

Linguistica e Filologia è inclusa in ERIH PLUS
(*European Reference Index for the Humanities and Social Sciences*)

Internet: <http://aisberg.unibg.it/handle/10446/6133>

I contributi contenuti nella rivista sono indicizzati nelle banche dati
Modern Language Association (MLA) International Bibliography
e Linguistics and Language Behaviour Abstracts (LLBA),
Directory of Open Access Journals (DOAJ) e Web of Science

Licenza *Creative Commons*:

This journal 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.



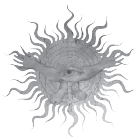
Volume pubblicato dal Dipartimento di Lingue, Letterature e
Culture Straniere e finanziato con fondi di Ateneo di ricerca.

ISSN: 1594-6517

Linguistica e Filologia

40

Dipartimento di Lingue, Letterature e Culture Straniere
UNIVERSITÀ DEGLI STUDI DI BERGAMO 2020



BERGAMO UNIVERSITY PRESS

sestante edizioni

Direzione della rivista

Giuliano Bernini, Università degli Studi di Bergamo, Direttore responsabile

Régine Delamotte, Université de Rouen

Klaus Düwel, Universität Göttingen

Edgar Radtke, Universität Heidelberg

Comitato editoriale

Maria Grazia Cammarota, Università degli Studi di Bergamo

Ada Valentini, Università degli Studi di Bergamo

Fulvio Ferrari, Università di Trento

Emilia Calaresu, Università di Modena e Reggio Emilia

Silvia Dal Negro, Libera Università di Bolzano

Maria Pavesi, Università di Pavia

Alessandro Zironi, Università di Bologna

Comitato Scientifico

Cecilia Andorno, Università di Torino

Alvise Andreose, Università e-Campus

David Ashurst, University of Durham

Sandra Benazzo, Université de Paris VIII

Gaetano Berruto, Università di Torino

Adriana Constăchescu, Universitatea din Craiova

Patrizia Giuliano, Università di Napoli ‘Federico II’

John McKinnell, University of Durham

Maria Grazia Saibene, Università di Pavia

Heidi Siller-Runggaldier, Universität Innsbruck

Andrea Trovesi, Università di Roma ‘La Sapienza’

Miriam Voghera, Università di Salerno

Marzena Wątarek, Université de Paris VIII

Maria Zaleska, Uniwersytet Warszawski

Lucia Avallone, Università degli Studi di Bergamo

Cécile Desoutter, Università degli Studi di Bergamo

Maria Gottardo, Università degli Studi di Bergamo

Dorothee Heller, Università degli Studi di Bergamo

Stefania Maci, Università degli Studi di Bergamo

Maria Chiara Pesenti, Università degli Studi di Bergamo

Comitato di Redazione

Jacopo Saturno, Università degli Studi di Bergamo

INDICE

Anna-Maria De Cesare, Begoña Sanromán Vilas <i>Restrictive focus adverbs in contemporary varieties of Italian and European Spanish: A contrastive, corpus-based study</i>	pag. 7
Laura Mori <i>La rappresentazione di scenari deontici e l'espressione della performatività nell'italiano delle leggi: dal diritto europeo alla legislazione nazionale</i>	» 45
Daniel Russo, Angela Andreani <i>Ogden's Basic English and its roots in the Early Modern English search for language simplicity</i>	» 99
Alessandro Zironi <i>Tracce e reminiscenze dei Goti in area ravennate tra il sesto e il decimo secolo: diversità e inclusione alla luce dei dati antroponimici</i>	» 129
Andrea Drocco, Neha Tiwari <i>A Pragmatic Approach to Compound Verbs in Hindi/Urdu: The Case of (Inter)subjectivity</i>	» 157
Mauro Maggi <i>Annotations on the Book of Zambasta, VIII</i>	» 197
<i>Elenco dei revisori per i numeri 37, 38 e 39</i>	» 223

ANNA-MARIA DE CESARE, BEGOÑA SANROMÁN VILAS*
(Universität Basel, Helsingin Yliopisto)

*Restrictive focus adverbs in contemporary varieties
of Italian and European Spanish:
A contrastive, corpus-based study*

This paper investigates the uses of Italian and Spanish restrictive focus adverbs, concentrating on the three cognates solo, solamente and soltanto / tan solo. Our aim is to highlight the discourse similarities and differences between these cognates in contemporary varieties of Italian and European Spanish through a contrastive corpus-based analysis. A further aim is to shed light on the factors that explain the differences from both an intra- and a cross-linguistic perspective. Our data is drawn from existing monolingual Italian and Spanish corpora as well as a self-assembled comparable corpus of academic texts written in Italian and Spanish.

Questo contributo indaga l'uso degli avverbi focalizzanti restrittivi in italiano e spagnolo, concentrandosi sulle forme solo, solamente e soltanto / tan solo. Il nostro obiettivo è quello di evidenziare le similitudini e le differenze nella distribuzione di questi avverbi nelle principali varietà contemporanee dell'italiano e dello spagnolo europeo attraverso un'analisi contrastiva, basata su corpora. Un ulteriore obiettivo consiste nel far luce sui fattori che spiegano le differenze riscontrate sia da un punto di vista intra- sia interlinguistico. I nostri dati sono tratti da corpora monolingui d'italiano e spagnolo già esistenti e da un corpus personale (costruito ad hoc) di testi accademici redatti in italiano e spagnolo.

1. Introduction

This contribution investigates the uses of Italian (It.) and Spanish (Sp.) restrictive (or exclusive) focus adverbs (henceforth RFA), starting from the prototypical adverb of the class: It. and Sp. *solo*.¹ Cross-linguistically,

* We would like to thank Davide Garassino for helping us with the statistical analyses of the data provided in Section 4 and Joan Nordlund for her careful proofreading of our contribution as a native speaker of English.

¹ In line with the current orthographic norm, we use the graphically non-accented form *solo* (also in *tan solo*) in the invented examples provided in this contribution. The reader should bear in mind that the graphically accented form (*sólo* / *tan sólo*) is also possible. Before 1959, the accented form (*sólo*) was used systematically to distinguish the adverb from the adjective *solo*, but since then, the orthographic norm has restricted the use of the accented form to contexts of ambiguity. Currently, new orthographic norms (RAE / ASALE 2010: 322) recommend the use of *solo* (without the accent) regardless of context ambiguities, which are assumed to be rare.

this Italian and Spanish adverb shares basic properties. As shown, in accordance with examples (1) and (2),² in both Romance languages *solo* is a simple adverb appearing in front of the syntactic constituent it modifies (in ex. 1, the subject), and it conveys the same meaning.

- (1) It. **Solo** Pietro lo sapeva.
- (2) Sp. **Solo** Pedro lo sabia.
'Only Paul knew it.'

It. and Sp. *solo* have two close equivalents, namely It. *solamente* and *soltanto* (Andorno 1999, 2000; Ricca 1999) and Sp. *solamente* and *tan solo* (Fuentes Rodríguez 2002; RAE / ASALE 2009: § 40.9). Italian and Spanish thus share a small set of similar adverbs to express restriction: *solo*, *solamente* (cross-linguistically equivalent both in form and meaning) and *soltanto* / *tan solo* (cross-linguistically similar in form and equivalent in meaning). As a consequence, given their similar morphological make-up and shared meaning component, considerable overlap is to be expected in the uses of It. and Sp. *solo*, It. and Sp. *solamente*, as well as It. and Sp. *soltanto* / *tan solo*. However, the Spanish translations of examples (3) and (4), provided in Alloa et al. (2008: 180), cast some doubt on this expectation, as we find that only in the first case is It. *solo* translated into Spanish as its closest equivalent form.

- (3) It. [...] aspettiamo che gli altri ci trovino, perché **solo** loro possono trovarci in tutto l'universo.
Sp. transl. [...] esperamos que los otros nos encuentren, porque **sólo** ellos pueden encontrarlos en todo el universo.
'[...] we will wait for the others to find us, because only they in the whole universe can find us.'
- (4) It. Noi ci siamo **soltanto** baciati.
Sp. transl. Nosostros **sólo** nos besamos.
'We only kissed each other.'

