and journalists (INM) chooses repetition of the HC, in that he repeats "the unit in an identical form" (Tanskanen 2004: 94).

(11) Senator Kerry, you've stated your concern for the rising cost of **health care**, yet you chose a vice presidential candidate who has made millions of dollars successfully suing medical professionals. How do you reconcile this with the voters?
[Senatore (.) Lei ha parlato delle sue preoccupazioni per i: i costi sempre più alti nell'**assistenza sanitaria** però il suo candidato vicepresidente e: (.) ha erm citato in giudizio er diversi medici e er industrie farmacologiche]
*Senator you expressed your concerns about increasing **health care** costs but your vice presidential candidate has been suing several doctors pharmacological industries*
(SkyTG24, 9.10.2004)

It would appear that, by doing so, INM is ratifying the HC put forward by INK, implicitly acknowledging his contribution, according to the principles of *lexical entrainment* (Brennan/Clark 1996: 1483) and *grounding* (see Clark/Brennan 1991). From a different perspective, the HC ratification may be seen as the result of "historical factors" (Brennan/Clark 1996: 1483), namely *recency* (Brennan/Clark 1996: 1483) of utterance of the lexical unit in terms of time: indeed, when an

equivalent is activated by previous use, it is more available in the interpreter's working memory, and has therefore higher chance of reappearing (see also Levelt/Kelter 1982).

This hypothesis is confirmed in the following turn (12), in which INK translates Kerry's "health care" almost always with the ratified HC.

(12) I have a plan to lower the cost of **health care** for you. I have a plan to cover all children. I have a plan to let you buy into the same **health care** senators and congressmen give themselves.
[il nostro piano invece abbatterà i piani dell'**assistenza sanitaria** per voi (.) per i vostri figli **l'assistenza sanitaria** per i vostri figli e anche avere lo stesso livello di **protezione** che abbiamo noi al Congresso (.)]
*Our plan, on the other hand, will dismantle **health care** plans for you for your children **health care** for your children and also have the same level of **protection** that we have at the Congress*
(SkyTG24, 9.10.2004)

What happens in the TG5 IT, however, disclaims this first explanation. As INK did in the SkyTG24 IT, Kerry's interpreter on TG5 (INK2), as well, selects the formally closest alternative to the ST term, which also becomes the HC (13).

(13) He can't come here and tell you that he's created **health care** for Americans because, what, we've got 5 million Americans who have lost their **health care**, 96,000 of them right here in Missouri.
[Vorrebbe: er appunto venirvi a dire che ha: creato nuovi: posti di lavoro ma non lo ha fatto er: cinque milioni di americani hanno er per:so cinque milioni di americani hanno perso l'**assistenza sanitaria** (.)]
He would like to come here to tell you that he created new jobs but he did not five million Americans have lost five million Americans have lost their **health care**. (TG5, 9.10.2004)

As we can see in (14), however, the INK2-produced HC is not ratified by the interpreter translating both mediator and journalists on TG5 (INM2).

(14) Senator Kerry, you've stated your concern for the rising cost of **health care**, yet you chose a vice presidential candidate who has made millions of dollars successfully suing medical professionals. How do you reconcile this with the voters?

> [Senatore Lei ha sottolineato le preoccupazioni per quanto riguarda la l'aumento dei costi della **sanità** tuttavia abbiamo avuto anche grandi cause nei confronti delle: dei medici che cosa mi dice a riguardo]
> *Senator you underlined your concerns about the increase in **health** costs but we had several lawsuits against doctors what do you have to say about this*
> (TG5, 9.10.2004)

In the passages that immediately follow (14), INK2's HC is not even reiterated by its own producer: INK2 keeps providing new alternatives in the form of contextual synonyms (*sistema sanitario nazionale; servizio sanitario nazionale*). In this case, a tendency to embedded correction (Jefferson 1987: 95) can be safely discarded, given the fact that the HC was indeed correct and perfectly acceptable.

Preliminary results show a clear tendency to variation in the interpreters' output, which results in greater lexical variation in the IT than in the ST. Occurrences of equivalents improving, worsening, or simply providing an acceptable alternative to the HC, respectively, are equally distributed throughout the analysed corpus. Interpreters were, however, expected to apply lexical variation in order to provide improving alternatives to the first occurrence, or simply repeat the latter or provide a contextual synonym, when correct. In other words, their lexical choices were expected to serve the purposes of communication exclusively, i.e. convey the message as clearly and as intelligibly as possible. Once again, such pragmatic function of language subsides in the face of a more specific (special?) and production-oriented purpose. What Straniero Sergio (in press) highlights in relation to DRs may as well be applied to the case of lexical variation: interpreters may indeed be constantly providing synonyms and alternatives for the sole purpose of foregrounding and ratifying their role of text producers.

4. Concluding remarks

At the stage we are now, it may appear that there can exist a special language interpreters use, which may resemble an "acquired

proficiency" (Straniero Sergio, in press), represented by a series of linguistic choices that are the result of application of strategies, respect of norms and reaction to limitations and constraints: in other words, the 'sum' of the individual choices that every interpreter makes and that are the operational result of a professional activity meant for serving a special purpose. On the other hand, some recurring features of the interpreters' output do not find any explanation in any of the above-mentioned aspects or functions; they have a "non-corrective, non-provisional, non-improving and non-explicating value" (Straniero Sergio, in press) and seem to reveal a deeper, perhaps less conscious level of language production, which exists before and beyond the purpose of the interpreting activity and yet can only exist in the output of interpreters at work. As Baker (2000) points out, translation – and, to some extent, interpreting, as well,

> has traditionally been viewed as a derivative rather than creative activity. The implication is that a translator cannot have, indeed should not have, a style of his or her own, the translator's task being simply to reproduce as closely as possible the style of the original. (Baker 2000: 244)

In fact, such claims have been amply disclaimed in the last decade. Interpreters do make a series of linguistic choices that make their output a text in its own right, the interpreted text, rather than a mere *transcodage* of the ST. This set of linguistic choices, the interpreter's thumbprint, are individual, and indicate a single interpreter's style. As shown in the studies illustrated in the present chapter, such a style does not necessarily respond to any communication purpose *per se*. This aspect is particularly evident in the context of television, where interpreters' visibility is much greater than in any other interpreting context (cf. Straniero Sergio 2007) and interpreters become fully-fledged elements of the media event they are interpreting.

How is style relevant to the research question the present chapter attempted to address? More specifically, is it possible to identify recurring features *foregrounding* the language of interpreters on television? Firstly, style in general is personal and idiosyncratic, and secondly, when it comes to interpreters, style really is first and foremost about language (lexical choices, discourse markers, etc.).

Stylistic features are recurrent in a single interpreter's output (Straniero Sergio 2012), over time (Straniero Sergio, in press), across-genres (Trisciuzzi 2010), and regardless of strategic translational choices (Dal Fovo 2012). They may therefore be considered elements that constitute an individual interpreter's very own language, which s/he develops for a special purpose, such as working on television:

> Tv interpreters are being encouraged, through the natural selection process, to enter the media habitus. Broadcasters expect interpreters not just to have the relevant linguistic skills, but also to be good performers. (Katan/Straniero Sergio 2003: 144)

Interpreters are required to make their lexical choices not merely on the basis of what their translational task requires, but also, and more importantly, focusing on the effect that the IT will have on the audience. Each interpreter working on television needs to develop his/her very own style, which results in his/her individual television interpreting language. The next research question that emerges from such conclusion is the following: is the development of a television interpreting style a shared feature of all television interpreters, thus suggesting the existence of a special use of language made by all interpreters working on television, an LSP of television interpreting? The following aspect is currently being investigated, following Straniero Sergio's (2012) suggestion about creating "a sort of identity card of […] individual interpreter[s], including [their] own typical features. To achieve this objective, correlations among different style indicators have to be found." Finding such shared patterns may eventually lead to the creation of an archive of profiles of interpreters working on television, across genres and throughout and may ultimately indicate the existence of an LSP of television interpreting.

References

Anderman, Gunilla M. / Rogers, Margaret (eds) 2008. *Incorporating Corpora. The Linguist and the Translator*. Clevendon: Multilingual Matters.

Baker, Mona 1993. Corpus Linguistics and Translation Studies. Implications and Applications. In Baker, Mona / Francis, Gill / Tognini-Bonelli, Elena (eds) *Text and Technology*. Amsterdam / Philadelphia: John Benjamins, 233-250.

Baker, Mona 1996. Corpus-based Translation Studies: The Challenges that Lie Ahead. In Somers, Harold (ed.) *Terminology, LSP and Translation Studies in Language Engineering: in Honour of Juan C. Sager*. Amsterdam / Philadelphia: John Benjamins, 175-186.

Baker, Mona 2000. Towards a Methodology for Investigating the Style of a Literary Translator. *Target* 12/2, 241-266.

Baker, Mona 2004. A Corpus-based View of Similarity and Difference in Translation. *International Journal of Corpus Linguistics* 9/2, 167-193.

Baumgarten, Nicole / Meyer, Bernd / Özçetin, Demet 2008. Explicitiness in Translation and Interpreting: A Critical View and Some Empirical Evidence (of an Elusive Concept). *Across* 9/2, 177-193.

Beccaria, Gian Luigi 2004. *Dizionario di linguistica e di filologia, metrica, retorica*. Torino: Einaudi.

Berretta, Monica 1982. Problemi testuali della traduzione: casi di ambiguità anaforica in Alice nel paese delle meraviglie. In Calleri, Daniela / Marello, Carla (eds) *Linguistica contrastiva*. Roma: Bulzoni, 229-254.

Bendazzoli, Claudio / Sandrelli, Annalisa 2005. An Approach to Corpus-based Interpreting Studies: Developing EPIC (European Parliament Interpreting Corpus). In Gerzymisch-Arbogast, Heidrun / Nauert, Sandra (eds) *Proceedings of the Marie Curie Euroconferences MuTra: Challenges of Multidimensional Translation - Saarbrücken 2-6 May 2005*.

<http://www.euroconferences.info/proceedings/2005_Proceedings/2005_proceedings.html>.

Bendazzoli, Claudio / Sandrelli, Annalisa 2009. Corpus-based Interpreting Studies: Early Work and Future Prospects. *Tradumatica* 7. *L'aplicació dels corpus linguistics a la traducció*.<http://webs2002.uab.es/tradumatica/revista/num7/articles/08/08art.htm>

Bendazzoli, Claudio / Sandrelli, Annalisa / Russo, Mariachiara 2011. Disfluencies in Simultaneous Interpreting: a Corpus-based Analysis. In Kruger, Alet / Walmach, Kim / Munday, Jeremy (eds) *Corpus-based Translation Studies: Research and Applications.* London / New York: Continuum.

Brennan, Susan E. / Clark, Herbert H. 1996. Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22/6, 1482-1493.

Cencini, Marco / Aston, Guy 2002. Resurrecting the Corp(us/se): Towards an Encoding Standard for Interpreting Data. In Garzone, Giuliana / Viezzi, Maurizio (eds) *Interpreting in the 21st Century*. Amsterdam / Philadelphia: John Benjamins, 47-62.

Clark, Herbert H. / Brennan, Susan E. 1991. Grounding in Communication. In Resnick, Lauren B. / Levine, John M. / Teasley, Stephanie D. (eds) *Perspectives on Socially Shared Cognition*. Washington: APA, 127-148.

Conte, Maria-Elisabeth 1999. *Condizioni di coerenza. Ricerche di linguistica testuale*. Alessandria: Edizioni dell'Orso.

Cortelazzo, Michele 1990. *Lingue speciali. La dimensione verticale*. Padova: Unipress.

Dal Fovo, Eugenia 2012. The More the Merrier? Teamwork and Lexical Variation in Simultaneous Interpretation on Television. In Zybatow, Lew N. / Petrova, Alena / Ustaszewski, Michael (eds) *Translation Studies: Old and New Types of Translation in Theory and Practice*. Frankfurt am Main: Peter Lang, 307-314.

Falbo, Caterina 2009. Un grand corpus d'interprétation: à la recherche d'une stratégie de classification. In Paissa, Paolo / Biagini, Marta (eds.) *Doctorants & Recherche 08. La recherche actuelle en linguistique française*. Milano: Lampi di stampa, 105-120.

Falbo, Caterina 2012. CorIT (Television Interpreting Corpus): Classification Criteria. In Straniero Sergio, Francesco / Falbo, Caterina (eds) *Breaking ground in corpus-based interpreting studies*. Bern: Peter Lang, 155-186.

Gile, Daniel 1995. Fidelity Assessment in Consecutive Interpretation: an Experiment. *Target* 7/1, 151-164.

Gotti, Maurizio 1991. *I linguaggi specialistici: caratteristiche linguistiche e criteri pragmatici*. Scandicci: La nuova Italia.

Gotti, Maurizio 2003. *Specialized Discourse: Linguistic Features and Changing Conventions*. Bern: Peter Lang.

Granger, Sylviane 2003. The Corpus Approach: A Common Way forward for Contrastive Linguistics and Translation Studies? In Granger, Sylviane / Lerot, Jacques / Petch-Tyson, Stephanie (eds) *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam / New York: Rodopi, 17-29.

Halliday, Michael A.K. 1978. *Language as Social Semiotic. The Social Interpretation of Language and Meaning.* London: Edward Arnold.

Hartley, John 1992. *The Politics of Pictures: the Creation of the Public in the Age of Popular Media*. London / New York: Routledge.

Katan, David / Straniero Sergio, Francesco 2001. Look Who's Talking: The Ethics of Entertainment and Talkshow Interpreting. *The Translator* 7, 213-237.

Katan, David / Straniero Sergio, Francesco 2003. Submerged Ideologies in Media interpreting. In Calzada Pérez, Maria *Apropos of Ideology*. Manchester: St.Jerome, 131-144.

Kenny, Dorothy 2001. *Lexis and Creativity in Translation. A Corpus-based Study*. Manchester: St. Jerome.

Jefferson, Gail 1987. On Exposed and Embedded Correction in Conversation. In Button, Graham / Lee, John R.E. (eds) *Talk and Social Organization*. Clevedon: Multilingual Matters, 86-100.

Laviosa, Sara 1998. The Corpus-based Approach: A New Paradigm in Translation Studies. *Meta* 43/4, 474-479.

Laviosa, Sara 2002. *Corpus-based Translation Studies. Theory, Findings, Applications*. Amsterdam / New York: Rodopi.

Levelt, Willem J.M. / Kelter, Stephanie 1982. Surface Form and Memory in Question Answering. *Cognitive Psychology* 14, 78-106.

Martin, Philippe 2005. Linguistique de l'oralité: description de la prosodie et analyse instrumentale. *The Interpreters' Newsletter* 13, 15-24.