² Two remarks should be made in relation to how we present our examples. First, for the sake of clarity, we always highlight in bold the RFA or linguistic expression in which we are interested. However, this does not mean that the item in bold is necessarily interpreted as carrying prosodic and information emphasis. Second, we generally provide a working translation of the original Italian/Spanish example. For longer stretches of text, we used deepl.com software in the first step of the translation.

Italian and Spanish RFAs have never been the object of an in-depth contrastive study, neither intra- nor cross-linguistically.³ Consequently, it is not currently known what (if any) parameters regulate the use of *solo*, *solamente* and *soltanto* / *tan solo* in Italian and Spanish. Our aim in this study is to uncover the similarities and differences between these items. We believe that gaining knowledge about these issues is relevant in different domains of applied linguistics, including bilingual lexicography, language teaching and acquisition (Italian/Spanish as L2), as well as translation studies. To shed light on these questions we investigate the following discourse-related features: the distribution of *solo*, *solamente* and *soltanto* / *tan solo* in different language varieties, their discourse frequency, as well as their occurrence in specific collocations.

Our study is semasiological in approach, our starting point being the three linguistic cognates expressing restriction, and our goal being to understand their uses and functions. It is also based on a non-oriented contrastive analysis of Italian and Spanish, meaning that the description of each language is equally important and serves to illuminate the properties of the other (for details on this approach, see Tekin 2012: 134-141). As far as the data is concerned, we rely on empirical evidence drawn from several corpora representing contemporary varieties of Italian and European Spanish.

Our contribution is organized as follows: in the first step (Section 2) we present the main properties of the class of RFAs, paying special attention to the semantic features of It./Sp. *solo*, *solamente* and *soltanto* / *tan solo*, and point to some differences between these items as acknowledged in monolingual dictionaries. Second (Section 3), we describe the distribution of these adverbs in different language varieties (written vs. spoken) based on the results found in the main monolingual corpora available for Italian and Spanish. Because the available corpora are not directly comparable, in step three (Section 4) we present the outcome of an analysis based on

³ RFAs have been described almost exclusively from a monolingual perspective, and considering only one adverb, generally *solo*. In relation to Italian, see Geerts 1977 on *solo* and La Forgia / Carreras i Goicoechea 2008 on the complex expression *anche solo*. In relation to Spanish, see Fuentes Rodríguez 2002 on *solamente* and Medina Gómez / Alarcón Neve 2017 as well as Medina Gómez / Fernández García 2018 on *solo* and *solamente*. Mention should also be made of contrastive studies devoted to one of the Romance languages in which we are interested (see Nicklaus 2015 on German, Italian and French restrictive expressions in literary texts).

a self-created comparable corpus of Italian and Spanish (PACOR_IS), comprising academic texts. Specifically, we provide insights into the discourse distribution of It./Sp. *solo*, *solamente* and *soltanto* / *tan solo* and analyze two collocation patterns. In discussing these RFAs, we also consider the role played by near-synonyms such as It. *unicamente*, *esclusivamente* and Sp. *únicamente*, *exclusivamente* ‘only, exclusively’. In the conclusion (Section 5), we summarize our main findings and suggest some steps that could be taken in the future to further enhance understanding of the uses of RFAs across Italian and Spanish.

2. Italian and Spanish restrictive focus adverbs: an overview

2.1. The meaning of restrictive focus adverbs

It. and Sp. *solo*, *solamente* and *soltanto* / *tan solo* are restrictive focus adverbs (RFAs) and belong to a class including, *inter alia*, It. *esclusivamente* and Sp. *exclusivamente* ‘exclusively’.⁴ Their main semantic contribution to the sentence in which they occur is based on the cognitive operation of ‘restriction’,⁵ and can be described starting from examples (5) and (6).

- (5) It. Lo sapeva **solo/solamente/soltanto** Pietro. (nessun altro)
- (6) Sp. Lo sabía **solo/solamente/tan solo** Pedro. (nadie más)
‘Only Paul knew it. (nobody else)’

RFAs such as It. and Sp. *solo*, *solamente* and *soltanto* / *tan solo* quantify over a set of alternative values to the one expressed by the element on which they operate, called *domain of association*, in short DA (in examples 5 and 6, the DA is *Pietro* / *Pedro*). Specifically, they restrict the validity of the proposition in which they enter to the value of their

⁴ For details on the class of RFAs, see Ricca (1999) and Andorno (1999, 2000) on Italian; Kovacci (1999), RAE / ASALE (2009: § 40.9) and García Pérez (2013) on Spanish; König (1991) for a broad typological perspective and De Cesare (2015) for a micro-typological view (based on French, Italian, English and German).

⁵ Some studies describe the main semantic feature of the class in terms of *exclusion* and, consequently, label the items in which we are interested *exclusive focus adverbs* or *quantifiers* (see König 1981 and 1991, as well as Manzotti 1984: 52-54 for Italian and Sánchez López 1999, Ferrari et al. 2011 and Montoro del Arco 2011, 2012 for Spanish).

DA, such that ‘no value other than the one expressed by their DA’ yields to a proposition that is true. Consequently, in the same discourse context as (5) and (6), a proposition such as It. *Lo sapeva Maria* / Sp. *Lo sabía María* ‘Maria knew it’ can only be false.

All in all, there are two major conditions regulating the use of RFAs. First, the DA must coincide with a subset of a larger set of alternative values. This subset may include one value, as in (5) and (6), or more than one value, as in (7) and (8). Second, all the values (of the larger set) that are not part of the subset of values denoted by the adverb’s DA are to be excluded as possible DA candidates.

- (7) It. Pensavo di averlo detto a Maria, Paola e Pietro, ma lo sapevano **solo/solamente/soltanto** Maria e Pietro.
- (8) Sp. Pensé habérselo dicho a María, Paula y Pedro, pero lo sabían **solo/solamente/tan solo** María y Pedro.
‘I thought I told Maria, Paola and Pietro, but only Mary and Pietro knew about it.’

Apart from their quantificational meaning component, which is an inherent feature of all RFAs, adverbs such as It. / Sp. *solo, solamente* and *soltanto / tan solo* are also compatible with a scalar reading, licensed by lexical items occurring in the same sentence (such as a quantitative expression) and/or the discourse context. The adverbs in examples (9) and (10) restrict the validity of the proposition to their DA (*un genio* ‘a genius’), which is also the highest value ranked on a scale of intelligence. In this case, the adverbs exclude all the alternative values ranked below their DA (‘a genius’).

- (9) It. **Solo/Solamente/Soltanto** un genio lo poteva sapere!
- (10) Sp. ¡**Solo/Solamente/Tan solo** un genio podía saberlo!
‘Only a genius could know it.’

There are contexts in which *solo, solamente, soltanto / tan solo* could optionally be interpreted as scalar, in which case their DA is to be interpreted as denoting a high (or low) point on an ordered scale. Sentences such as (11) and (12) could be used to communicate either that no one else attended the meeting (in this case they are purely restrictive) or that no one with a higher rank did (the example is from RAE / ASALE

2009, § 40.9f). The set of alternatives excluded in both cases does not overlap: only in the first reading is the body of students excluded from the group of people who participated in the meeting.

- (11) It. A questa importante riunione ha partecipato **solo** il decano.
- (12) Sp. A esa importante reunión **solo** asistió el decano.
'Only the dean attended this important meeting'.

Although *solo* etc. are compatible with a scalar interpretation, other RFAs are not. This can be observed in (13) and (14), including It. [...] *esclusivamente* and Sp. *exclusivamente*, respectively.

- (13) It. ***Esclusivamente** un genio lo poteva sapere!
- (14) Sp. ***Exclusivamente** un genio lo podía saber!
Lit. Exclusively a genius could know it.
'Such knowledge is exclusive to geniuses.'