Martin, Philippe 2009. *Intonation du français*. Paris: Armand Colin.

Marzocchi, Carlo 2005. On Norms and Ethics in the Discourse on Interpreting. *The Interpreters' Newsletter* 13, 87-107.

Olohan, Maeve 2004. *Introducing Corpora in Translation Studies*. London / New York: Routledge.

Papa, Carlo 2010. *Il fenomeno del self-repair nell'interpretazione simultanea in televisione [Self-repair in TV simultaneous interpreting]*. Unpublished master dissertation, University of Trieste,

Petite, Christelle 2005. Evidence of Repair Mechanisms in Simultaneous Interpreting: A Corpus-based Analysis. *Interpreting* 7, 27-49.

Pinker, Steven A. 1994. *The Language Instinct*. London: Penguin Books.

Russo, Mariachiara / Bendazzoli, Claudio / Sandrelli, Annalisa 2006. Looking for Lexical Patterns in a Trilingual Corpus of Source and Interpreted Speeches: Extended Analysis of EPIC (European Parliament Interpreting Corpus). *Forum* 4/1, 221-254.

Sandrelli, Annalisa / Bendazzoli, Claudio 2005. Lexical Patterns in Simultaneous Interpreting: a Preliminary Investigation of EPIC (European Parliament Interpreting Corpus). In *Proceedings from the Corpus Linguistics Conference Series*, 1:1, ISSN 1747-9398, online: <http://www.corpus.bham.ac.uk/PCLC/>.

Setton, Robin 2011. Corpus-based Interpretation Studies (CIS): Reflections and Prospects. In Kruger, Alet / Walmach, Kim / Munday, Jeremy (eds) *Corpus-based Translation Studies: Research and Applications*. London / New York: Continuum, 33-75.

Shlesinger, Miriam 1998. Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies. *Meta* 43/4, 486-493.

Simone, Raffaele 2001. *Fondamenti di linguistica*. Bari: Editori Laterza.

Straniero Sergio, Francesco 2003. Norms and Quality in Media Interpreting: the Case of Formula One Press Conferences. *The Interpreters' Newsletter* 12, 135-74.

Straniero Sergio, Francesco 2007. *Talkshow Interpreting: la mediazione linguistica nella conversazione spettacolo*. Trieste: EUT.

Straniero Sergio, Francesco 2012. Using Corpus Evidence to Discover Style in Interpreters' Performances. In Straniero Sergio, Francesco / Falbo, Caterina (eds) *Breaking Ground in Corpus-based Interpreting Studies*. Bern: Peter Lang, 211-230.

Straniero Sergio, Francesco, in press. Double Renditions in Simultaneous Interpreters' Output. A Corpus-based Research.

Straniero Sergio, Francesco / Falbo, Caterina (eds) 2012. *Breaking Ground in Corpus-based Interpreting Studies*. Bern: Peter Lang.

Tanskanen, Sanna Kaisa 2004. Cohesion and Collaboration. In Aijmer, Karin / Stenström, Anna-Brita (eds) *Discourse Patterns in Spoken and Written Corpora*. Amsterdam / Philadelphia: John Benjamins, 89-110.

Tissi, Benedetta 2000. Silent Pauses and Disfluencies in Simultaneous Interpretation: A Descriptive Analysis. *The Interpreters' Newsletter* 10, 103-127.

Trisciuzzi, Marco 2010. *L'interprete televisivo: uno stile [Television Interpreter: a style]*. Unpublished master dissertation, University of Trieste.

Van Besien, Fred / Meuleman, Chis 2008. Style Differences among Simultaneous Interpreters: A Pilot Study. *The Translator* 14/1, 135-155.

Vanelli, Laura 1981. Il meccanismo deittico e la deissi nel discorso. *Studi di grammatica italiana* 10, 293-311.

Wadensjö, Cecilia 1998. *Interpreting as Interaction*. London / New York: Longman.

Yagi, Sane M. 2000. Studying Style in Simultaneous Interpretation. *Meta* 45/3, 520-547.

CÉCILE DESOUTTER

16. La prise en compte linguistique des femmes dans les discours électoraux : une étude sur corpus.

1. Introduction

L'analyse de corpus permet de mettre en évidence la variété des usages dans les limites permises par la culture et par la langue. C'est précisément dans cette optique que nous abordons ici la prise en compte linguistique des femmes à travers l'observation de deux corpus de discours électoraux français ayant trait aux élections présidentielles de 2012 et prud'homales de 2008. Après avoir rappelé les diverses positions qui s'opposent dans les pays francophones sur la question du genre grammatical, nous préciserons les principes de l'écriture non sexiste tels qu'ils sont formulés dans les guides de rédaction publiés en Suisse et au Canada. Nous nous attacherons ensuite à relever, dans nos corpus recueillis en France, les contextes discursifs et pragmatiques ainsi que les domaines référentiels qui tendent à favoriser la visibilité linguistique des femmes. Nous observerons parallèlement les procédés utilisés dans les discours analysés pour éviter le masculin générique et considérerons les conséquences au niveau de la cohérence et de la cohésion du texte.

2. A l'origine du débat sur la féminisation : genre grammatical *vs* genre biologique

La catégorie du genre grammatical repose sur la division des noms en classes, dites classes nominales. En français, comme dans d'autres langues romanes dotées de deux genres (masculin et féminin), il y a lieu de faire une distinction entre la sous-classe des référents non animés et celles des référents animés. La première sous-classe possède un genre arbitraire, fruit de l'étymologie et de l'histoire: on dit *une chaussure* et *un soulier*, *le soleil* et *la lune*, sans que ceci n'ait rien à voir avec un éventuel "sexe" des mots. En revanche, pour ce qui concerne la sous-classe des noms animés, et plus en particulier celle des êtres humains, le genre grammatical correspond le plus souvent au sexe biologique. Comme le font remarquer la plupart des grammaires, les exceptions à cette motivation sont peu nombreuses. Riegel *et al.* (2006 : 172), par exemple, les regroupent en quatre séries: quelques noms féminins désignent des fonctions habituellement exercées par des hommes (*recrue, sentinelle*…), un nombre également restreint de noms masculins (*alto, bas-bleu, laideron, mannequin* …) s'appliquent à des femmes, des noms masculins (*auteur, modèle, otage*...) et des noms féminins (*connaissance, personne, vedette*…) désignent indifféremment des personnes des deux sexes.

2.1. Valeur spécifique vs *valeur générique*

Hors les quelques exceptions indiquées précédemment, le féminin est toujours particularisant et exclusif : il désigne uniquement un individu de sexe féminin. En revanche, si le masculin est également particularisant, désignant un individu de sexe masculin, c'est par ailleurs un genre "indifférencié" (Wagner / Pinchon 1962 : 56), "commun, neutre ou asexué" (Grévisse 1986 : 799), c'est-à-dire inclusif et généralisant. A cet égard, il peut indiquer aussi bien des êtres masculins que féminins.

Cette valeur générique correspond, d'un côté, à un emploi au singulier, en compréhension, et, de l'autre, au pluriel, en extension. En compréhension, le masculin grammatical permet de désigner, par exemple, l'espèce entière sans distinction de sexe (*l'homme est un loup pour l'homme*) ou de faire référence à une fonction (*l'assistance d'un avocat est facultative*) et non à une personne en particulier (*je suis venue avec mon avocate*). En extension, il permet de référer à la fois à l'homme et à la femme : dans l'énoncé *les candidats sont priés d'attendre leur tour*, les candidats peuvent être uniquement de sexe masculin ou des deux sexes. Par le phénomène de l'accord, la règle scolaire du masculin qui l'emporte sur le féminin s'étend à l'adjectif épithète ou attribut ainsi qu'au pronom représentant le nom. Par conséquent, "si les donneurs d'accord ne sont pas du même genre, le receveur se met au genre indifférencié, c'est-à-dire au masculin" (Grévisse 1986 : 720). On dit donc : *les hommes et les femmes sont égaux.*

Le double statut du masculin grammatical en français, tour à tour spécifique ou générique, peut provoquer des ambiguïtés. Avant tout, il y a risque d'assimilation de la forme particularisante et de la forme généralisante. Ainsi, si l'on vous dit que *les Français sont forts en handball*, vous pouvez légitimement vous demander s'il est uniquement question d'équipes masculines ou si les équipes féminines sont également concernées. Certes, le contexte permet en général de dissiper assez rapidement le doute. Il n'en demeure cependant pas moins que le masculin générique tend à occulter les femmes, ce qui est loin d'être indifférent du point de vue sociologique. D'ailleurs, cette vision linguistiquement masculine du monde n'a pas toujours été de mise. Que l'on parle du nom des professions au masculin ou de la règle de l'accord lorsque les référents ont des genres différents, le principe du masculin indifférencié en français est historiquement daté. Sous l'Ancien régime, les termes par lesquels les femmes étaient désignées (titre ou profession) étaient toujours au féminin. On rencontrait ainsi des baronnesses, des commanderesses, des chirurgiennes, des cabaretières… (Diester / Moreau 2009 : 13). De même, Grévisse fait-il remarquer que les auteurs du XVII^{e} pratiquaient assez souvent l'accord de l'adjectif non pas avec le nom masculin mais avec le nom le plus proche, suivant en cela l'usage

dominant en ancien français. Et si “Malherbe critiquait cette façon de faire […] Vaugelas l’admettait” (Grévisse 1986 : 721). Ce n’est qu’au XVIII[e] siècle que le masculin générique s’impose pour des raisons plus idéologico-culturelles que linguistiques. “Le plus fort l’emporte pour dire que les plus puissants ont toujours l’avantage. Masculin qualifie le plus noble des genres, celui qui a ordinairement l’avantage sur l’autre”, affirme le lexicographe Antoine Furetière (Fracchiola 2008 : 8) et le grammairien Nicolas Beauzée de renchérir tout simplement que “le genre masculin est réputé plus noble que le féminin à cause de la supériorité du mâle sur la femelle” (Fracchiola 2008 : 8). De cette époque date donc ce que l’on pourrait appeler la “masculinisation” linguistique.

2.2. Des formulations attentives au genre biologique

Depuis la deuxième moitié du XX[e] siècle, le souci de rendre les femmes linguistiquement visibles, pour tenir compte de leurs place et rôle effectifs dans la société, est repérable dans deux domaines bien distincts : celui du lexique et celui du discours.

2.2.1. Dimension lexicale

La dimension lexicale concerne avant tout les noms de titre, de fonction ou de métier qui désignent une personne. Il s’agit de rendre le genre congruent avec le sexe de cette personne. On pourra ainsi dire ***la*** *ministre Cécile Duflot, l’écrivain****e*** *Benoite Groult*. Les personnes favorables à cette mise au féminin la justifient par des raisons sociologiques : la place grandissante de la femme dans la société et l’importance d’être nommée pour avoir une identité (Yaguello 1978) mais aussi par des arguments plus pragmatiques de clarté du discours et de bonne interprétation de l’information (Matthey 2000). En revanche, comme le font remarquer Dister / Moreau (2009), ceux ou celles qui s’opposent à la féminisation de ces désignateurs, préférant les formes génériques *le ministre Cécile Duflot, l’écrivain Benoite Groult,* évoquent plutôt des considérations linguistiques. Leur point de vue repose essentiellement sur ce que nous avons précisé

précédemment – et sur quoi a tranché l'Académie française à diverses reprises (en 1984 et en 2002) – à savoir que le genre dit couramment "masculin" est le genre non marqué, qui peut représenter des éléments aussi bien masculins que féminins. Par ailleurs, les adversaires de la féminisation linguistique estiment que, en contexte, les noms de profession ou de titre désignent le plus souvent la fonction et non la personne.

2.2.2. Dimension discursive

La dimension discursive concerne l'évitement du masculin générique et, au-delà du lexique, elle a trait à l'emploi d'une syntaxe et d'une grammaire tenant compte de la parité linguistique. Différentes appellations sont utilisées pour désigner l'approche rédactionnelle qui vise à mieux manifester la présence des femmes dans la langue et nous aurons recours aux unes et aux autres. La revue *Nouvelles Questions féministes* (Baider *et al* 2007) emploie "parité linguistique", Perry (2004) parle de "désexisation", Abbou (2011) d'"antisexisme linguistique", tandis que les administrations suisses et canadiennes optent pour "formulation non-sexiste" (Chancellerie fédérale suisse) ou plus souvent pour "langage épicène" ou "rédaction épicène" (Etat de Genève, Office québécois de la langue française). "Epicène" se dit "des noms communs aux deux sexes, tels que *enfant*, *aigle*, *caille*, etc." (Petit Larousse). Toutefois, quand on parle d'un texte épicène, on se réfère à "un texte qui met en évidence de façon équitable la présence des femmes et des hommes" (Vachon-L'Heureux / Guénette 2007 : 17). La rédaction épicène repose sur deux procédés rédactionnels spécifiques : d'une part la particularisation (on parle aussi souvent de féminisation[1]) et d'autre part la neutralisation.

Dans le premier cas, on particularise le discours en faisant apparaitre les oppositions entre forme marquée masculine et forme

1 Personnellement, nous préférons le terme le terme de "particularisation" emprunté à Abbou (2011). D'une part, parce qu'il peut s'agir – quoique rarement – d'ajouter un élément de masculinisation à un terme féminin, d'autre part parce que le but pourrait être, "moins de faire apparaître systématiquement les femmes, que de pouvoir mettre simultanément les genres côte à côte" (Abbou 2011 : 113).

marquée féminine au moyen d'une formulation bi-genrée. Ce "double-marquage" (Abbou 2011) implique de recourir à des doublets intégraux (*les collaborateurs et collaboratrices, ceux et celles...*) ou abrégés par le biais d'artifices typographiques divers : parenthèses (*les client(e)s*), tirets (*les clients-es*), barres obliques (*les client / e / s*).

Dans le second cas, on neutralise le discours, en faisant d'une certaine façon disparaitre le genre, par le recours à des noms génériques (*les personnes*) ou collectifs (*le personnel*), à des formes non personnalisées, à des formes passives, à des ellipses…

Le premier procédé a pour effet d'inclure les genres masculin et féminin de manière explicite, le second de manière implicite. Nous verrons plus loin que la rédaction épicène consiste à combiner les deux procédés de la meilleure manière possible.

2.3. Les guides de rédaction dans les pays francophones

A partir des années 1970, suite à la poussée des mouvements féministes en particulier en Amérique du Nord, des institutions francophones se sont progressivement souciées de trouver des solutions permettant de donner une plus grande visibilité linguistique aux femmes. Cette préoccupation s'est en particulier traduite par la publication de guides de féminisation dans un premier temps et de formulation non sexiste dans un deuxième temps.

Pays / Région	**Date de publi-cation**	**Titre des guides**	**Editeur**
Québec	1979	*Féminisation des titres – Affichage des postes*	OQLF[2]
	1986	*Titres et fonctions au féminin. Essai d'orientation de l'usage*	OQLF
	1991	*Au féminin – Guide de féminisation*	OQLF

2 Office québécois de la langue française

		des titres de fonction et des textes	
	2007	*Avoir bon genre à l'écrit – Guide de rédaction épicène*	OQLF
Suisse	1986	*Formulation non sexiste des actes législatifs et administratifs*	Chancellerie fédérale
	1991 (mis à jour en 1999)	*Dictionnaire féminin-masculin des professions, des titres et des fonctions*	Cantons de Genève et du Jura
	2000	*Guide de formulation non sexiste des textes administratifs et législatifs de la Confédération*	Chancellerie fédérale
	2002	*Ecrire les genres – Guide romand d'aide à la rédaction administrative et législative épicène*	État de Genève, DF-SPPÉgalité
Belgique	1994 (mis à jour en 2005)	*Mettre au féminin – Guide de féminisation des noms de métier, fonction, grade ou titres*	Conseil sup. de la langue française de la Communauté française de Belgique
Luxem-bourg	1998 (mis à jour en 2002 puis en 2006)	*Féminin-masculin, Métiers, Titres et Fonctions*	Ministère de l'Égalité des chances
France	1999	*Femme j'écris ton nom – Guide d'aide à la féminisation des noms de métiers, titres, grades et fonctions*	CNRS et INaLF[3]

Tableau 1. Quelques guides de féminisation et de rédaction non sexiste

3 Centre national de la recherche scientifique et Institut national de la langue française

2.3.1. Les guides de féminisation des noms de titre et de profession

Le Canada a lancé l'initiative dès 1979 avec un premier *Avis de recommandation sur la féminisation des titres et l'affichage des postes* de l'Office québécois de la langue française (OQLF). A sa suite, dans les pays européens francophones, les pouvoirs publics ont émis plus ou moins rapidement des réglementations ou avis allant dans ce sens et les institutions chargées de la gestion de la langue ont publié des guides de féminisation afin d'orienter l'usage (Suisse 1986, Belgique 1994, Luxembourg 1998, France 1999)[4]. Ces guides se composent le plus souvent d'une présentation des règles de mise au féminin et d'un répertoire des solutions préconisées, réhabilitant de la sorte certaines formes féminines inusitées et en développant de nouvelles. Le fait que les guides aient été rédigés à des moments distincts et que, pour quelques mots, ils puissent diverger légèrement d'un pays à l'autre, traduit les sensibilités différentes de ces derniers en matière de féminisation. Dans la plupart des cas, les brochures, aisément accessibles en ligne, ont fait l'objet de mises à jour à quelques années de leur publication (Cf. Tab. 1) de façon à mieux coller à l'usage qui ne cesse d'évoluer.

Comme le font remarquer Mougeon *et al.*, réhabiliter ou créer des substantifs féminisés là où ils n'existaient pas est un changement qui affecte très peu la langue française, "puisque les règles grammaticales du français se prêtent bien à cette gymnastique lexicale" (Mougeon *et al.* en ligne : 1). C'est sans doute ce qui explique que le recours à un lexique féminisé pour désigner les femmes dans leur activité soit aujourd'hui presque généralisé dans certains pays (ex. Canada) et partiellement admis dans d'autres (ex. France[5]). En revanche, le passage de la féminisation de certains mots à

4 Nous nous limitons ici à citer les guides principaux publiés dans 5 pays francophones mais les institutions internationales (ONU, Conseil de l'Europe, UNESCO, UE, Organisation internationale de la Francophonie) ont également publié leur propres guides ou recommandations.

5 La Commission générale de terminologie et de néologie dans son *Rapport sur la féminisation des noms de métier, fonction, grade ou titre* (1998) ainsi que le Premier Ministre Jospin, dans sa préface au Guide *Femme j'écris ton nom*

la rédaction non sexiste représente un pas plus difficile à franchir. Comme on peut l'observer dans le tableau 1, ce n'est qu'au Québec puis en Suisse que la réflexion sur la parité linguistique a mené, au-delà de la publication de listes de désignateurs au féminin, à la production de guides institutionnels pour promouvoir et rendre praticable une rédaction épicène.

2.3.2. Les guides de rédaction épicène

La rédaction non sexiste, telle qu'elle est appréhendée aujourd'hui dans ces pays, ne consiste pas à féminiser systématiquement des textes écrits au masculin. A la différence des premiers opuscules des années 1980 "qui ne se penchaient que sur la féminisation de chaque phrase indépendamment des autres, tout en offrant peu de stratégies pour le faire" (Mougeon *et al* en ligne : 1), les guides québécois et suisses du début des années 2000, consultables en ligne, proposent des solutions "pour penser son texte de façon épicène et l'écrire de sorte qu'il s'adresse d'emblée aux femmes et aux hommes" (Vachon-L'Heureux / Guénette 2007 : 17).

Avant de fournir des solutions pratiques illustrées d'exemples, les guides de rédaction épicène insistent sur le précepte de base qui doit rester celui de la clarté du message et la cohérence de l'écriture. Ainsi, l'OQLF liste des principes généraux qui servent de balises à une rédaction de qualité:

- Abandonner la mise au masculin habituelle du texte
- Penser épicène et rédiger épicène
- Préserver la lisibilité du texte
- Assurer l'intelligibilité du texte
- Veiller à une juste répartition des formes féminines
- Évaluer la pertinence du recours aux marques du genre
- Adapter la rédaction épicène à chaque type de textes

(CNRS / INLF 1999), distinguent les noms de métiers, qui sont personnels et doivent donc être accordés en genre avec la personne qui exerce le métier donné, et les noms de fonctions, qui sont neutres et ne peuvent donc être employés qu'au masculin générique.

Compte tenu de ces recommandations générales partagées par les institutions suisses et canadiennes, chaque guide de rédaction évoque plus ou moins les mêmes techniques. Par exemple, le *Guide de formulation non sexiste des textes législatifs et administratifs de la Confédération* (2000), publié par la Chancellerie fédérale suisse, propose une douzaine de possibilités en recommandant d'éviter de recourir à la dernière d'entre elles (qui consiste à conserver le masculin générique). Il est ainsi conseillé d'utiliser par ordre décroissant :

- des termes épicènes (*tout être humain* plutôt que *tout homme ; apte* plutôt que *qualifié*)
- des doublets intégraux (*Chère lectrice, cher lecteur ; les citoyens et citoyennes…*)
- des articles féminisés, dédoublés ou supprimés (*le juge ou la juge*; *le / la juge* plutôt que *le juge)*
- des singuliers collectifs (*équipe, groupe, personnel, service, corps*…)
- le pluriel dans le cas de termes épicènes (*les propriétaires* plutôt que *le propriétaire*)
- des formes non personnalisées (*la Présidence* plutôt que *le Président* ; *le tribunal* plutôt que *le juge ; les secours* plutôt que *les sauveteurs*)
- la forme passive sans complément d'agent
- l'ellipse
- la personnalisation (rédaction d'une version d'un document au masculin et d'une autre au féminin)
- la barre oblique à l'intérieur des mots (*signature du / de la titulaire*)
- une explication à l'intérieur du texte ("*pour faciliter la lecture du document, le masculin générique est utilisé pour désigner les deux sexes*")
- le masculin générique.

A cette liste, le guide roman d'aide à la rédaction administrative épicène *Ecrire les genres* (2001) ajoute les solutions d'utiliser

l'infinitif ou de s'adresser le plus possible directement au destinataire (*vous* au lieu de *il* ou *elle*).

2.3.3. La rédaction non sexiste : une forme de variation diatopique

Les conseils fournis par les guides s'appliquent à la rédaction de documents et donc à l'écrit. Ainsi, alors que la variation diatopique du français concerne essentiellement l'oral, la présence de marques de féminisation explicites (les doublets intégraux ou abrégés) est une des rares traces de variation à l'écrit qui peut laisser supposer qu'un texte les contenant provient du Canada ou de Suisse plutôt que de France ou de Belgique. Dans les pays qui la promeuvent, les principes de la rédaction épicène doivent avant tout être suivis par les administrations et ils s'imposent à des degrés très divers. La pratique effective reste par conséquent fluctuante. Thérèse Moreau signale qu'en Suisse "si la féminisation lexicale est bien comprise et utilisée partout, l'école reste fermée à ce qui est la parité syntaxique" (Dumais 2009 : 178). De même, Pierrette Vachon-L'Heureux fait-elle remarquer qu'au Québec "la féminisation des titres et des fonctions est acquise ; la féminisation des textes est entreprise, mais doit être poursuivie, raffinée, travaillée" (Dumais 2009 : 174). A cet égard, elle signale une absence de consensus quant au type de rédaction acceptable et remarque que l'usage est balloté entre le tout et le rien. Le tout, c'est, selon elle, la féminisation de chaque substantif, adjectif ou pronom susceptible de l'être et qui rend les textes longs et la lecture lourde. Le rien, c'est la note explicative placée en début du texte pour excuser l'emploi du masculin générique. A ce tout ou rien, la linguiste canadienne oppose donc la rédaction épicène, laquelle exige un effort de la part de ceux et celles qui rédigent et devrait permettre de parvenir à des résultats acceptables pour tous.

Il apparait ainsi que la prise en compte linguistique des femmes n'entraine pas nécessairement pas une féminisation tous azimuts. De fait, dans la mesure où l'expression de l'opposition masculin-féminin n'est pas "adaptée à n'importe quel contexte cognitif ou argumentatif" (Béguelin / De Pietro 2000), l'exigence de rendre femmes et hommes linguistiquement visibles ne s'applique pas indifféremment à tous les genres de textes. Ainsi, dans le contexte de la Suisse romande,

Béguelin (1999 : 45) considère que la “rédaction ‘politiquement correcte’ ne concerne que certains genres bien précis, textes officiels, discours politiques, tracts électoraux, offres d’emploi…”. Vacheron-L’Heureux (en ligne 2004) retient pour sa part que “la féminisation des textes est acceptable et acceptée lorsqu’elle porte sur certains types de documents. Par exemple la carte professionnelle, la lettre personnalisée, le formulaire, le contrat…”. La linguiste canadienne précise aussi que la féminisation de documents qui ne sont pas destinés à la lecture mais plutôt à la consultation est généralement acceptable et acceptée. Dans le même esprit, la brochure sur *La féminisation au Gouvernement de l’Ontario* (2003) précise, dans ses recommandations, qu’en matière de féminisation chaque ministère, chaque organisme, chaque société émet ses propres consignes, qui peuvent varier selon la nature du texte et qu’il importe de se faire confirmer ces consignes. En somme, pour reprendre un titre de Thérèse Moreau (2007), le mot d’ordre serait “Prière de ne pas épicéner partout”.

En France, aucune volonté politique ou aménagement linguistique ne pousse à la prise en compte linguistique des femmes. Tout au contraire, comme nous l’avons signalé plus haut, le purisme académique conseille fortement de s’en tenir au genre dit “non marqué” qu’est le masculin générique. Cela dit, dans la pratique, il existe quelques exceptions. Ainsi, on relève un effort de parité linguistique dans les annonces d’emploi dont le code du travail (Art. L1142-1) exige qu’elles ne soient pas sexuellement discriminatoires. De même, la présence d’une candidate aux élections présidentielles de 2007 avait-elle permis une introduction timide de la parité linguistique avec des formulations bi-genrées.

Puisque la France se distingue par une propension très faible − mais néanmoins non nulle − à l’écriture épicène, nous avons cherché des traces de cette pratique dans des genres de discours pour lesquels elle serait conseillée ou adoptée en Suisse romande ou au Canada. Il s’agit de deux genres relativement stabilisés appartenant au domaine politique ou professionnel : les déclarations (ou professions de foi) des candidats et candidates aux élections présidentielles et les tracts syndicaux incitant les personnes salariées ou demandeuses d’emploi à voter lors des élections prud’homales. A défaut de pouvoir nous

référer à des consignes françaises pour analyser les choix de particularisation ou de neutralisation des genres grammaticaux dans ces écrits, nous prendrons pour repères les guides suisses et québécois cités plus haut.

3. Corpus et méthodologie

3.1. Recueil des corpus

3.1.1. Corpus Elections présidentielles

En 2012, toutes les personnes de nationalité française inscrites sur les listes électorales ont reçu par voie postale la profession de foi des 7 candidats et 3 candidates à la Présidence de la République : soit 10 feuilles doubles imprimées en couleurs. Par ailleurs, les candidats et candidates ont pu accéder aux adresses électroniques de l'électorat résidant à l'étranger et contacter par courriel les personnes inscrites sur les listes consulaires. Pour constituer notre corpus, nous avons rassemblé les divers documents qui nous ont été personnellement adressés à l'occasion de ces élections en tant que Française résidant en Italie. Il s'agit donc d'abord des 10 documents papier représentant 40 pages illustrées, auxquels s'ajoutent 7 courriels (3 candidats n'ayant pas opté pour cette forme de communication)[6]. Pour l'essentiel, ces courriels sont composés d'un message directement adressé à l'électorat, se présentant sous forme épistolaire avec d'éventuels liens à un programme plus détaillé. Nous n'avons tenu compte que du message.

---

6 Lorsque les candidats ou candidates ont adressé plusieurs courriers ou courriels, nous n'avons tenu compte que du premier reçu, après vérification que les suivants ne présentaient pas de différences notables du point de vue de l'évitement du masculin générique.

3.1.2. Corpus Elections prud'homales

Les conseils de prud'hommes sont des tribunaux spécialisés dans la résolution des conflits individuels du travail entre employeurs et salariés. Ils sont composés paritairement de juges élus ou élues pour 5 ans[7] et appartenant au monde économique. Ce sont en général les syndicats qui présentent des candidats et candidates. Nous nous intéressons ici aux tracts rédigés par les syndicats à l'intention de l'électorat salarié ou demandeur d'emploi à l'occasion des élections prud'homales du 3 décembre 2008. Pour constituer le corpus, nous avons consulté les sites internet des 5 confédérations syndicales dites représentatives : CGT, CFDT, CGT-FO, CFTC, CFE-CGC[8] et de trois syndicats dits minoritaires : UNSA, Solidaires, CNT[9]. Nous avons ainsi recueilli 47 tracts en format Word ou PDF.