Mention should also be made to It. *unicamente* and Sp. *únicamente*. Although the former is claimed in the literature to be incompatible with a scalar reading (see Andorno 2000: 87), the authentic example in (15), found in a newspaper article, seems to indicate that it is in fact allowed in a context in which the value encoded by the adverb's DA is positioned on an ordered scale. Thus, with all due caution, we could suggest that It. *unicamente* is compatible with a scalar reading at least in some cases.

- (15) It. [...] Egli [Bramante], ancora oggi, sa dunque riunire le persone, ciò è in grado di fare, molto tempo dopo la sua morte, **unicamente** un genio. [<https://pesaronotizie.com/2014/08/02>]
'He [Bramante], still today, therefore, knows how to bring people together, which only a genius is able to do long after his death.'

Sp. *únicamente* is also compatible with a scalar reading, as examples (16) and (17) show. In (16), the adverb excludes other options than *un experto* 'an expert', a value occupying the highest rank of the ordered scale. As pointed out for the three adverbs under scrutiny (RAE / ASALE 2009: § 40.9e), a hierarchy is easier to perceive when the adverb's DA is a quantitative expression, as in (17).

- (16) Sp. Liberto Pech nunca se ha preguntado cuánto podría sacar en la calle por la *padparadscha*. **Únicamente** un experto apreciaría su valor. [CORPES XXI: Merino, Olga, (2004): *Espuelas de papel*. Madrid: Alfaguara]
 ‘Liberto Pech never asked himself how much he could get in the street for the *padparadscha*. Only an expert would appreciate its value.’
- (17) Sp. En enero de aquel mismo año la Junta Superior de Cataluña protestaba por haber recibido **únicamente**, en todo el curso de la guerra, 43 ó 44 millones de reales, cuando las necesidades mensuales del ejército eran de 11 millones de reales. [CORPES XXI: Canales Gili, Esteban. *Hispania Nova*. Madrid: hispanianova.rediris.com, 2003-01-01]
 ‘In January of that same year the Higher Board of Catalonia complained about having received only 43 or 44 million reales in the course of the entire war, even though the army’s monthly needs were 11 million reales.’

Given the semantic properties of Italian and Spanish RFAs discussed so far, namely restrictive quantification and scalarity, we propose a classification of these adverbs along the lines of Table 1. Both languages possess additional RFAs, such as It. *meramente*, *semplicemente* and *puramente* (on the last two forms, see Andorno 2000: 86-88) and Sp. *puramente*, *simplemente* and *sencillamente* (cf. Alloa et al. 2008: 180),⁶ as well as *nada más (que)* and *exclusive*. These forms are not considered in this study.

	Italian	Spanish
scalar compatible	solo/soltanto/solamente únicamente	solo/tan solo/solamente únicamente
not scalar compatible	esclusivamente	exclusivamente

Tab. 1. A semantic classification of restrictive focus adverbs in Italian and Spanish

⁶ Kovacci (1999: 773) refers to Sp. *puramente*, *simplemente* and *sencillamente* as exclusive focalizers (“focalizadores exclusivos”), whereas RAE / ASALE (2009: § 40.9) include these adverbs in the group of particularising adverbs (“adverbios particularizadores”), together with *precisamente*, *exactamente*, etc.

2.2. It. /Sp. solo, solamente, soltanto /tan solo: intra-linguistic variation

As shown in § 2.1, It./Sp. *solo*, *solamente* and *soltanto* / *tan solo* are semantically equivalent, as all adverbs are inherently restrictive and compatible with a scalar interpretation. In this section, we highlight further similarities and some differences between these items in Italian (§ 2.2.1) and Spanish (§ 2.2.2), primarily relying on the descriptions provided in monolingual dictionaries.

2.2.1. It. solo, solamente and soltanto

The semantic equivalence between It. *solo*, *solamente* and *soltanto* is clearly reflected in monolingual dictionaries of Italian. The three dictionaries we consulted (*DISC*, *NVDB*, *TRECCANI*), resort to synonyms to describe the adverbial meaning of *solo*, first and foremost *solamente* (*NVDB*) and *soltanto* (*DISC*), while using the same circular defining procedure – *mutatis mutandis* – for *solamente* and *soltanto*.

These dictionaries provide general information on the morphological properties of the three adverbs: *solo* is presented as a simple adverb (based on a single lexical morpheme, going back to the Latin *sōlu(m)*), *solamente* as a derived adverb (formed with the adjective *solo* to which the *-mente* suffix is attached), and *soltanto* as a compound adverb (based on the grammatical and orthographic conjunction of the morphemes *solo* and *tanto*). From a historical perspective, both *solo* and *solamente* were attested as RFAs as early as in the 13th century (*NVDB*). *Soltanto* seems to have arisen somewhat later, but no clear date is provided (*NVDB* just indicates ‘before 1375’).

It is not clear from the dictionaries we consulted whether there are differences in usage with reference to *solo*, *solamente* and *soltanto*, be it in terms of register, discourse distribution or collocations. Not only is the information on these aspects rather scarce, it is also partly contradictory and based on different choices in describing variation. According to one view, *solo*, *soltanto* and *solamente* belong to the *vocabolario fondamentale* ‘basic vocabulary’, in other words to the 2,000 most frequently used words in contemporary Italian (see *NVDB*), whereas according to another, *solo* is more familiar than both *solamente* and *soltanto* (*TRECCANI*).

As is evident from this succinct presentation, the information provided in monolingual dictionaries does not allow the highlighting of

any clear differences in the use of It. *solo*, *solamente* and *soltanto*. We are thus left to wonder how we can account for the polymorphism of Italian RFAs.

2.2.2. Sp. *solo*, *solamente* and *tan solo*

As acknowledged in § 2.1, Sp. *solo*, *solamente* and *tan solo* are semantically equivalent. Most of the Spanish monolingual dictionaries we consulted treat *solo* and *solamente* as absolute synonyms, using one of the adverbs within the definition of the other (see *LEMA*, *CLAVE*, *DUE* and *DLE*). In specialized lexicographical works, such as *DCOE* and *DP*, we also read that *solo* has the same values as *solamente*. In relation to their frequency, Davies's (2006) list of the 5,000 most commonly used words in Spanish ranks *solo* at 102 and *solamente* at 336.

Lexicographically, *tan solo* is not described in the same way as *solo* and *solamente*. None of the dictionaries we consulted have an independent entry for *tan solo*, which is mentioned in the entry of *solo* (see *DEA*, *DSLE* and *DUE*). *DEA* describes *tan solo* as an emphatic form of *solo* and *tan solamente* of *solamente*. In contradiction with this, *DSLE* claims that, unlike *solo*, *solamente* does not normally take the intensifier adverb *tan*. Finally, *DUE* equates *tan solo* with *solamente*, whereas *solamente* is circularly defined through *solo*.

From a morphological perspective, Sp. *solo* is described as a simple adverb, originating from the homophone adjective, and *solamente* as a *-mente* derived adverb. The item *tan solo* is considered here as a compound adverb on the grounds that it has a single meaning and is pronounced with a single accent.⁷

This short presentation of Sp. *solo*, *solamente* and *tan solo* as described in monolingual dictionaries reveals that there are several questions to clarify. One of these concerns the status of *tan solo* as a full-fledged adverb, as it is not recognized as such in any of the dictionaries (and studies) we consulted. More importantly, given that these dictionaries do not tackle any of these issues, we still lack a clear picture on the differences between the three Spanish RFAs.

⁷ This is one of the factors distinguishing the adverbial expression *tan solo* 'only', pronounced with a single accent [tan'solo], from the adjectival phrase *tan solo* 'so lonely', where the adverb *tan* is used to intensify the adjective *solo* and has two accents [tan'solo].