3.2. Caractéristiques des corpus

Nos deux corpus présentent des caractéristiques distinctes que nous présentons ci-dessous de façon succincte (Tab. 2) :

	Elections présidentielles	**Elections prud'homales**
Genre de discours	Profession de foi (déclaration) électorale	Tract syndical
Emetteurs	Candidats ou candidates aux élections présidentielles	Syndicats

7 Les élections objet de la présente étude datent de 2008, les suivantes qui auraient dû se dérouler en 2013 ont été reportées au plus tôt à 2015 pour éviter de multiplier les scrutins sur une même période (Article 7, loi n° 2010-1215 du 15 octobre 2010).

8 Confédération générale du travail, Confédération française démocratique du travail, Force ouvrière, Confédération française des travailleurs chrétiens, Confédération française de l'encadrement–Confédération générale des cadres.

9 Union nationale des syndicats autonomes, Solidaires (auquel est fédéré le syndicat SUD), Confédération nationale du travail.

Destinataires	Personnes inscrites sur les listes électorales pour les élections présidentielles	Personnes salariées ou demandeuses d'emploi inscrites sur les listes électorales prud'homales
Ouverture / fermeture du corpus	Fermé	Semi-fermé
Période	2012	2008

Tableau 2. Caractéristiques des deux corpus

Toutefois, au-delà de ces différences, ils possèdent de nombreux aspects en commun (Tab. 3) :

Canal	Ecrit
Support	Papier + électronique
Domaine	Discours électoral
Cadre	Institutionnel
Type de discours	Argumentatif
But illocutoire	Convaincre
Registre	Formel
Diffusion	Très large

Tableau 3. Points communs aux corpus

3.3. Finalité de l'étude

Compte tenu du faible intérêt manifesté en France pour la parité linguistique telle que nous l'avons décrite précédemment, il s'agit d'en repérer les traces, même minimales, dans les discours électoraux objet de l'étude. Sur le plan linguistique, nous nous interrogeons sur la façon dont est évité le masculin générique : quels sont les procédés les plus utilisés et leurs effets sur la cohésion et cohérence du texte ? Sur le plan sociolinguistique, nous cherchons à identifier les supports, contextes, orientations politiques favorisant l'écriture épicène.

4. Résultats

4.1. Le corpus Elections présidentielles

4.1.1. Les formules d'ouverture (corpus papier et électronique)

Les déclarations électorales sur support papier ou électronique commencent le plus souvent par une formule d'appel. D'un point de vue pragmatique, celle-ci a pour effet de ménager la personne qui (é)lit et de solliciter sa bienveillance. Elle constitue également le premier lieu où il est possible au candidat ou à la candidate de montrer une éventuelle sensibilité à l'adresse de son électorat féminin. Notons à cet égard que traditionnellement – mais contrairement à la conception largement partagée que le masculin est un genre non marqué – les discours officiels des présidents de la V^{e} République commencent par la formule *Françaises, Français*.
Le tableau ci-dessous indique les formules d'adresse utilisées (Tab. 4) :

Candidat Candidate	**Sexe**	**Orienta-tion politique**	**Déclaration papier**	**Déclaration électronique**
Artaud	♀	Gauche	*Travailleurs, travailleuses*	Pas de courriel
Bayrou	♂	Centre	*Madame, Monsieur, chers concitoyens*	*Chère Madame, Cher Monsieur*
Dupont-Aignan	♂	Droite	*Mes chers compatriotes*	*Cher(e)s compatriotes*
Cheminade	♂	Droite	*Ø*	Pas de courriel
Hollande	♂	Gauche	*Madame, Monsieur*	*Cher-e compatriote,*
Joly	♀	Vert	*Citoyennes, citoyens; Madame, Monsieur*	*Chères concitoyennes, Chers concitoyens***,**
Le Pen	♀	Droite	*Mes chers compatriotes*	*Françaises et Français de l'étranger, mes chers compatriotes,*
Mélanchon	♂	Gauche	*Ø*	*Chère concitoyenne, cher concitoyen*

Poutou	♂	Gauche	Ø	Pas de courriel
Sarkozy	♂	Droite	*Mes chers compatriotes*	*Mes chers amis*

Tableau 4. Les formules d'appel des déclarations papier et électroniques

Il ressort de ce relevé que les termes d'adresse sont diversifiés. Dans les déclarations papier, sur 7 formules d'appel, 4 sont bi-genrées et 3 se limitent au masculin générique. Ces trois dernières recourent au même terme *compatriotes,* qui est certes épicène mais ici précédé de l'adjectif *chers* au masculin. En revanche sur 7 formules d'appel des déclarations électroniques, 1 seule conserve le masculin générique.

A cette exception près, et tout en tenant compte de l'exiguïté du corpus, il apparait donc que le support électronique favorise les formules d'appel doublement marquées. De fait, Nicolas Dupont-Aignan et Marine Le Pen passent du masculin générique (Mes chers compatriotes) dans leur profession de foi papier à une formulation avec double-marquage dans la version électronique : *Cher(e)s compatriotes* (Dupont-Aignan); *Françaises et Français* (Le Pen). De même relève-t-on dans le corpus électronique la présence de formes abrégées qui ne figurent pas dans le corpus papier : F. Hollande passe ainsi de *Madame, Monsieur* à *Cher(e) compatriote* . Dans ce tableau, N. Sarkozy apparait comme le seul réfractaire au double-marquage. En utilisant le terme d'adresse *chers amis* au lieu de *chers compatriotes,* il précise dans son écrit électronique la nature amicale de la relation mais ne rend pas pour autant les femmes plus visibles.

4.1.2. Les candidates et la parité linguistique (corpus électronique)

Pour observer si les femmes tendent à produire une écriture plus épicène que les hommes, nous avons comparé les courriels envoyés par Marine Le Pen et par Eva Joly à l'électorat résidant à l'étranger (Tab. 5). A cet effet, nous avons compté[10] dans leur texte électronique

10 Pour les besoins de l'analyse, les données quantitatives ont été extraites en partie manuellement et en partie avec le logiciel AntConc 3.2.4w.

les termes d'adresse ou désignateurs de référents humains susceptibles de donner lieu à une réflexion sur le choix du genre :

Candidate	**Nombre de mots du courriel**	**Termes d'adresse ou de désignateurs en nombre**	**Termes d'adresse ou de désignateurs en pourcentage**
Eva Joly	397	5	1,25 %
Marine Le Pen	605	10	1,65 %

Tableau 5. Rapport nombre de désignateurs ou termes d'adresse / nombre de mots

Il est intéressant de remarquer que dans les deux textes, le nombre de mots susceptibles de particularisation ou de neutralisation est extrêmement réduit (moins de 2 %), tout comme il l'est d'ailleurs dans l'ensemble des déclarations papier ou électroniques. Voyons à présent quels sont ces mots et le traitement qui leur est appliqué (Tab. 6) :

Eva Joly		
Doublet	**Terme épicène**	**Masculin générique**
Chers concitoyens, chères concitoyennes		*Français de l'étranger* (3 fois)
Nos concitoyen-ne-s		

Marine Le Pen		
Doublet	**Terme épicène**	**Masculin générique**
Françaises et Français de l'étranger	*des compatriotes*	*Français* (4 fois)
		chers compatriotes (3 fois)
		chers amis

Tableau 6. Traitement appliqué aux désignateurs et termes d'adresse par les candidates

De ces tableaux, on relève qu'Eva Joly évite le masculin générique dans le cas de *concitoyen* (utilisé comme terme d'adresse ou comme désignateur). En revanche, elle conserve le masculin pour *Français de l'étranger* – un peu comme s'il s'agissait d'une formule figée (et donc non susceptible de double-marquage) – à la fois pour s'adresser à ses

compatriotes de façon directe (*Français de l'étranger, votre expérience*…) ou indirecte (*une vidéo à l'adresse des Français de l'étranger*).

En revanche, Marine Le Pen interpelle son électorat avec le doublet Françaises et Français de l'étranger, mais elle l'abandonne par la suite et utilise dans le corps du texte Français expatriés ou d'autres masculins génériques.

4.1.3. Style rédactionnel et parité linguistique (corpus papier)

Dans les déclarations papier, au-delà d'une éventuelle formule d'appel bi-genrée, la parité linguistique – comme on pouvait s'y attendre – n'a pas mobilisé les candidats ou candidates. Toutefois, le style rédactionnel utilisé peut aboutir à une relative neutralisation du discours. Ainsi, le recours aux déictiques personnels permet d'éviter d'utiliser un désignateur à la troisième personne : *Nous voulons pouvoir décider de notre avenir* (Poutou). Des locutions comme *personnes handicapées* (Bayrou) ou *en situation de handicap* (Joly), des termes génériques comme *le peuple français* (Le Pen) contribuent à la neutralisation. De même, *création de 60.000 postes dans l'éducation* (Hollande) est plus neutre que *on pourra financer la formation des enseignants* (Cheminade). Tout comme *10% de la richesse produite est passée des poches du travail à celles du capital* (Mélanchon) évite d'introduire le désignateur *travailleurs*.

En terme de neutralisation (voulue ou fortuite ?), les déclarations de J.-L. Mélanchon (Front de gauche) et de N. Dupont-Aignan (Droite Républicaine) offrent un résultat remarquable. Les deux discours se présentent sous forme de listes de propositions (à l'infinitif pour Dupont-Aignan, avec des nominalisations pour Mélanchon), accompagnées d'un message dans lequel prédomine le déictique *nous*. Par ailleurs, dans les deux professions de foi, le nombre de désignateurs de référents humains est extrêmement réduit, ce qui évite le choix du genre (Tab.7).

Candidat	Nombre de mots	Nombre de désignateurs et termes d'adresse au masculin générique	Nombre de mots épicènes	Nombre de formes bi-genrées
Mélanchon	1006	6	1	3
Dupont-Aignan	1254	5	3	1

Tableau 7. Un nombre réduit de désignateurs et termes d'adresse

Si un effet de neutralisation ressort ainsi de certaines déclarations, la particularisation en revanche est plus rare. Elle est réalisée presque exclusivement par le recours à un double-marquage de pronoms indéfinis ou démonstratifs : toutes et tous : *partager le travail entre toutes et tous* (Poutou), chacune et chacun : *c'est considérer chacun et chacune d'entre nous comme une chance* (Hollande), celles et ceux : *la vie n'a jamais été rose pour celles et ceux qui n'ont ni capitaux, ni rentes* (Arthaud). A côté de cela, l'unique substantif bi-genré présent dans le corps de 8 des 10 déclarations est *hommes-femmes,* le plus souvent en collocation avec *égalité salariale* ou *parité*.

Ceci étant, il est intéressant de relever chez Nathalie Arthaud l'usage de la forme bi-genrée comme élément de reprise après une liste de désignateurs au masculin générique, que ce soit sous forme d'anaphore pronominale : *je m'adresse aux électeurs, aux ouvriers, aux employés, aux retraités, [...], aux paysans qui n'exploitent qu'eux-mêmes. Je m'adresse à celles et ceux qui …* ou bien nominale : *il y a parmi les travailleurs des hommes et des femmes qui …* Ici, la présence du double-marquage, outre le fait que l'énumération contient essentiellement des noms de profession, amène à se demander si les termes au masculin générique désignent des personnes ou bien des fonctions. Ces énoncés sont par conséquent un bon exemple de la difficulté qu'il peut parfois y avoir à faire la distinction.

D'après Dister / Moreau (2006), les partis de gauche ont plus tendance que ceux de droite à mettre les noms de titre et de profession au féminin pour désigner les candidates (féminisation lexicale). Dans notre corpus, les choses ne sont pas aussi tranchées en ce qui concerne l'aspect plus général de la prise en compte linguistique des femmes (rédaction épicène). En effet, plus les discours contiennent de

désignateurs de référents humains, plus ils sont susceptibles de s'éloigner de la parité linguistique, si ceux-ci sont au masculin générique. Or, on constate que les trois désignateurs les plus fréquemment utilisés sont : *Français*, *salariés*, *travailleurs*. Les discours de droite et plus encore d'extrême-droite ont tendance à prendre en compte les sentiments nationalistes de l'électorat et à recourir abondamment au désignateur ou au terme d'adresse *Français*. En revanche, les discours d'extrême-gauche font appel ou référence au peuple des *travailleurs*, des *salariés* et également, quoique plus rarement, des *chômeurs*.

Nous avons calculé les occurrences de *Français*, *travailleur(s)* et *salarié(s)* par rapport à l'ensemble des substantifs désignateurs de référents humains au masculin générique présents dans les déclarations. Il ressort de ce calcul que ce sont deux femmes, qui à l'extrême-gauche (N. Arthaud) et à l'extrême-droite (M. Le Pen), recourent le plus souvent aux termes cités ci-dessus : Arthaud l'emporte avec 16 occurrences de *salariés* et *travailleurs* (sur 43 désignateurs) tandis que Le Pen détient le record de *Français* avec 9 occurrences dans sa déclaration qui ne contient au demeurant que 18 désignateurs. Il apparait ainsi que la rhétorique des partis extrémistes tend à ne pas favoriser la rédaction épicène.

Il est par ailleurs intéressant de relever que ces mêmes partis qui pratiquent abondamment l'usage du masculin générique sont également ceux qui n'accordent pas d'attention spécifique aux femmes dans le contenu des programmes. De fait, alors que les uns et les autres promettent d'œuvrer pour l'égalité des salaires, la parité entre hommes et femmes, la défense du droit à l'avortement… Marine Le Pen et Nathalie Arthaud ne consacrent aucune ligne de leur discours à des projets visant particulièrement les femmes.

4.2. Le corpus Elections prud'homales

Les tracts diffusés par les syndicats à l'occasion des élections prud'homales se présentent généralement sur une page simple ou recto-verso, souvent illustrée de dessins qui peuvent faire apparaitre

des hommes et / ou des femmes. La plupart du temps, ils comportent aussi une accroche qui ressort typographiquement du reste du texte.

Dans la mesure où la particularisation du discours est plus immédiatement perceptible que la neutralisation, dans ce corpus, nous avons concentré notre attention sur le double-marquage.

4.2.1. Double-marquage des désignateurs

Sur les 47 tracts recueillis, 17 présentent des traces explicites d'évitement du masculin générique. Ce sont ces derniers que nous avons analysés.

Il apparait à la première observation que le choix de faire apparaitre les deux genres côte à côte ne s'applique pas indifféremment à tous les désignateurs. L'extrait suivant en est une bonne illustration : *Il est donc important que tous les adhérent-e-s et militant-e-s de l'UNSA, y compris de la Fonction publique, se mobilisent en proposant à des salarié-e-s du secteur privé d'être candidat-e-s, que ce soit leur conjoint, amis ou relations* (UNSA-Anpe). Il ressort ainsi que les mots qui figurent sous forme doublement marquée appartiennent presque exclusivement au champ lexical du monde du travail et des élections. Dans le sous-corpus des 17 tracts, nous avons relevé la présence de doublets pour les substantifs suivants (Tab. 8) :

Désignateur	**Formes du doublet**
salarié	*salarié(e)* *salarié(e)s* *salarié-e* *salarié-es* *salarié-e-s*
candidat	*candidat-e-s*
retraité	*retraité-es*
français	*français-es*
immigré	*immigré-es*
militant	*militant-e-s*
adhérent	*adhérent-e-s*
représentant	*représentant-e-s*
chômeur	*chômeurs / ses* *chômeurs et chômeuses*

exclu	*exclu-e-s*

Tableau 8. Les désignateurs doublement marqués dans les tracts

Comme on peut le remarquer dans le tableau, une certaine anarchie règne quant au choix de la forme du doublet (pleine ou abrégée, trait d'union, barre oblique ou parenthèses). Des variations apparaissent d'un tract à l'autre mais aussi à l'intérieur d'un même tract : *un combat solidaire avec les chômeurs / ses et les exclu-e-s* (Solidaires-national) ; *français-es et immigré-es, salarié-es du privé et du public, chômeurs et chômeuses* (Solidaires-Var). Précisons, à propos des formes abrégées, qu'il n'y a pas de véritable harmonie au sein de la Francophonie. De fait, les guides québécois, ou plus généralement canadiens, déconseillent d'avoir recours aux parenthèses, trait d'union ou barres obliques pour l'inscription du féminin. L'OQLF le justifie par le fait que ces formes télescopées ne sont pas conformes aux règles grammaticales et nuisent à la lisibilité des textes et à la clarté de la communication. Les guides suisses en revanche n'ont pas un avis unanime sur le sujet. *Ecrire les genres* recommande l'usage du trait d'union ("et non la parenthèse ou la barre d'exclusion") "pour ne pas avoir à réécrire les substantifs et les adjectifs faisant leur féminin en *e*", tandis que le *Guide de formulation non sexiste* conseille de limiter l'usage des barres obliques aux documents de style "lettre type" ou "décision type" et de ne pas utiliser les parenthèses et traits d'union.

4.2.2. Effet du double-marquage sur la cohérence du texte

C'est dans le titre ou la formule d'accroche qu'on rencontre le plus de formes doublement marquées, celles-ci n'étant pas systématiquement reprises dans le reste du texte. Ainsi l'UNSA-Sictame titre : *L'UNSA veut pour chaque salarié(e) des conseils de Prud'hommes simples, proches et rapides*, mais par la suite utilise 7 fois sur 9 le terme *salarié* au masculin. En revanche, Solidaires-national titre : *Des élections très importantes pour les salarié-e-s* et conserve cette forme bi-genrée 6 fois sur 10 occurrences du mot. Pour sa part, Solidaires-Var, exhorte les salariés : "*Votez pour un syndicalisme au plus près des salarié-es*" et maintient ensuite systématiquement la forme bi-genrée (8 fois). On pourrait certes voir l'absence de régularité dans le

double-marquage d'un même mot comme un manque de cohérence. Toutefois, ce serait ne pas prendre en compte que, dans les tracts, l'accroche a une fonction tout à fait fondamentale : celle d'attirer l'attention du lectorat. Or, par ses caractères graphiques particuliers, le double-marquage, surtout sous forme abrégée, attire l'œil parce qu'il est inhabituel en français. Dans ce sens, il peut être considéré comme un signe de la dynamique linguistique des rédacteurs (ou rédactrices ?) qui jouent avec les possibilités créatives de la langue pour attirer l'attention à l'endroit désiré. Une fois l'effet produit, il semble qu'il n'y ait pas unanimité quant à l'exigence que le double-marquage soit rigoureusement conservé à chacune des occurrences du désignateur dans le corps du texte.

Il est intéressant de noter par ailleurs l'usage du pronom bi-genré *celles et ceux* dans l'accroche des tracts de la CFE-CGC : *Aux élections prud'homales, choisissez celles et ceux qui vous défendent* ! Cette formulation, qui permet de rendre linguistiquement visibles candidats et candidates, ne s'applique en revanche pas aux individus amenés à voter : *Pour la CFE-CGC, le droit et la défense des salariés ne s'arrêtent pas aux murs de l'entreprise.* Sans doute le syndicat souhaite-t-il insister davantage sur la parité hommes-femmes au sein des conseils prud'homaux que sur le fait que l'électorat auquel il s'adresse est composé de salariés des deux sexes.

4.2.3. Effet du double-marquage sur la cohésion du texte

Le recours à la féminisation implique une adaptation de la syntaxe qui doit en principe respecter le choix de la forme bi-genrée du désignateur en tête de proposition. Toutefois cette forme de cohésion semble élastique. Ainsi, on relève des cas où adjectifs et participes passés suivent le double-marquage du substantif auxquels ils se réfèrent : *Tout-e salarié-e qui s'estime lésé-e par son employeur …* (Solidaires-national), *une minorité où chacun et chacune serait soi-disant seul(e) responsable de sa situation* (Solidaires-Var), et d'autres cas où ces mêmes éléments restent au masculin générique: *condamnant à verser plusieurs dizaines de milliers d'euros à des salarié-e-s lésés* (SUD Virgin), *SALARIE(E)S "SAISONNIERS" il est possible de gagner !* (CGT). Là encore, les guides suisses et canadiens

ne donnent pas une orientation nette. Conformément à l'usage en vigueur au Moyen-âge et jusqu'au XVII[e] siècle (cf. § 2.1), *Ecrire les genres* recommande l'accord au plus proche. En revanche, dans sa Banque de dépannage linguistique, l'Office québécois de la langue française, ne retient que l'accord au masculin pluriel.

4.2.4. Stratégie plus que système

Ainsi que nous venons de l'observer, notre corpus de tracts révèle que, même en l'absence de toute incitation de politique ou aménagement linguistique, certains syndicats français sont soucieux de rendre les femmes linguistiquement visibles. Toutefois il ressort que les marques de cette visibilité restent hétérogènes. D'un tract à l'autre, et souvent même à l'intérieur d'un même tract, on constate une grande irrégularité. Plutôt que d'y voir un manque de cohérence ou de cohésion, nous préférons considérer, tout comme le fait Abbou à propos d'un corpus de brochures libertaires, que : "s'il n'y a pas d'uniformité dans le marquage, c'est qu'il s'agit davantage de stratégie que de système" (Abbou 2011 : 129). En effet, les syndicats qui recourent au double-marquage cherchent sans doute plus à montrer une attention particulière à l'égard des salariées ou des demandeuses d'emploi et à attirer leur bienveillance qu'à appliquer rigoureusement un système de féminisation syntaxique. A propos de stratégie et de système, il est d'ailleurs intéressant de remarquer que notre corpus peut être scindé en deux. D'un côté, on trouve les grandes confédérations syndicales reconnues représentatives (CGT, CFDT, FO) qui sont les plus soucieuses de respecter les règles du système, de la grammaire traditionnelle et qui ne rendent pas les femmes linguistiquement visibles (hormis quelques tracts de la *CFE-CGC* où l'on relève la présence de *celles et ceux*). De l'autre, il y a les syndicats minoritaires (en particulier Solidaires et dans une moindre mesure l'UNSA) pour qui la présence du "petit e- qui traîne partout et parasite nos textes" peut être considérée comme "un acte aussi politique que d'aller manifester contre le Grand Capital, d'écrire des textes contre les oppressions des états [*sic*] ou de créer des alternatives durables..." (Solidaires-Sud-Education 2007). Si dans les discours de ces syndicats le double-marquage se révèle parfois irrégulier, peut-être

faut-il dès lors y voir l'expression d'un syndicalisme de lutte et de transformation sociale qui invite à "s'organiser pour […] éviter les phénomènes d'institutionnalisation, […] d'experts, de superstructures que connaissent les grandes confédérations" (Solidaires-Var), tant en ce qui concerne le contenu discursif que la forme expressive.

5. Remarques conclusives

Dans un contexte français peu enclin à la prise en compte linguistique des femmes dans les textes, la présente analyse sur corpus a montré que les discours électoraux n'y sont pas cependant pas toujours indifférents. Certes, cette parité peut parfois être fortuite lorsqu'elle est atteinte par la neutralisation. Mais les occurrences de formes doublement marquées dans plusieurs écrits papiers ou électroniques analysés traduit chez certains une évidente volonté de rendre les femmes visibles comme citoyennes, électrices, candidates, travailleuses, salariées, chômeuses… à côté des hommes.

Les observations d'ordre sociolinguistique et linguistique auxquelles l'analyse a donné lieu montrent l'intérêt didactique que peut présenter une telle étude. D'un point de vue sociolinguistique, les corpus utilisés sont de bons instruments pour faire réfléchir les apprenants aux représentations et au rôle social de la langue. Ils ouvrent aussi le regard vers la variation diatopique, permettant, par la comparaison entre pays, d'observer d'intéressantes différences entre le français écrit de France, de Suisse, de Belgique ou du Canada. D'un point de vue linguistique et textuel, ils peuvent donner lieu à un travail sur le genre grammatical, la cohésion et la cohérence textuelle. A cet égard, ils se prêtent aussi bien à la didactique du français langue maternelle qu'à celle du français langue seconde ou étrangère. Quant au choix du genre de textes sur lesquels faire porter l'analyse, en dehors de ceux que nous avons abordés, il est large et s'étend des annonces d'emploi – où l'écriture épicène est grandement encouragée par l'obligation de rédaction non discriminatoire – aux formulaires de

toutes sortes, en passant par les contrats et les lettres ou courriels circulaires.

Références bibliographiques

Abbou, Julie 2011. *L'Antisexisme linguistique dans les brochures libertaires : Pratiques d'écriture et métadiscours*, Thèse de Doctorat en Sciences du Langage. Université d'Aix-Marseille.

Académie Française, 2002. *Féminisation des noms de métiers, fonctions, grades et titres*. <http://www.academie-francaise.fr>

Baider, Fabienne / Khaznadar, Edwige / Moreau, Thérèse (eds) 2007. Parité linguistique. *Nouvelles Questions féministes* 26 / 3.

Béguelin, Marie-Josée 1999. Les consignes de "féminisation" du lexique et du discours : l'exemple de la Suisse romande. In Almeida, Elisete Maillard, Michel (eds) *O Femino nas Linguas, Culturas e Literaturas*, Madeira: Universidade da Madeira. Coleçâo Metagrama, 37-49.

Béguelin, Marie-Josée / De Pietro, Jean-François 2000. Le féminin dans la langue : espace de variation et de réflexion. *Français et société* 10, 30-44.

Bureau de l'égalité 2002. *Ecrire les genres. Guide romand d'aide à la rédaction administrative et législative épicène*. < http://www.egalite.ch/langage-epicene.html>

CNRS / INLF 1999, *Femme j'écris ton nom… Guide d'aide à la féminisation des noms de métiers, titres, grades et fonctions*. Paris : La Documentation française. <http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/994001174/0000.pdf>

Commission générale de terminologie et de néologie 1998. *Rapport sur la féminisation des noms de métier, fonction, grade ou titre.* <http://www.dglf.culture.gouv.fr/cogeter/feminisation/sommaire.html>

Dister, Anne / Moreau, Marie-Louise 2009. *Féminiser ? Pas sorcier !*. Bruxelles : De Boeck- Duculot.

Dister, Anne / Moreau, Marie-Louise 2006. "Dis-moi comment tu féminises, je te dirai pour qui tu votes". Les dénominations des candidates dans les élections européennes de 1989 et de 2004 en Belgique et en France, *Langage et société* 116, 5-46.

Gouvernement de l'Ontario 2003. *La féminisation au gouvernement de l'Ontario*. Toronto : Service de traduction du gouvernement.

Grévisse, Maurice, 1986. *Le bon usage*. Paris-Gembloux : Ducolot.

Chancellerie fédérale 2000. *Guide de formulation non sexiste des textes administratifs et législatifs de la confédération*. <http://www.bk.admin.ch/dokumentation/sprachen/04908/05037/index.html?lang=fr>

Matthey, Marinette 2000. Féminisation du lexique et du discours en Suisse romande. Un état des lieux. *Bulletin VALS-ASLA* 72, 63-79.

Mougeon, Françoise. Préface. In Ordre des enseignantes et des enseignants de l'Ontario *Féminisation des documents en français* <http://www.oct.ca/~/media/PDF/Feminization%20of%20French%20Publications/FR/fem_policy_guide_f.ashx>

Moreau, Thérèse 2007. Prière de ne pas épicéner partout. In Baider *et al.* (eds) 2007. Parité linguistique. *Nouvelles Questions féministes* 26 / 3.

Vachon L'Heureux, Pierrette / Guénette, Louise 2007. *Avoir bon genre à l'écrit : guide de rédaction épicène*. Québec : Les Publications du Québec.

Perry, Véronique (ed.) 2004. *Désexisation des textes et parité linguistique : le cas de la langue française*. Toulouse : ANEF.

Riegel, Martin / Pellat, Jean-Christophe / Rioul, René 2006. *Grammaire méthodique du français*. Paris : PUF.

Solidaires-Sud-Education 2007. *Le pourquoi de la féminisation des textes* <http://www.sudeducation.org/Le-pourquoi-de-la-feminisation-des.html>

Vachon-L'Heureux, Pierrette 2004. Féminisation des titres et des textes. *Correspondance* 10 / 2. <http:/correspo.ccdmd.qc.ca/Corr10-2/Feminisation.html>

Wagner, Robert-Léon / Pinchon, Jacqueline 1962. *Grammaire du français classique et moderne*. Paris : Hachette.

DANIO MALDUSSI

17. Anisomorphisme et relation de converse à l'épreuve des corpus spécialisés : le couple "créance"/ "crédit" par opposition à "credito"

1. Introduction

L'*anisomorphisme*, en tant que divergence constatée entre les unités lexicales de deux langues différentes, s'oppose à l'*isomorphisme*. Causé par "the differences in the organisation of the designate in the individual languages and by the differences between languages" (Manual of Lexicography 1971 : 294), l'*anisomorphisme*, peut être entendu ici comme un découpage différent du monde extralinguistique qui se reflète dans un découpage différent du sens des termes. De ce manque de superposition parfaite découlent des écarts lexicaux issus de visions du monde différentes ou de pratiques sociales différentes et parfois radicalement opposées. Cette définition est renforcée par celle de De Mauro dans son *Grande dizionario italiano dell'uso*. Dans cet ouvrage, l'anisomorphisme, dont la marque d'usage est TS, sigle qui indique les termes techniques spécialisés appartenant, dans ce cas d'espèce, au domaine de la sémiologie, est un

> fenomeno per cui lingue diverse esprimono in forma differente lo stesso senso, in modo tale che parole d'ambito d'uso molto simile hanno significati non sovrapponibili in modo univoco (per es. la parola fr. *bois* e le italiane *bosco* e *legno*, o la parola it. *fiume* e le francesi *fleuve* e *rivière*). CONTR. Isomorfismo (2000 : 293).