3. *It. / Sp. solo, solamente, soltanto / tan solo across contemporary language varieties: a contrastive analysis based on monolingual corpora*

To clarify the uses of *It. / Sp. solo, solamente, soltanto / tan solo*, we report the results of a contrastive, corpus-based analysis. The corpora we consulted (presented in § 3.1) give some idea of the distribution of the three RFAs in different varieties of Italian and Spanish. However, as we emphasize throughout this section (in particular in § 3.2), these results are not fully comparable because the main available corpora of contemporary Italian and European Spanish differ somewhat in design.⁸ Given that the corpora representing written and spoken Italian and Spanish are therefore not directly comparable, we add data from another corpus, DB-IPIC, which represents informal spoken communication (§ 3.3).

3.1. *Available corpora for comparing Italian and Spanish*

Table 2 shows the major monolingual corpora representing contemporary Italian and Spanish. As we point out below, these corpora are not fully comparable.

	Written	Spoken
Italian	CORIS (150 million words)	BADIP (490,000 words)
Spanish	CORPES XXI 90% written; 10% spoken (≈ 286 million words)	

Tab. 2. Monolingual Italian and Spanish corpora

One of the most important corpora representing Italian is CORIS, which includes commonly occurring written texts totaling 150 million words. Most of these texts were published in Italy in the 1980s and 1990s. The corpus comprises five subcorpora featuring five language varieties / genres, including journalistic texts (daily newspapers, periodicals and supplements), administrative-legal texts and academic texts. One of the

⁸ Mention should be made to the existence of a fully comparable multilingual corpus, named CONTRAST-IT (for a detailed description, see De Cesare 2019), which includes a subcorpus of Italian and Spanish written journalistic texts. The results based on CONTRAST-IT are not included in the present study because we did not find sufficient occurrences of some of the RFAs in which we were interested.

most commonly used corpora representing spoken Italian, in turn, is BADIP (based on De Mauro et al.'s LIP, 1993). This corpus, which includes 490,000 words, represents different communicative situations (both formal and informal) and four regional varieties of Italian (the data was collected in Milan, Rome, Naples, and Florence).

The main reference corpus for Spanish is CORPES XXI, which is pan-Hispanic and includes European and American varieties of Spanish covering the period 2001-2016. The data representing European Spanish is less extensive than that representing American Spanish: of the total 286 million words, 86 million are from European Spanish (30%), the rest (70%) represent non-European (mostly American) written and spoken varieties. The CORPES XXI website points out that 90% of the total 286 million words (including European and American Spanish) represent the written language, featuring different genres (journalistic, academic, legal), thus only 10% represents spoken language. Unfortunately, there is no indication whether this proportion (90% written and 10% spoken texts) is the same for every diatopic variety of Spanish represented in the corpus. If we assume that this proportion also holds for European Spanish, that is for the language variety we are analyzing in this study, then 90% of the 86 million words represent the written language (roughly 77 million words) and 10% represent the spoken language (roughly 8 million words). We should point out that we are not certain about these figures. In fact, as our results indicate, the data obtained for spoken Spanish casts some serious doubts on this assumption. Another point worth making about the data on spoken European Spanish in CORPES XXI is that it is biased towards formal registers.

In light of their corpus design, we can highlight several differences between the Italian and Spanish corpora indicating that the sets of data they represent are not fully comparable. Crucially, these corpora (and subcorpora) are not comparable in terms of size. Given that the size of a corpus plays an important role in assessing the distribution of *solo*, *solamente*, *soltanto* / *tan solo* in different contemporary varieties of Italian and Spanish, in Table 3 we specify relevant information on the data used in § 3.2 to compare the three RFAs cross-linguistically.⁹

⁹ As we show below (§ 3.2), CORPES XXI also suffers from a series of technical shortcomings, which are pointed out, e.g., in Corpas Pastor (2017: 33-34): “[T]he CORPES in-built corpus system is rather unstable and slow in terms of processing, data downloading is not possible and access to the data is not flexible enough.”

	Written	Spoken
Italian	CORIS (150 million words)	BADIP (490,000 words)
European Spanish	CORPES XXI (77 million words)	CORPES XXI (8 million words, mainly from formal registers)

Tab. 3. Monolingual corpora used to compare Italian and Spanish

3.2. *It. / Sp. solo, solamente, soltanto / tan solo across written and spoken language varieties*

Table 4 reports the normalized frequency (per 100,000 words) of *It. / Sp. solo, solamente, soltanto / tan solo* in data representing written and spoken language varieties, respectively.

		Written	Spoken
Italian	<i>solo</i>	110	118
	<i>soltanto</i>	26	43
	<i>solamente</i>	2	11
	TOT.	138	172
Spanish	<i>solo</i>	119	1.8
	<i>tan solo</i>	6.6	0.8
	<i>solamente</i>	4.1	0.06
	TOTAL	129.7	2.66

Tab. 4. The frequency of restrictive focus adverbs in written and spoken Italian and Spanish
(normalized frequency: N. occurrence/100,000 words)

First, we look at Table 4 vertically, considering the distribution of the three adverbs intra-linguistically. *Solo* is by far the most frequent item in both language varieties of Italian and Spanish (*viz.* written and spoken), with a similar frequency in the data representing spoken Italian as was calculated for written texts (118 vs. 110). This result is, of course, largely expected given that *solo* belongs to basic vocabulary, meaning that it is among the 2,000 most frequently used words in present-day Italian (see § 2.2.1). The second most commonly used form in Italian, irrespective of the language variety, is *soltanto* (which turns out to be more frequent in the spoken than in the written language: 43 vs. 26), and the least is *solamente* (again, clearly more common in spoken than in written Italian: 11 vs. 2). We observed a similar pattern in Spanish. Again, regardless of the language variety, the second most commonly used form is *tan solo* and the third is *solamente* (6.6 vs. 4 in written Spanish and 0.8 vs. 0.06 in spoken Spanish). With regard to spoken Spanish, the exceptionally low figures obtained in CORPES XXI for all three restrictive adverbs should be considered more critically. These figures, in particular the one related to *solo*, lead us to believe that the subcorpus representing spoken European Spanish is probably smaller than we assumed (*i.e.* it does not include 8 million words and cover 10% of the data, as mentioned in § 3.1).

Let us now turn to the cross-linguistic frequency of each cognate pair. As far as It. and Sp. *solo* are concerned, we observe a similar distribution in written texts (110 vs. 119); for the reason mentioned above, we do not make a comparison between them in oral communication. With regard to It. and Sp. *solamente*, the adverb is rarely used, but its distribution seems to differ in the two languages: it is twice as frequent in Spanish as in Italian (4 vs. 2 occ.). The extremely low figure for spoken Spanish does not allow for sound conclusions: it is improbable that the gap between It. and Sp. *solamente* is so wide. Finally, in the case of *soltanto* / *tan solo* we observe the reverse pattern. The adverb is more frequent in written Italian than in written Spanish (the gap is quite wide here: 26 vs. 6.6 occ.), whereas the low data in the CORPES XXI inflates the cross-linguistic differences such that the result is unrealistic.

All in all, it turns out that the distribution of It. and Sp. *solo* is similar. It is the most frequently used adverb in both languages and appears to have a similar frequency (we do not take the figure related to spoken Spanish into account). Conversely, there are both similarities and differences between the adverbs belonging to the other two pairs. In order

of preference, It. *soltanto* and Sp. *tan solo* are the second most commonly used adverbs, followed by *solamente*. However, It. *soltanto* is more frequent than Sp. *tan solo*, and the reverse is true for *solamente*.

3.3. It. / Sp. solo, solamente, soltanto / tan solo in informal spoken language (DB-IPIC)

To better understand how sensitive these adverbs are to register variation (i.e. to the degree of formality of the discourses in which they occur), let us now consider additional data, taken from the DB-IPIC corpus. This is a small comparable corpus representing informal spoken language. The figures for each restrictive adverb are reported in Table 5. Note that we normalized the frequency of each adverb to 10,000 words (whereas Table 4 includes normalized frequencies to 100,000 words).

	DB-IPIC_It (32,500 words)		DB-IPIC_Sp (40,500 words)
<i>solo</i>	6.5 (21)	<i>solo/sólo</i>	5.2 (21: 18/3)
<i>soltanto</i>	2.2 (7)	<i>tan solo</i>	0
<i>solamente</i>	3.4 (11)	<i>solamente</i>	2 (8)
TOTAL	11	TOTAL	7

Tab. 5. The frequencies of restrictive focus adverbs in informal spoken Italian and Spanish (normalized frequencies: N. occurrence/10,000 words; absolute frequency in parentheses)

The order of preference of the three RFAs is again similar in informal spoken Italian and Spanish. Unsurprisingly, the most frequent adverb is *solo*, followed by *solamente* and It. *soltanto* / Sp. *tan solo* (the latter form does not appear at all). Table 5 reveals other cross-linguistic similarities. The frequencies of It. and Sp. *solo* (6.5 vs. 5.2), and of It. and Sp. *solamente* to a lesser extent (3.4 vs. 2), are quite close. Interestingly, *solamente* is preferred over *soltanto* / *tan solo* in informal spoken Italian and Spanish, which is in clear contrast with the results described in § 3.2.

We also observe from Table 5 that, although *solo* is the most frequently used adverb cross-linguistically, the difference between this and the other two restrictive expressions is not as marked as in written Italian and Spanish (we do not consider spoken language). Even if Italian *solo* (6.5)

is twice as frequent as *solamente* (3.4), which in turn is almost twice as frequent as *soltanto* (2.2), the differences between *solo* and the other two adverbs are not as large in informal spoken Italian as in written texts (adapting the data to match a relative frequency of 10,000 words, we arrive at 11 vs. 0.2 and 2.6). The same is true for Spanish: *solo* (5.2) is twice as frequent as *solamente* (2), but the difference in frequency between both forms is not as large as in written texts (12 vs. 0.4).

3.4. *It. / Sp. solo, solamente, soltanto / tan solo across contemporary language varieties: a discussion of the corpus-based results*

The results presented in §§ 3.2 and 3.3 shed some light on the polymorphism of Italian and Spanish RFAs, in particular on some general distributional properties of It. and Sp. *solo*, *solamente* and *soltanto / tan solo*. The picture that arises is, understandably, much more nuanced than the one based on lexicographical sources (§ 2.2). The three RFAs show strong similarities and differences in usage across the language varieties.

The adverb *solo* is by far the most frequently used form in Italian and Spanish in all language varieties. Its frequency is relatively stable across the written and spoken language (confirmation for spoken Spanish is needed here). The stable distribution is, of course, expected because this adverb belongs to basic contemporary Italian and Spanish, and it is known that the basic vocabulary of a language tends not to vary across varieties and genres (Chiari / De Mauro 2012: 23). Therefore, the lower frequency rate of It. and Sp. *solo* in informal spoken communication, with only half of the occurrences, is rather puzzling. This finding seems to imply that *solo* is sensitive to extreme register variation and leads us to hypothesize that there might be other factors explaining its distribution in informal language. The lower frequency in the data representing informal spoken Italian could be related to the somewhat high frequency of *solamente* (65 and 34 occ. / 100,000 words).¹⁰

The distribution of the other two pairs of adverbs is more complex and subject to variation. It turns out that *soltanto / tan solo* is the second most preferred option in both written and spoken Italian and Spanish, and that *solamente* is the adverb that is used the least. Two points are worthy of note in relation to these pairs of adverbs. The first is that *soltanto / tan*

¹⁰ Another parameter that could play a role in explaining the data, at least in relation to Italian, is the strong bias of the DB-IPIC towards Florentine (a diatopic variety of Italian).

solo differ a great deal in their frequency: in Italian, the adverb *soltanto* is a vital form in all three language varieties investigated (i.e. written, spoken and informal spoken), whereas in Spanish, *tan solo* is rather rare. The second point is the higher frequency of *solamente* with respect to *soltanto* / *tan solo* in informal spoken Italian and Spanish. This finding is surprising given that *-mente*-derived adverbs are considered typical of written and formal language varieties (although Hummel 2013: 26 points out that *solamente* has popular origins). In fact, we observed the exact opposite in our data: the adverb *solamente* is much more frequent in informal spoken than in written Italian and Spanish. In the case of Italian, there is a steady increase in the frequency of *solamente* across the three language varieties: written Italian, spoken Italian and informal spoken Italian (2, 11, and 34 occurrences / 100,000 words, respectively). The same pattern occurs in two varieties of the Spanish data (again, due to problems with the data, we cannot say anything about spoken Spanish): compared to its frequency in written texts, *solamente* is five times more frequent in informal spoken Spanish (4 vs. 20 occ. / 100,000 words).

4. It. / Sp. *solo*, *solamente*, *soltanto* / *tan solo* in contemporary academic texts: a contrastive analysis based on the comparable corpus PACOR_IS

In the last section of this contribution, we investigate the discourse distribution of It. / Sp. *solo*, *solamente*, *soltanto* / *tan solo* in more depth, considering their frequency, function and use in a single text type representing academic writing. First, we describe the empirical data used, which is drawn from a self-assembled comparable corpus called PACOR_IS (§ 4.1), then we report the discourse distribution of these forms in the data (§ 4.2), specifically with regard to their frequency and text dispersion. Finally, we describe the uses and function of the three adverbs in specific collocational patterns (§ 4.3).

4.1. PACOR_IS: a self-assembled comparable corpus of Italian and Spanish

PACOR_IS (an acronym standing for *Parallel Academic CORpus_Italian_Spanish*) is a corpus including two subcorpora of comparable academic texts written in Italian and Spanish.¹¹ The comparability of the two

¹¹ This (privately owned) corpus was created in 2018-2019 in joint collaboration between the two authors of this paper.

subcorpora can be observed at different levels of their design. First, both parts comprise academic articles belonging to the same fields of study: Italian Linguistics, Spanish Linguistics and, more generally, Romance Philology. Second, they both include the same number of texts, namely 100 articles, all of which were published in the same timeframe, from the year 2000 onwards. To ensure stylistic variation and to avoid analyzing idiosyncratic uses of *solo*, *solamente* and *soltanto / tan solo* in Italian and Spanish, we selected articles written by different authors. Moreover, to ensure that the texts represented written European Spanish, we restricted the selection to articles written by authors affiliated to a European University.¹²

Table 6 lists the journals used to create the two subcorpora of PACOR_IS and specifies the number of articles selected in each journal.

PACOR_ITALIAN (100 articles)	PACOR_SPANISH (100 articles)
<i>Italiano LinguaDue</i> (25)	<i>Revista de Filología Románica</i> (31)
<i>Studi Linguistici e Filologici Online</i> (21)	<i>Revista Electrónica de Didáctica del Español como Lengua Extranjera</i> (21)
<i>Linguistica e Filologia</i> (20)	<i>Cuadernos de filología italiana</i> (9)
<i>Cuadernos de filología italiana</i> (10)	<i>Revista de Filología Española</i> (8)
<i>E-JournALL. EuroAmerican Journal of Applied Linguistics and Languages</i> (9)	<i>Estudios Románicos</i> (7)
<i>Romanische Forschungen</i> (4)	<i>E-JournALL. EuroAmerican Journal of Applied Linguistics and Languages</i> (5)
<i>Revue Romane (open access until 2003)</i> (3)	<i>Linguistik online</i> (5)
<i>Estudios Románicos</i> (3)	<i>Revue Romane (open access until 2003)</i> (4)
<i>Chimera. Romance Corpora and Linguistic Studies</i> (1)	<i>Romanische Forschungen</i> (4)
<i>Lettere italiane</i> (1)	<i>Chimera. Romance Corpora and Linguistic Studies</i> (3)
<i>Revue de linguistique romane électronique</i> (2)	<i>Revue de linguistique romane électronique</i> (2)
<i>Linguistik online</i> (1)	<i>Revista Electrónica del Lenguaje</i> (1)

Tab. 6. Sources of PACOR_IS (Journals and numbers of articles selected)

All the articles have been uploaded on the Sketch Engine platform (for details on this electronic resource, see Kilgarriff et al. 2014), compiled by

¹² This, of course, is not a sufficient criterion to exclude non-European Spanish authors working in Europe. However, as we know most of the authors personally, we can claim that the Spanish texts are – to the best of our knowledge – written by Spanish authors and represent written academic European Spanish.

the software available for Italian and Spanish, respectively. Table 7 shows the resulting properties of the compiled PACOR_IS corpus: although both subparts include the same number of articles (100), the Italian part (i.e. PACOR_I) is slightly larger in terms of the numbers of words. In order to compare the data drawn from the two sections of PACOR_IS (as well as the other corpora considered in this study), we will present our results in the form of normalized frequencies (in the form: N. of occurrences / 100,000 words).