L'anisomorphisme des langues a donné lieu dans le passé, parmi ses effets les plus néfastes et improductifs, à la conviction de l'impossibilité de la traduction et donc d'une véritable communication

entre différentes visions du monde. Or, comme l'écrit Eco, c'est bien avec la contribution de la sémiotique de Hjelmslev (1943) que ce problème a été définitivement résolu. En effet, "incommensurabilità [di sistemi del contenuto] non significa incomparabilità" (Eco 2003 : 41). Ce débat, profondément enraciné, est assimilable à celui qui a caractérisé les premiers pas de la terminologie. En effet, selon la Théorie générale de la terminologie proposée par l'Ecole de Vienne, dont Wüster a été l'initiateur, certains phénomènes tels que l'homonymie et la polysémie étaient entourés d'une aura négative car susceptibles d'empêcher une communication claire. Or, cette conception négative a changé du fait de la prise de conscience de l'importance des éléments connotatifs dans la communication spécialisée[1]. D'une vision du lexique en tant que nomenclature, on passe aujourd'hui à la reconnaissance de la valeur discursive de la polysémie[2]. Comme l'écrit Prandi, des phénomènes tels que l'anisomorphisme, l'homonymie, la polysémie et la synonymie, tout en présentant des caractéristiques spécifiques sujettes à des conséquences variables dans la communication, représentent des "spie del radicamento endocentrico dei concetti". Or, toujours d'après le même auteur :

> L'anisomorfismo non si lascia facilmente ridurre, ma in compenso non ha ricadute serie sulla comunicazione interlinguistica. Un traduttore ne deve essere consapevole, ma senza sopravvalutarlo. Se i singoli termini non hanno corrispondenti diretti nella lingua di arrivo, le relative aree concettuali e le pratiche e le istituzioni sulle quali poggiano sono largamente condivise. Per un buon traduttore economico-finanziario dall'italiano al francese, ad esempio, la presenza del solo termine *credito* nella lingua di partenza non è un ostacolo all'uso corretto degli equivalenti *crédit* e *créance* nella lingua d'arrivo (Prandi 2009).

---

1 Sur le concept de "flou terminologique" lié à la notion de connotation, voir Magris / Musacchio (2009).

2 Sur l'inclusion des aspects liés à la variation linguistique dans la communication spécialisée, voir la théorie communicative de la terminologie proposée par Cabré (1998), la théorie sociocognitive de Temmerman (2000) et la socioterminologie de Gaudin (2003).

Pour analyser les différences de comportement entre l'italien et le français relatives aux technicismes "créance" et "crédit" pour le français et l'italien "credito", nous avons construit des corpus spécialisés de type comparable dans les deux langues, à savoir des corpus portant sur le même contenu spécialisé et partageant le même registre et la même typologie textuelle, mais qui ne sont pas la traduction les uns des autres[3]. Le domaine considéré est celui du surendettement des ménages, un domaine sociétal de forte actualité. Dans notre article, nous ne nous attarderons pas sur les aspects techniques tel que le volet informatique et le traitement par concordancier, sujets sur lesquels il existe déjà une vaste littérature[4]. Nous nous focaliserons en revanche sur la valeur ajoutée d'une didactique de la traduction technique supportée par l'exploration de corpus spécialisés comparables[5]. Gardant toujours à l'esprit l'analyse

---

3 La première définition de "corpus comparable" est celle de Baker (1995 : 234): "two separate collections of texts in the same language : one corpus consists of original texts in the language in question and the other consists of translations in that language from a given source language or languages. […] both corpora should cover a similar domain, variety of language and time span, and be of comparable length". La notion de "original texts" a été l'objet de précisions ultérieures de la part de Baker. La définition de "corpus comparable" que nous retenons ici est celle que nous avons explicitée dans notre texte, à savoir des corpus portant sur le même contenu spécialisé et partageant le même registre et la même typologie textuelle mais qui ne sont pas la traduction les uns des autres.

4 Les deux corpus comparables que nous avons construits pour notre analyse grâce au logiciel en ligne BootCat (http://bootcat.sslmit.unibo.it/), le premier consacré aux occurrences de "credito" et le deuxième à celles du couple "créance / crédit", contiennent respectivement 95.983 et 97.252 mots. Comme le confirme la littérature sur la linguistique de corpus, les corpus spécialisés comparables présentent une dimension réduite mais offrent, par contre, l'avantage d'être ciblés sur un domaine ou un sous-domaine. Pour un approfondissement de cette problématique appliquée au domaine de la traduction, voir l'article de Gavioli / Zanettin (2000).

5 Sur la linguistique de corpus appliquée au domaine de la traduction, nous signalons : Bernardini / Zanettin (2000), Bernardini *et alii* (2003), Sinclair (1991), Baker (1995), Olohan (2004) et le numéro monographique *Meta* (1998).

de Prandi, nous montrerons, à l'appui de plusieurs exemples concernant les occurrences des termes "créance" et "crédit" pour le français et "credito" pour l'italien, tout l'intérêt qu'il y a, pour l'étude de la langue, à se fonder sur des bases de données textuelles authentiques tirées de corpus représentatifs de l'objet d'étude (approche *corpus-driven* ou guidée sur corpus) et qui permettent donc une prise directe sur les différents comportements langagiers. Cet appui ne se limite pas à fournir des indications sur la productivité des phénomènes étudiés. Face au flou terminologique qui entoure le couple "créance" / "crédit" dans une perspective intra et interlinguistique, l'étude des corpus nous offre des contextes réels qui permettent de faire émerger les différents découpages du sens des termes et de rendre plus accessibles les concepts véhiculés par ces derniers. La valeur ajoutée d'une analyse contrastive sur base textuelle authentique permet d'étudier et de répertorier des comportements langagiers et des stratégies de production du texte écrit différents et parallèlement de repérer des stratégies de traduction adéquates au texte cible, notamment dans le cas d'écarts lexico-sémantiques patents. Notre étude adopte donc une démarche pragmatique et professionnalisante de type inductif. Pour reprendre les mots de Prandi, l'anisomorphisme est un obstacle si l'on accorde une importance excessive au mot isolé; si, au contraire, nous explorons un contexte plus large, notamment grâce à la construction de corpus comparables spécialisés, nous nous apercevons que cette différence s'estompe, s'amenuise et que la traduction est plus immédiate, sans oublier que c'est bien la traduction qui nous met dans l'obligation de déconstruire les concepts génériques. Pour reprendre le titre très évocateur de l'essai de Hédiard *L'analyse guidée sur corpus : une approche "décloisonnante" pour accéder au sens* (2007), l'approche par corpus est *décloisonnante* par excellence, mais nous ajoutons que c'est bien l'opération traduisante qui fait éclater le phénomène des anisomorphismes, ces apparentes contradictions inscrites dans le code génétique des langues. Ainsi, nous pouvons affirmer que la traduction est, à son tour, une opération décloisonnante, en ce sens qu'avant d'entrer en contact par la traduction, ces mêmes anisomorphismes appartenaient à des systèmes linguistiques parfaitement cohérents.

2. Le surendettement des ménages

Dans notre étude, nous allons explorer le couple "créance / crédit" *vs* "credito" dans le contexte du surendettement des ménages. Le phénomène du surendettement, qui accable de nos jours un nombre croissant de ménages, se caractérise par la difficulté, voire l'impossibilité, de rembourser les dettes personnelles contractées au fil du temps. En France, la prise en charge des dossiers de surendettement des ménages revient à une commission spécifique appelée Commission de surendettement, sous l'égide de la Banque de France[6]. L'Italie, de son coté, a promulgué en 2011 le Décret-loi du 22 décembre 2011 n. 212 "Rateizzazione dei debiti e Disciplina del sopraindebitamento"[7] et successivement la Loi du 27 janvier 2012, n.3, "Disposizioni in materia di usura e di estorsione, nonché di composizione delle crisi da sovraindebitamento"[8], des mesures qui témoignent de l'internationalisation du phénomène. Deux raisons principales motivent le choix de ce sujet : en premier lieu son actualité que la crise sociale et financière qui sévit autour du globe ne fait qu'accroître et, en deuxième lieu, son degré de technicité qui ressort de l'usage simultané de technicismes financiers tels que le couple créance / crédit qui se prêtent bien à une analyse contrastive en didactique de la traduction spécialisée, appuyée par l'exploration de corpus spécifiquement construits. Comme l'explique le site Rachat du

---

6 Pour plus de détails, voir : <http://www.banque-france.fr/la-banque-de-france/missions/protection-du-consommateur/surendettement.html> [consulté le 25 mars 2013].

7 <http://www.gazzettaufficiale.it/gunewsletter/dettaglio.jsp?service=1&datagu=2011-12 22&task=dettaglio&numgu=297&redaz=011G0255&tmstp=1325582228682> [consulté le 8 septembre 2013].

8 Le texte est disponible en ligne à l'adresse suivante : <http://www.odcec.roma.it/index.php?option=com_content&view=article&id=1469:crisi-da-sovraindebitamento&catid=18:comunicazioni-dallordine&Itemid=61INS> [consulté le 8 septembre 2013].

crédit[9], il existe deux typologies de surendettement : "le surendettement passif lié à un accident de vie (cause de divorce, décès, perte d'emploi, maladie…)" et "le surendettement actif suite à un excès de crédits". Le Décret-loi italien du 22 décembre 2011 définit, pour sa part, la situation de surendettement :

> una situazione di perdurante squilibrio tra le obbligazioni assunte e il patrimonio liquidabile per farvi fronte, nonché la definitiva incapacità del debitore di adempiere regolarmente le proprie obbligazioni[10]

Or, toujours sur le site Rachat du crédit, nous trouvons une description des crédits et des dettes souscrits par le particulier surendetté que nous considérons particulièrement intéressante à exploiter du point de vue didactique et sur laquelle nous centrerons notre attention dans la présente étude :

> Le dépôt d'un dossier de surendettement devant la banque de France (bdf) est destiné pour les particuliers endettés et surendettés qui ont souscrit trop de crédits et dettes. Les créances impayées sont des crédits revolving, cartes de crédits permanents, emprunt immobilier, prêt à la consommation, crédit automobile, des emprunts pour travaux, pensions alimentaires, retard d'impôts, surendettement immobilière…pour les propriétaires ou locataires.

En particulier, nous allons nous concentrer sur la partie centrale de cette définition : "[…] qui ont souscrit trop de crédits et dettes. Les créances impayées sont des crédits revolving […]" qui présente une intéressante reprise anaphorique [Les créances impayées], à même de mettre en question la stabilité de la relation d'antonymie qui caractérise les définitions classiques de "créance" et de "dette" proposées par les dictionnaires monolingues. Reprise anaphorique que

---

9 <http://www.rachatducredit.com/dossier-surendettement.html> Le rachat de crédit est une solution au problème du surendettement qui peut être proposée par une banque ou un organisme de crédit et qui consiste en la renégociation d'un crédit souscrit à un taux défavorable [consulté le 11 septembre 2013].

10 <http://www.camera.it/_dati/leg16/lavori/stampati/html/decretolegge%5Carticoli%5Ccomponiarticoli.asp?CodiceBarre=16PDL0056790&NomeFile=16PDL0056790_Capo_I_Articolo_1_.html> [consulté le 8 septembre 2013].

nous verrons dans le prochain paragraphe. Comme l'écrit Hédiard dans son article précédemment cité, "[n]ous nous trouvons en présence d'un microsystème non isomorphe entre l'italien et le français"[11]. Le système "créance / crédit" par opposition à "credito" explique la difficulté de comprendre les caractéristiques d'emploi de ces deux termes. Or, en classe, comment expliquer aux étudiants cette apparente contradiction, qui voit la reprise anaphorique d'un antécédent, "dette" considéré comme antonymique de "créance", et en même temps une reprise anaphorique qui assure la fonction de cataphore en listant les produits qui composent les "créances impayées" ? Comment faire prendre conscience aux étudiants de l'importance de reconstruire le tissu conceptuel sous-jacent à cette définition et comment prouver l'utilité de recourir au bagage encyclopédique individuel et à la compétence culturelle ? Du point de vue didactique, montrer aux étudiants le degré d'interpénétration de ces points sensibles et leur participation stratégique à la compréhension finale s'avère une problématique de taille. La première démarche sera donc de pénétrer en profondeur dans le fonctionnement de ces termes dans les codes linguistiques respectifs en partant des définitions classiques qu'en donnent les dictionnaires monolingues. Nous analyserons ensuite comment l'étude sur corpus est à même d'étoffer et de préciser les contours de la relation d'antonymie qui, d'après les définitions classiques susmentionnées, s'instaurerait entre créance et dette.

11 Hédiard signale un autre exemple de microsystème non isomorphe : les adverbes "entre" et "parmi" (2007 : 229).

3. Créance, crédit, crédito : l'apport des définitions classiques

Selon la définition qu'en donne le CNRTL (Centre National de Ressources Textuelles et Lexicales)[12], qui reprend celle du TLF, une créance (marque d'usage, Droit) est le

> [d]roit que possède une personne d'exiger l'exécution d'une obligation, en particulier le paiement d'une somme d'argent. Recouvrer une créance (cf. A. France, Jocaste,1879, p. 101). Anton. Dette;

par contre un crédit est un

a) Acte par lequel une personne, généralement un banquier, met à la disposition d'une autre personne, une somme d'argent ou un bien qui devra être restitué, ou son équivalent. Ouverture de crédit, crédit privé. Crédit bancaire (cf. Chenot, Entr. national.,1956, p. 60). Lettre de crédit (cf. Baudhuin, Crédit et banque,1945, p. 125).
b) P. ext. Prêt consenti, généralement par une banque.

Le Grand Robert de la Langue française atteste "créance" dans l'acception de "droit en vertu duquel une personne (créancier) peut exiger qqch. De qq. Et, spécialement, une somme d'argent (opposé à dette)" (777). En italien, le terme polysémique " credito" indique à son tour

a) Operazione e rapporto di scambio in cui le due prestazioni sono separate nel tempo, e consistente in genere nella cessione attuale di una somma di denaro da parte del creditore contro la promessa da parte del debitore di controprestazione futura, a scadenza determinata o indeterminata, di altra somma di denaro (in genere superiore alla prima di una percentuale corrispondente all'interesse), oppure nella fornitura di merci con dilazione di pagamento, con vendita a rate, ecc. (in genere a prezzi superiori di quelli in contanti)

12 <http://www.cnrtl.fr> [consulté le 27 juillet 2013].

b) Il diritto, già scaduto o ancora da scadere, di pretendere dal debitore il pagamento di una somma di denaro, o di pretendere l'adempimento di una prestazione di carattere patrimoniale; anche la somma stessa[13].