	PACOR_I	PACOR_S
Words	1,060,320	883,937
Sentences	41,591	52,002
Articles	100	100

Tab. 7. PACOR_IS corpus design (based on Sketch Engine counts)

4.2. *It. / Sp. solo, solamente, soltanto / tan solo: frequency and dispersion*

In this paragraph we describe two distributional features of *It. / Sp. solo, solamente* and *soltanto / tan solo* in the PACOR_IS corpus: general frequency (§ 4.2.1) and text dispersion (§ 4.2.2).

4.2.1. *Corpus frequency*

Table 8 reports the frequency of *It. / Sp. solo, solamente* and *soltanto / tan solo* in PACOR_IS. According to the Pearson's Chi-squared test, the difference in the distribution of these adverbs in the Italian and Spanish data is statistically significant (Chi = 95.54, df = 2, p-value < 2.2e-16 [p < .0001]).

	PACOR_I		PACOR_S
<i>solo</i>	120 (1209)	<i>solo/sólo</i>	83 (736)
<i>soltanto</i>	20 (205)	<i>tan solo</i>	3 (25)
<i>solamente</i>	5 / 3* (49 / 32*)	<i>solamente</i>	8 / 7* (73 / 68*)
TOTAL	143	TOTAL	93

Tab. 8. Normalized (and absolute) frequency of restrictive focus adverbs in academic Italian and Spanish; the figure followed by an asterisk is the data analyzed¹³

¹³ The data related to *It. and Sp. solamente* has been cleaned: we excluded all instances in which the adverb does not occur in the text written by the author of the article, but is found in stretches of text quoted from other authors, in examples belonging to other language varieties (such as oral

In line with the main results obtained from the other corpora we consulted (see Table 4 in § 3.2), the most frequently occurring form is It. *solo* and Sp. *solo/sólo*, followed by *soltanto* and *solamente* in Italian and *solamente* and *tan solo* in Spanish. The order of preference of the three RFAs mirrors the one found earlier in relation to written Italian, but not for written Spanish, in which *tan solo* occurs slightly more often than *solamente*. In contrast, *solamente* is roughly three times more frequent than *tan solo* in PACOR_S. Moreover, It. *solo* is also more frequent in PACOR_I than the equivalent form in Spanish (120 vs. 83 occ. / 100,000 words). As far as the two other restrictive adverbs are concerned, It. *soltanto* is much more frequent than Sp. *tan solo*, which occurs only marginally in the data (20 vs. 3). In turn, and in line with what we previously observed in relation to written texts, Sp. *solamente* is twice as frequent as its equivalent form in Italian (7 vs. 3)¹⁴.

A comparison of the frequency of the three RFAs in the Italian and Spanish data in the form of a Cohen-Friendly association plot (Figure A¹⁵) allows to confirm some of the claims made on the basis of Table 8. This plot shows very clearly that *soltanto* and *solamente* are overrepresented in the Italian and Spanish data, respectively. By contrast, *solamente* in Italian and *tan solo* in Spanish are underrepresented in the data. The most significant results (shown by the size of the boxes and the shade of their color) are the ones obtained, in order of importance, for Sp. *tan solo*, It. *soltanto* and Sp. *solamente*. Interestingly, the figures associated to It. *solo* and Sp. *solo/sólo* (see the small grey boxes) are not very important in explaining the overall result of the test (the values of the residuals for each adverb are close to 0; for details, see the bar to the right).

speech), or in texts belonging to other historical periods of Italian and Spanish. Given the large number of occurrences, we did not follow this procedure for *solo* and *soltanto*. Consequently, the data for these two forms is not as ‘clean’.

¹⁴ These results are based on a thorough check of the corpus and the correction of a series of shortcomings in the Sketch Engine POS-tagging. Some adverbial uses of It. / Sp. *solo* (without an accent) end up being POS-tagged as ‘adjective’: whereas there are only two cases (out of 79 occ. of *solo* as ADJ) in PACOR_I, there are as many as 213 in PACOR_S (out of a total of 347 occ. of *solo* tagged as ADJ). The high margin of error may be related to the lack of accent on Sp. *solo* (see Fn 1). Conversely, some adjectival uses of *solo* are wrongly labeled ‘adverb’ in the corpus: there are 24 cases (out of 1,209 occ. of *solo* as ADV) in PACOR_I and only four (out of 522) in PACOR_S.

¹⁵ Based on the assumption that the variables analyzed are statistically independent, a Cohen-Friendly association plot (based on Pearson’s residuals) allows us to visually display the deviations between the expected frequency and the actual frequency observed in the data. Specifically, it helps us understand the contribution of each cell to the overall result of the test. When observed frequencies

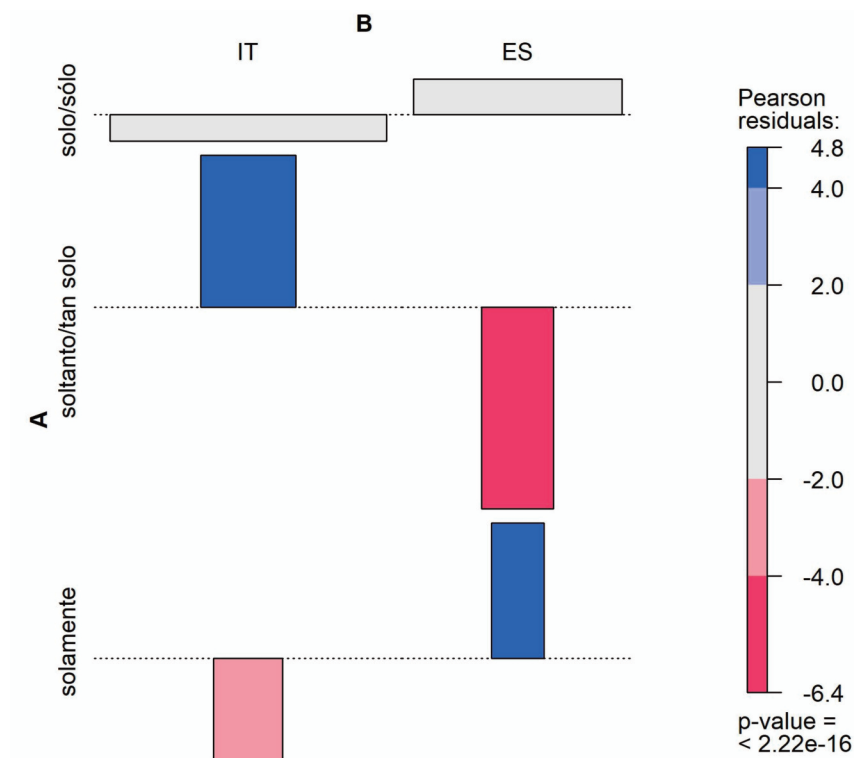


Fig. A. Comparing the distribution of restrictive focus adverbs in academic Italian and Spanish (Cohen-Friendly association plot)

In relation to the figures in the last row of Table 8, we observe another substantial difference between the two Romance languages: overall, Italian uses the three restrictive adverbs more often than Spanish (143 vs. 93). This result is primarily attributable to the large discrepancy in the frequency of It. and Sp. *solo*, which is somewhat puzzling in that it does not mirror what we found in corpora representing written Italian and

are greater than expected (i.e., when residuals are positive), the boxes rise above the baseline. Conversely, when observed frequencies are smaller than expected (in case of negative residuals), the boxes go below the baseline. The size of the box provides clues on the importance of the deviation of each result (or cell) and the shading on the level of significance.

Spanish (§ 3.2.), namely that the frequency of *solo* was very close (110 vs. 119). Explaining the different results (in Table 8) is not, of course, an easy task. Is restriction expressed in Spanish academic texts by drawing on other linguistic means, in particular other adverbials or lexical expressions (such as verbs)?¹⁶ To throw some light on this issue, in Table 9 below we report the frequency of two other RFAs in the PACOR_IS corpus, namely It. *unicamente* and *esclusivamente* and Sp. *únicamente* and *exclusivamente* (the results of the Pearson's Chi-squared test show that the difference in the distribution of these adverbs in the Italian and Spanish data is statistically significant: $\chi = 81.203$, $df = 1$, $p\text{-value} < 2.2e\text{-}16$ [$< .0001$]).

	PACOR_I		PACOR_S
<i>unicamente</i>	1.8 (19)	<i>únicamente</i>	10 (89)
<i>esclusivamente</i>	12 (126)	<i>exclusivamente</i>	5 (45)
TOTAL	13.8 (145)	TOTAL	15 (134)

Tab. 9. Normalized (and absolute) frequency of restrictive focus adverbs in academic Italian and Spanish

The numbers in Table 9 do not throw much light on the discrepancy between It. and Sp. *solo* in PACOR_IS, as the overall frequency of It. *unicamente* and *esclusivamente* on the one hand and Sp. *únicamente* and *exclusivamente* on the other is fairly close (13.8 for Italian and 15 for Spanish). What is interesting, however, is the cross-linguistic preference for a different adverb: Italian clearly favors *esclusivamente* over *unicamente* (the former is six times more frequent than the latter), whereas Spanish prefers *únicamente* over *exclusivamente* (the former is twice as frequent as the latter). Cross-linguistically, we observe that Sp. *únicamente* is roughly five times more frequent than It. *unicamente*, and that It. *esclusivamente* is almost three times more frequent than Sp. *exclusivamente*.

¹⁶ Syntactic constructions encoding restriction, in particular cleft sentences, seem less likely to play a role. We know from other contrastive corpus-based studies (see, in particular, De Cesare *et al.* 2016: 217-261) that Spanish uses cleft sentences, i.e. a group of syntactic structures associated with an exhaustiveness component (as in *Es Stella quien se ha comido la tarta* 'It's Stella who ate the cake'), less often than Italian (*È Stella che mangia la torta*). We found twice as many cleft sentences in Italian in a comparable corpus of news drawn from online daily newspapers (De Cesare *et al.* 2016: 228). There are, respectively, 45 and 21 Italian and Spanish clefts in 100,000 words.

A comparison of the frequency of the two RFAs in the Italian and Spanish data in the form of a Cohen-Friendly association plot (Figure B) allows to confirm the preferences highlighted on the basis of Table 9 across the two Romance languages. This plot shows very clearly that the use of *esclusivamente* and *unicamente* is overrepresented in the Italian and Spanish data, respectively. Conversely, we observe that the use of *unicamente* in Italian and *esclusivamente* in Spanish is underrepresented in the data.

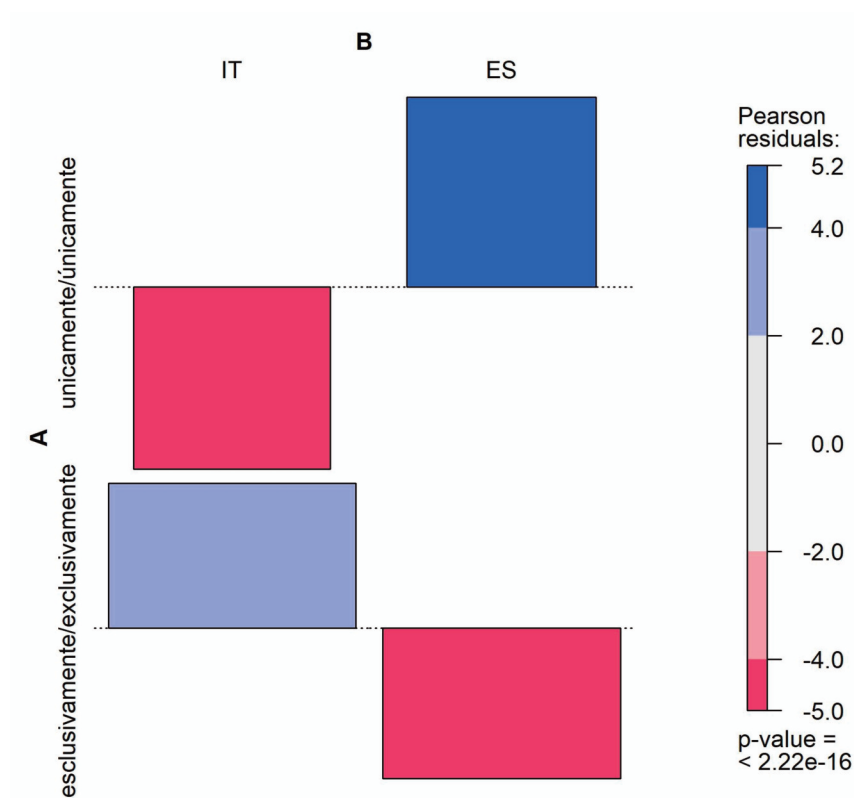


Fig. B. Comparing the distribution of restrictive focus adverbs in academic Italian and Spanish (Cohen-Friendly association plot)

4.2.2. *It. / Sp. solo, solamente, soltanto / tan solo: dispersion*

To shed more light on the distribution of *It. / Sp. solo, solamente* and *soltanto / tan solo* in the PACOR_IS corpus, let us consider their dispersion, i.e. their distribution across the corpus parts (see the data in Table 10).¹⁷

		n. of occ.	n. of documents	Average occurrences per document
PACOR_I	<i>solo</i>	1209	98	12
	<i>soltanto</i>	205	53	3.9
	<i>solamente</i>	32	18	1.8
PACOR_S	<i>solo</i>	736	92	8
	<i>tan solo</i>	25	10	2.5
	<i>solamente</i>	68	29	2.4

Tab. 10. The dispersion of three restrictive focus adverbs in academic Italian and Spanish (PACOR_IS)

It. and Sp. solo are found in almost every document in the PACOR_IS corpus (it is missing in only two documents in PACOR_I and eight in PACOR_S). However, the frequency per document of *It. solo* is, on average, higher than that of its Spanish counterpart (12 vs. 8), a result that is directly related to the higher frequency of *solo* in Italian (see Table 8). Turning to the other two restrictive adverbs, we observe that *It. soltanto* is present in 50% of the documents and *tan solo* in only 10%, whereas *It. solamente* occurs in 20% of them and *Sp. solamente* in 30%. The differences in distribution between *soltanto* and *tan solo* are reflected in part in the average number of occurrences of the two adverbs per document (3.9 vs. 2.5).

These results support the idea that *It. and Sp. solo* is the most basic, or unmarked restrictive focus adverb. They also confirm that *It. soltanto* is a rather common form, being used less frequently than *solo* but more often

¹⁷ These results should be interpreted with caution, as PACOR_I is larger than PACOR_S and the texts included in each subcorpus are not of the same length.

than *solamente*, that Sp. *solamente* is the second most commonly used form, and that the use of *tan solo* is rather restricted.