Comme on peut le constater à partir des définitions susmentionnées [le soulignement est le nôtre] et comme nous le verrons plus clairement ensuite, à partir des données extraites des corpus, "créance" et "crédit" sont deux termes spécifiques qui précisent un concept plus général ; c'est à dire qu'ils ne sont pas reconductibles à un terme commun qui puisse les contenir. Par contre, l'italien regroupe le sens des deux termes spécifiques sous le terme polysémique et générique "credito". Une créance est un droit personnel permettant à son titulaire d'exiger d'une autre personne l'exécution d'une obligation (une personne a une créance vis-à-vis de quelqu'un) ; le crédit est un financement consenti, une somme allouée, un prêt. Comme nous le verrons avec l'analyse de quelques exemples choisis, l'exploration du corpus permet de mettre en évidence les conditions et les restrictions d'emploi des lexèmes "créance", "crédit" et "credito". La signification de la base "credito" (398 occurrences), par exemple, – droit d'exiger le paiement d'une somme ou somme allouée – est précisée plus fréquemment par le collocatif placé à gauche, avec une fréquence nettement supérieure pour les substantifs tels que "accesso", "mercato", suivis par "apertura", "rinuncia" et ce par rapport au verbe à l'infinitif, "accedere" (six occurrences), suivi par "trasferire", "recuperare", "rinunciare", "esigere". En revanche, la présence d'un collocatif placé à droite, comme par exemple le participe passé, est moins significative, comme le montre la fréquence des occurrences de "credito vantato" (créance) et de "credito erogato" (crédit)[14].

---

13 <http://www.treccani.it/vocabolario/credito> [consulté le 27 juillet 2013].

14 La base "crediti" (199 occurrences), par contre, est précisée plus fréquemment par le collocatif à gauche "recupero" (77 occurrences), suivi par "compensazione" (38) et "accertamento" (10). Dans dix cas, le sens de "crediti" est spécifié par le participe passé placé à droite ("acquisiti", "accesi", "prescritti", "vantati", "contratti" etc.).

3.1. Créance, crédit, credito à l'épreuve des corpus comparables

Nous allons maintenant observer le comportement de la langue en contexte grâce à des exemples choisis de contextes élargis. Comme nous l'enseigne la littérature existante sur le sujet, le recours au corpus permet, entre autres, de recueillir des données utiles quant à la fréquence d'une occurrence particulière ou d'éclaircir ses contraintes syntaxiques et sémantiques. Notre recherche se focalisera sur la fonction du contexte dans une perspective contrastive, une démarche stratégique à notre sens à même d'illustrer et de comprendre finement un phénomène linguistique tel que celui de l'anisomorphisme des langues.

Analysons, grâce à la capture d'écran, quelques exemples du fonctionnement du terme "credito" en contexte :

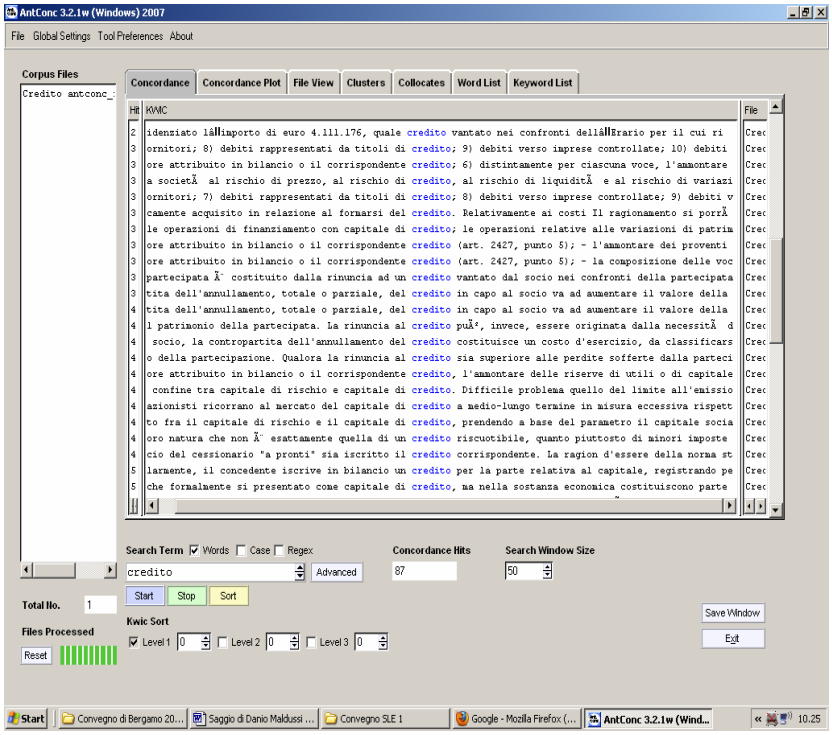


Fig.1 "Credito" dans son environnement linguistique.

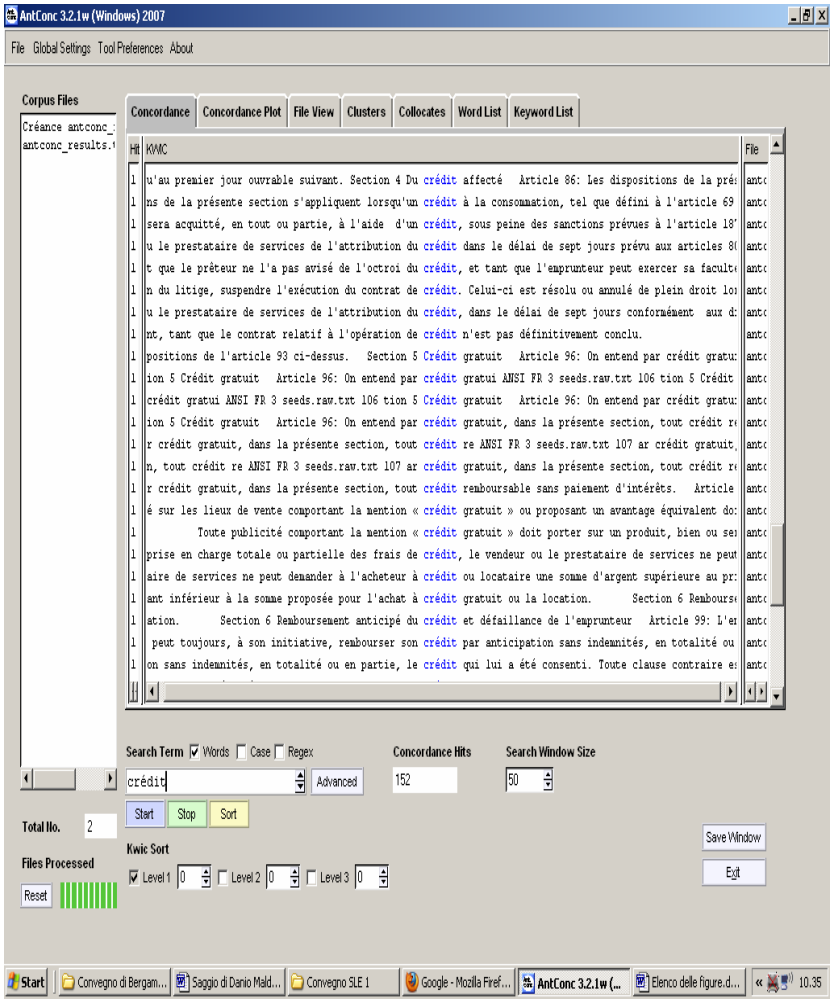


Fig. 2 "Crédit" dans son environnement linguistique.

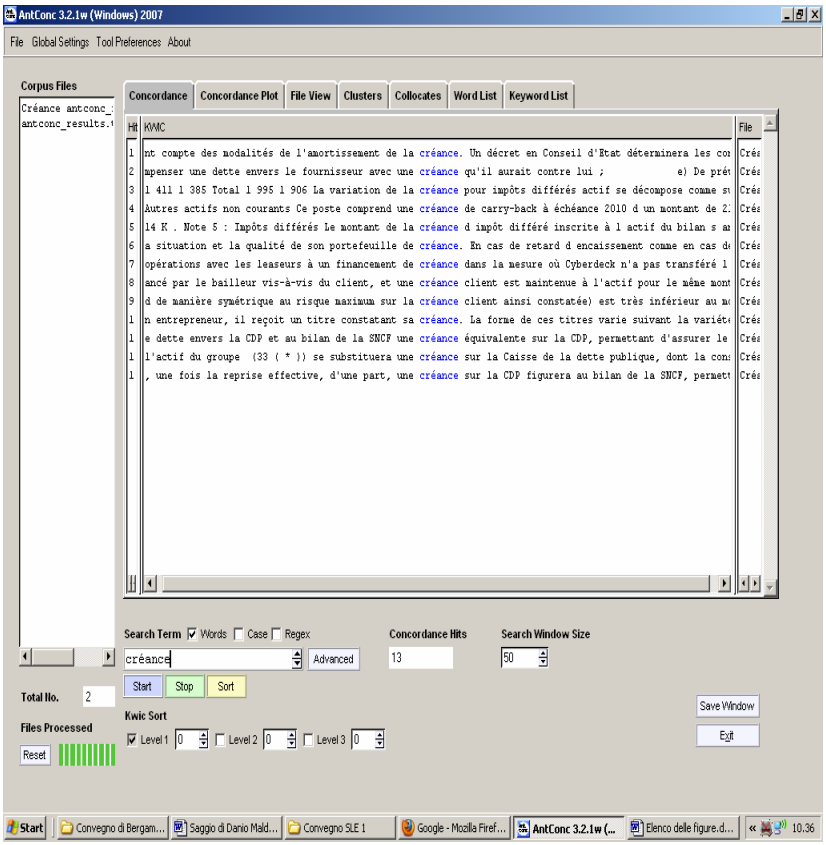


Fig. 3 “Créance” dans son environnement linguistique.

Une première exploration des exemples susmentionnés nous permet d’avancer une remarque importante concernant le comportement des deux langues : en italien, comme le montrent les exemples de la figure 1 (“rinuncia ad un credito”, “rischio di credito” ou “capitale di credito”, “iscrive in bilancio un credito”), le contexte a la fonction de préciser la valeur pertinente du terme polysémique “credito”. Il en va tout autrement en français où le contexte est cohérent avec une valeur déjà spécifiée au niveau lexical, comme en témoignent les exemples “le crédit qui lui a été consenti”, “achat à crédit”, “octroi du crédit”,

"l'attribution du crédit dans le délai de sept jours" (Fig. 2) et "compenser une dette envers le fournisseur avec une créance qu'il aurait contre lui", "il reçoit un titre constatant sa créance", "le montant de la créance d'impôt différé inscrite à l'actif" (Fig. 3). Dans les exemples tirés de la figure 2 le crédit est un prêt, un financement. Dans les exemples de la figure 3, il s'agit du droit que possède une personne d'exiger le paiement d'une somme d'argent, droit qui finit par compenser une dette précédemment contractée. Cette première exploration à partir d'un contexte élargi nous permet donc de "visualiser" le sens des termes et d'en recenser le comportement authentique.

L'exploration systématique guidée sur corpus nous permet, par ailleurs, d'avancer une autre remarque de type didactique en classe de traduction, qui confirme la polysémie du terme "credito". Si les syntagmes verbaux "ricorrere al mercato del capitale di credito" et "ricorrere al credito locale" répérés dans notre corpus peuvent se lier indistinctement au terme "credito" et à son expansion nominale où "capitale" est le noyau et "credito" le complément, comme dans l'exemple "mercato del capitale di credito", il est fort probable que ces compléments du verbe soient en rapport de synonymie. Les exemples "ricorrere al mercato del capitale di credito" et "ricorrere al credito locale" laissent supposer une condition d'interchangeabilité, induisant une relation de polysémie. Le terme "credito" peut donc indiquer la somme que l'on reçoit ou que l'on donne ou le système qui permet d'y accéder. La vérification de l'existence dans le dictionnaire de cette acception nous permet de confirmer, à l'appui de données authentiques, la polysémie du terme "credito".

4. Les contours flous de la relation d'antonymie entre créance et dettes

Revenons maintenant sur la relation entre "créance" et "dette", que les dictionnaires monolingues considèrent comme un cas d'antonymie, et

confrontons-la à la définition de surendettement telle que nous l'avons anticipée au début et que nous reprenons ici:

> Le dépôt d'un dossier de surendettement devant la banque de France (bdf) est destiné pour les particuliers endettés et surendettés qui ont souscrit trop de crédits et dettes. Les créances impayées sont des crédits revolving, cartes de crédits permanents, emprunt immobilier, prêt à la consommation, crédit automobile, des emprunts pour travaux, pensions alimentaires, retard d'impôts, surendettement immobilière…pour les propriétaires ou locataires.

La partie sur laquelle nous allons focaliser notre attention est la partie centrale de la phrase, celle qui contient la reprise anaphorique "créances impayées". Pour ce faire, nous reprenons également la distinction entre "créance" et "crédit" proposée par Prandi qui éclaire particulièrement bien le phénomène :

> Nell'altro senso, all'italiano *credito* corrispondono in francese due termini, *crédit* e *créance*. All'interno della complessa rete di rapporti, obblighi e aspettative che caratterizza una stessa transazione, i due termini introducono prospettive di segno opposto. *Crédit* focalizza la somma effettivamente erogata dal creditore al debitore: in questo senso si parla, ad esempio, di *crédit à l'exportation*. Si tratta di una visione retrospettiva che porta su un dato reale. *Créance*, viceversa, focalizza la somma che il debitore deve al creditore: in questo senso si parla, ad esempio, di *créance douteuse*. Si tratta di una visione prospettiva che investe un dato non reale, oscillando tra una modalizzazione deontica – ciò che è dovuto – e una epistemica: ciò che forse sarà restituito. Come accade nei lessici naturali, il radicamento intralinguistico è inversamente proporzionale alla facilità di circolazione interlinguistica. Non c'è corrispondenza biunivoca tra termini, e il confronto, come la traduzione, richiede un passaggio attraverso la struttura del campo nella sua totalità (Prandi: 2008).

Si nous élargissons le contexte des termes au niveau textuel et si nous analysons la totalité des concepts qui définit le "surendettement", nous remarquons que le syntagme "créances impayées" fait fonction de reprise anaphorique tant de "crédit" que de "dettes". Or, si les deux concepts de "créances" et "dettes" étaient véritablement des antonymes, il serait impossible de reprendre anaphoriquement le couple "crédits / dettes" par "créances". L'hypothèse que nous avançons ici est que "créances" et "dettes", à l'instar du couple "crédit / dettes", entretiennent une corrélation converse. La corrélation de

converse (Lyons 1977 : 279-280), appelée opposition relationnelle par Palmer 1976 : 82), entraîne "pairs of words which exhibit the reversal of a relationship between items". Considérés isolément, deux lexèmes converses semblent former une opposition ; cependant, s'ils sont saturés de façon adéquate, ils forment des structures sémantiques équivalentes. Deux verbes converses comme *prêter* et *emprunter*, par exemple, forment le même procès vu dans des perspectives opposées : *Jean a prêté cent euros à Luc – Luc a emprunté cent euros à Jean*. Deux noms converses, également, décrivent dans des perspectives opposées la même relation : *Luc est père de Jean – Jean est fils de Luc*. Deux lexèmes converses sont logiquement inséparables, ce qui amène Aristote (*Catégories*, 7b) à parler de corrélatifs : "La corrélation entre maître et esclave ne continuera pas d'exister, car sans maître il n'y a plus d'esclave"[15]. Il est utile de souligner, à ce propos, que la reprise anaphorique "créances impayées" fonctionne elle aussi comme si elle était saturée au niveau contextuel par ses arguments appropriés, les mêmes qui, dans des positions différentes, saturent les antécédents "crédit" et "dettes".

L'intérêt de la corrélation de converse en lexicologie est énorme : comme la définition même de la structure lexicale comporte la référence à une construction complexe, spécifiquement à une phrase, elle ouvre le chemin à l'idée que l'unité minimale de la description lexicale n'est pas le lexème isolé mais la structure de la phrase, comme l'affirme par exemple Gross (2012 : Ch. 1, 2). Les termes "créance" et "crédit" qui, comme nous l'avons dit plus haut, introduisent des perspectives différentes dans un même concept, entretiennent la même relation de converse avec "dette" : le bénéficiaire d'un crédit active une dette et "créance" reprend "dette": le binôme crédit-dette génère une créance, à savoir le droit d'exiger un crédit-dette car la relation qui s'instaure est de type converse. La relation de conversion entre "créances" et "dettes" ressort assez

15 La traduction que nous utilisons ici est celle de Nau, disponible en ligne à l'adresse suivante : <http://www.documentacatholicaomnia.eu/03d/-384_322,_Aristoteles,Categories_%28Pascale_Nau%29,_FR.pdf> [consulté le : 4 septembre 2013]

clairement de la définition juridique de “créance” qu’en donne le Cabinet Arc :

> Une créance relève du droit des obligations. Une créance détenue par un créancier correspond à une dette détenue par un débiteur donc, lorsque la dette est soldée, la créance correspondante est également annulée.[16]

Cette définition a le mérite de nous faire entrevoir la complexité du réseau de relations entretenues par “créance” et “dette”, réseau difficilement réductible à une relation d’antonymie. En effet, comme l’écrit Hédiard, les réalisations effectives “remettent en cause les distinctions tranchées entre acceptable et non acceptable” (2007: 227). En observant le comportement du terme “créance” dans des contextes plus amples, il est possible d’éclaircir la relation qui unit “créances” à “dette” et donc, de mettre en évidence le contenu référentiel des différentes “créances” auxquelles la description du surendettement fait allusion. Voici quelques exemples significatifs [le soulignement est le nôtre] qui illustrent bien la valeur ajoutée liée à l’exploration du corpus :

a) L'effacement partiel des créances, dont les dettes fiscales, combiné avec les mesures imposées. Les créances qui ont été payées par la caution ou le coobligé du débiteur ne peuvent faire l'objet d'un effacement. Les créances figurant dans l'état du passif définitivement arrêté par la commission ou le juge ne peuvent produire d'intérêts ni entraîner de pénalités de retard (article L331-7-1 du code de la consommation).

b) Dans les procédures de traitement du surendettement, les créances des dettes de loyer sont réglées prioritairement aux créances des établissements de crédit (article L333-1-1 du code de la consommation).

c) […] l'interdiction pour le débiteur de faire tout acte qui aggraverait son insolvabilité, de payer, en tout ou partie, une créance autre qu'alimentaire, dont les découverts bancaires, de désintéresser les cautions qui acquitteraient des créances nées antérieurement, […][17]

16 <http://www.cabinet-arc.com/creances.html> [consulté le : 27 juillet 2013].

Comme on peut le constater, les trois exemples permettent de mieux comprendre le contenu référentiel du concept de " créance". Toutefois, l'exemple b) présente un degré de difficulté d'interprétation plus élevé: la préposition "de" trace une relation entre dettes et loyer mais présente un contenu sémantique pauvre. Le repérage de la relation pertinente passe donc forcément à travers une procédure d'inférence et cela est encore une fois rendu possible par l'utilisation d'un corpus qui fournit des contextes élargis de matériel conceptuel. Les corpus matérialisent, sous forme d'expressions, les inférences cohérentes qui peuvent remplir la relation "de". La reprise anaphorique avec "créance" instaure à son tour une relation de type métonymique car il s'agit du droit qui dérive du fait de posséder un crédit ; ce qui nous autorise à l'utiliser en tant qu'expression anaphorique. Les créances impayées sont donc des droits qui n'ont pas été réalisés. La métonymie, nous le rappelons, est une figure qui consiste à remplacer un substantif par un autre qui entretient avec le premier un rapport de contiguïté. Plus précisément, les termes d'une relation métonymique représentent ou des rôles d'un procès (le lieu pour l'agent) ou un rôle et un procès (l'instrument pour l'action) ou deux procès (l'effet pour la cause). Les créances impayées se configurent donc comme des concepts qui ne sont pas saturés et dont les arguments se doivent d'être reconstruits à travers un procédé inférentiel. Le syntagme "créances impayées" est une anaphore dont le sens réside dans un antécédent qui n'est explicable que par le contexte global. En même temps, c'est également une cataphore car il introduit et canalise l'attente à l'égard de ses composantes. Du point de vue didactique, reconstruire les arguments du concept de "créances impayées" et approfondir la reprise anaphorique s'avère stratégique car cela remet en question la relation d'antonymie figurant dans les définitions respectives.

Un tout dernier exemple, toujours tiré du site Rachat du crédit, peut éclaircir ultérieurement le contenu sémantique du concept de "créances impayées" :

> Créances impayées
> Il est toujours possible de rembourser les dettes en vendant la voiture ou d'autres biens. Même si des particuliers ont des créances ou mensualités

> impayés chez des grandes surfaces, commerçants ou des banques…il est possible de se faire racheter toutes ses dettes par un établissement de crédit ou une société de rachat de crédit.

L'exemple confirme encore une fois que les deux concepts de " créance" et " dettes" ne sont pas antonymes mais plutôt caractérisés par une corrélation de conversion, en relation avec un troisième concept: ce qui est une dette pour le débiteur représente un crédit pour le prêteur (relation de conversion), ce dernier, suite au prêt, acquiert le droit d'exiger ce que nous appelons une dette, ou une créance. L'exploration directe des corpus permet donc de reconstruire le réseau complexe de relations et d'enrichir les définitions classiques des dictionnaires, nous confirmant que la notion d'antonymie est excessivement schématique.

Quant à la traduction en italien du syntagme "créances impayées", une vérification approfondie des sources fiables en italien repérées sur la base de mots-clés tels que, par exemple, "sovraindebitamento" et "composizione delle crisi" et de la définition même de "sopraindebitamento" fournie par le Décret-loi italien du 22 décembre 2011, nous suggèrent deux équivalents possibles, "crediti non soddisfatti" et le syntagme avec fonction d'hypéronyme "obbligazioni non soddisfatte", nettement moins fréquent mais que nous allons enregistrer, dans notre hypothétique fiche terminologique, parmi les variantes.

4. Conclusion

La didactique par corpus a des implications didactiques fortes. Son approche non dirigiste est formatrice par antonomase et va du texte authentique à la proposition authentique. L'approche guidée sur corpus, qui s'oppose à une approche dirigiste (*top-down*) basée sur l'utilisation de dictionnaires et de répertoires terminologiques sans contextes et que nous rejetons ici, favorise la prise de conscience du fonctionnement du lexique spécialisé. La prise de conscience de ses implications est essentielle à la compréhension globale du texte et,

dans notre cas particulier, du concept de surendettement et de la reprise anaphorique "créances impayées". Les corpus comparables indiquent donc d'une part que la fonction du contexte est différente en français et en italien. D'autre part, ils permettent d'extraire des variantes de traduction en contexte. La valeur ajoutée d'une analyse contrastive d'une base textuelle authentique par une approche guidée sur corpus permet en effet de répertorier et d'étudier des comportements langagiers et des stratégies de production du texte écrit différents et parallèlement de repérer des stratégies de traduction adéquates au texte cible, notamment dans le cas d'écarts lexico-sémantiques patents. Ainsi, l'exploration des différents corpus met en évidence l'anisomorphisme. Ce phénomène n'est pas perçu si l'on accorde une importance excessive au mot isolé et devient alors un obstacle à la compréhension et à la restitution du sens. En revanche, si nous explorons un contexte élargi, nous nous apercevons que cette différence s'amenuise et que la traduction est plus immédiate. C'est donc bien la traduction qui nous met dans l'obligation de déconstruire les concepts génériques. "Créance" et "crédit" sont deux signes qui représentent deux concepts. La tautologie fonctionne parfaitement en langue française : l'existence de deux signes suppose celle de deux concepts. Les langues naturelles possèdent des tautologies différentes et c'est bien dans l'acte de traduire que ces contradictions apparentes explosent et sont rendues comparables.

Références bibliographiques

Aristote. *Catégories*. Traduction de Pascale Nau : <http://www.documentacatholicaomnia.eu/03d/-384_-322,_Aristoteles,_Categories_%28Pascale_Nau%29,_FR.pdf>

Baker, Mona 1995. Corpora in Translation Studies : An Overview and some Suggestions for Future Research. *Target* 7 / 2 223-243.

Bernardini, Silvia / Stewart, Dominic / Zanettin, Federico 2003. *Corpora in Translator Education*. Manchester / Northampton : St. Jerome.

Bernardini, Silvia 2000. *I corpora nella didattica della traduzione. Corpus Use and Learning to Translate*. Bologna : CLUEB.

Cabré, Maria Teresa 1998. *La terminologie. Théorie, méthode et applications*, (adaptation française du volume La terminologia. La teoria, els mètodes, les aplicacions, 1992, Barcelone: Ed Empuries), Paris : Armand Colin; Ottawa : Les presses de l'Université d'Ottawa.

De Mauro, Tullio 2000. *Grande dizionario italiano dell'uso*, Torino : UTET.

Eco, Umberto 2003. *Dire quasi la stessa cosa. Esperienze di traduzione*. Milano : Bompiani.

Gaudin, François 2003. *Socioterminologie : une approche sociolinguistique de la terminologie*. Bruxelles : Duculot.

Gavioli, Laura / Zanettin, Federico 2000. I corpora bilingui nell'apprendimento della traduzione. Riflessioni su un'esperienza pedagogica. In Bernardini, Silvia *I corpora nella didattica della traduzione. Corpus Use and Learning to Translate*. Bologna : CLUEB, 61-80.

Gross, Gaston 2012. *Manuel d'analyse linguistique*. Villeneuve d'Ascq : Presses Universitaires du Septentrion.

Habert, Benoît / Nazarenko, Adeline / Salem, André 1997. *Les linguistiques de corpus*. Paris : Armand Colin / Masson.

Hédiard, Marie 2007. L'analyse guidée sur corpus : une approche "décloisonnante" pour accéder au sens. In Bellati, Giovanna *et al.* (eds) *Un paysage choisi, Mélanges de linguistique française offerts à Leo Schena.* Torino : L'Harmattan Italia, 225-238.

Le Grand Robert de la langue française 2001. Rey, Alain (éds), (2ème édition). Paris : Dictionnaires Le Robert.

Lyons, John (1977). *Semantics*. Cambridge : Cambridge University Press.

Magris, Marella / Musacchio, Maria Teresa (2009), Vaghezza e connotazioni: elementi di disturbo nelle definizioni terminologiche?. *MediAzioni 7*, <http://mediazioni.sitlec.unibo.it>

Meta 1998. *L'approche Basée sur le Corpus/The Corpus-based Approach*. Special Issue 43 / 4.

Olohan, Maeve 2004. *Introducing Corpora in Translation Studies*, London / New York : Routledge.

Palmer, Frank Robert 1976. *Semantics*. Cambridge : Cambridge University Press.

Prandi, Michele 2009. *Segni e termini : descrizione e normalizzazione*. In Terminologia, ricerca e formazione, Publif@rum 9 <http://www.publifarum.farum.it/ezine_articles.php?art_id=10>

Sinclair, John 1991. *Corpus, Concordance, Collocation*. Oxford : Oxford University Press.

Temmerman, Rita 2000. *Towards New Ways of Terminology Description.The sociocognitive approach*. Amsterdam / Philadelphia : John Benjamins.

TLF (Trésor de la langue française) 1978. Imbs, Paul (eds), Paris : Editions du centre national de la recherche scientifique.

Zgusta, Ladislav (éds) 1971. *Manual of Lexicography*. Academia, Prague : Publishing House of the Czechoslovak Academy of Sciences.

Sitographie

Banque de France :
<http://www.banque-france.fr/la-banque-de-france/missions/protection-du-consommateur/surendettement.html>

Gazzetta ufficiale :
<http://www.gazzettaufficiale.it/gunewsletter/dettaglio.jsp?service=1&datagu=2011-12-22&task=dettaglio&numgu=297&redaz=011G0255&tmstp=1325582228682>

BootCat : <http://bootcat.sslmit.unibo.it>

Cabinet ARC : <http://www.cabinet-arc.com/creances.html>

Centre national de ressources textuelles et lexicales : <http://www.cnrtl.fr/>
Decreto-legge 22 dicembre 2011 : <http://www.camera.it/_dati/leg16/lavori/stampati/html/decretolegge%5Carticoli%5Ccomponiarticoli.asp?CodiceBarre=16PDL0056790&NomeFile=16PDL0056790_Capo_I_Articolo_1_.html>
Dossier de surendettement en ligne : <http://www.rachatducredit.com/dossier-surendettement.html>
Ordine dei dottori commercialisti e degli esperti contabili di Roma : <http://www.odcec.roma.it/index.php?option=com_content&view=article&id=1469:crisi-da-sovraindebitamento&catid=18:comunicazioni-dallordine&Itemid=61INS>
Treccani.it : <http://www.treccani.it/vocabolario/credito>