The difference between It. *soltanto* and *solamente* on the one hand, and Sp. *tan solo* and *solamente* on the other is further illustrated in the data presented in Table 11. (The data in this table reads as follows: It. *soltanto* occurs once in 14 documents, twice in six documents, three times in 12 documents, and so on).

	n. occ. / document	1	2	3	4	5	6	7	8	9	20
		Documents (N)									
PACOR_I	<i>soltanto</i>	14	6	12	5	6	4	2	2	1	1
	<i>solamente</i>	9	6	1	2	0	0	0	0	0	0
PACOR_S	<i>tan solo</i>	8	2	0	0	0	0	0	0	0	0
	<i>solamente</i>	16	4	4	3	1	1	1	0	1	0

Tab. 11. The dispersion of two restrictive focus adverbs in academic Italian and Spanish

As the results for PACOR_I show, It. *soltanto* occurs up to 20 times in a single article; It. *solamente* occurs a maximum of four times in the same document, but only once in most of the documents in which it does appear. We thus observe again that *solamente* is much more restricted in use than *soltanto*. The reverse pattern is evident in PACOR_S: *solamente* occurs up to nine times in the same document, and *tan solo* only twice. The index of dispersion confirms the cross-linguistic difference between It. and Sp. *solamente*: 0.55 for Italian and 0.88 for Spanish.

4.3. It. / Sp. solo, solamente, soltanto / tan solo: two semantic operations

Having considered the discourse frequency of It. / Sp. *solo*, *solamente* and *soltanto* / *tan solo* in the PACOR_IS, we now turn to their uses and function in academic texts. We describe two syntactic patterns related to two different semantic operations in more detail: negating (§ 4.3.1) and reinforcing (§ 4.3.2) the restrictive meaning component. We show that It. / Sp. *solo*, *solamente* and *soltanto* / *tan solo* do not

accomplish these operations in the same way across the two Romance languages, which points to the divergent restrictions imposed on the discourse context.

4.3.1. *Negating restrictive focus adverbs*

A frequent pattern in PACOR_IS is the occurrence of RFAs after negation, namely It. *non* and Sp. *no*. Table 12 shows the frequency of the collocational pattern ‘Neg. + RFA’, in which the adverb directly follows the negation.¹⁸ The data reads as follows (taking the first line as example): It. *non solo* occurs 261 times in the corpus (absolute frequency); it is a pattern that occurs in 22% of all instances of *solo* found in the corpus; and in relation to the data on all three patterns (i.e. *non solo*, *non soltanto* and *non solamente*, right-hand column), it covers 85% of the cases. The results of the Pearson’s Chi-squared test show that the difference in the distribution of these adverbs in the Italian and Spanish data is statistically significant (Chi = 50.521, df = 2, p-value = 1.07e-11 [p < .0001]).

	Collocations	n. of occ. (and %)	Total relative frequency (in %)
PACOR_I	<i>non solo</i>	261 (22)	85
	<i>non soltanto</i>	45 (22)	15
	<i>non solamente</i>	0	0
PACOR_S	<i>no solo</i>	192 (26)	94
	<i>no solamente</i>	13 (19)	6
	<i>no tan solo</i>	0	0

Tab. 12. ‘Negation + restrictive focus adverbs in Italian and Spanish (PACOR_IS)

The It. negation *non* is a stable lexical item before two RFAs, namely *solo* and *soltanto* (in both cases, the pattern ‘Neg. + RFA’ covers as much as a fifth of the occurrences), but it never occurs immediately before

¹⁸ The analysis does not include cases in which the negation is not adjacent to the RFA (e.g., *non è solo x* lit. ‘not is only x’).

solamente. The T-score¹⁹ associated with the first element before the adverb confirms this finding and allows us to claim that the chance of finding *non* before *solo* is greater than before *soltanto* (T-score: *non solo* = 15.49/*Non solo* = 2.69; *non soltanto* = 6.57).

The Sp. negation *no* is also a stable lexical item before two RFAs (T-score: *no solo* = 6.82/*no sólo* = 7.98; *no solamente* = 3.51), in this case *solo* and *solamente* (occurring with roughly the same frequency in relation to each other and to Italian). However, it was not found at all before *tan solo*. Cross-linguistically, these results highlight a strong similarity in the use of It. and Sp. *solo*. They also reinforce the claim that It. *soltanto* is closer to Sp. *solamente* than to *tan solo*, once more revealing the discrepant behavior of cross-linguistic cognates: It. *soltanto* and Sp. *tan solo* on the one hand, and It. / Sp. *solamente*, on the other.

A comparison of the frequency of the pattern ‘Neg. + RFA’ in the Italian and Spanish data in the form of a Cohen-Friendly association plot (Figure C; see next page) allows to show that *non soltanto* and, to a lesser extent, *no solamente* are overrepresented in the Italian and Spanish data, respectively. These results thus confirm that It. *soltanto* is close to Sp. *solamente*. Conversely, we observe that *non solamente* in Italian and, even more clearly, of *no tan solo* in Spanish are underrepresented in the data. The most significant results (shown again by the size of the boxes and the shade of their color) are the ones obtained, in order of importance, for Sp. *no tan solo*, It. *non soltanto* and Sp. *no solamente*. Interestingly, exactly as we observed in Figure A, the results associated to It. and Sp. *solo* (see the small grey boxes) are not very important in explaining the overall result of the test.

Let us now turn to the formal and functional properties of the collocation ‘Neg. + RFA’ in the PACOR_IS data. There are stable features in the discourse context in which we find this collocational pattern, in both Romance languages. Formally, the pattern typically occurs in a correlative construction (see examples 18 to 21 below). Specifically, the collocation ‘Neg. + RFA’ introduces the first member of the correlative

¹⁹ The definition of T-score provided by Sketch Engine is as follows: “T-score expresses the certainty with which we can argue that there is an association between the words, i.e. their co-occurrence is not random”.

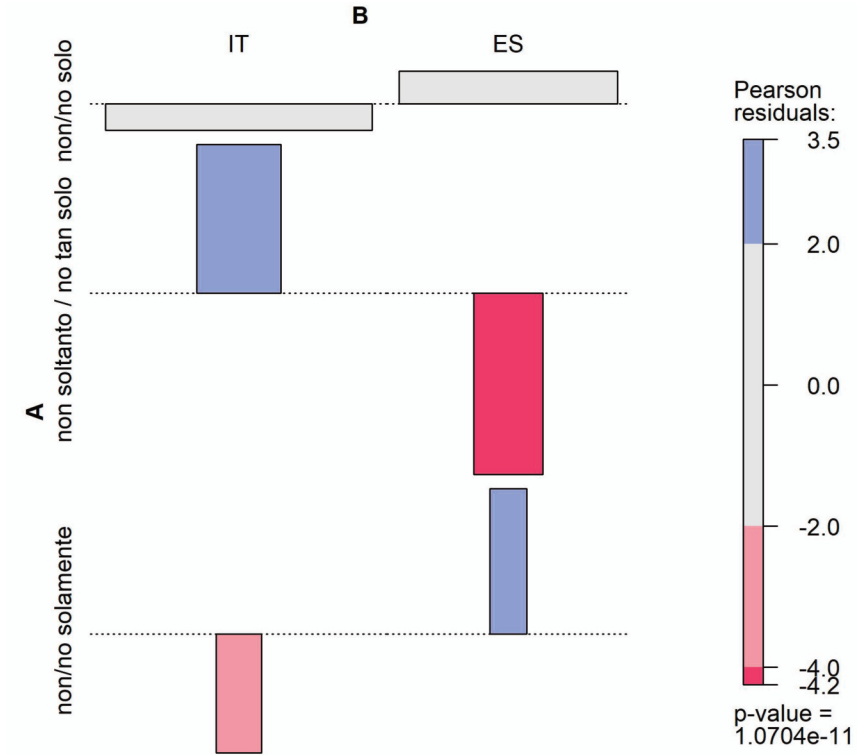


Fig. C. Comparing the distribution of restrictive focus adverbs in academic Italian and Spanish (Cohen-Friendly association plot)

construction and projects a second member, opened by an adversative conjunction (It. *ma*, Sp. *sino*) optionally followed by an additive and/or scalar focus adverb (It. *anche* ‘also’ / *anche e soprattutto* ‘also and foremost’; Sp. *también* ‘also’ / *además* ‘furthermore’). In most of the cases found in PACOR_IS, the collocation ‘Neg. + RFA’ is followed by the DA of the adverb, whereas one (or more) alternative of the DA is explicitly expressed in the second part of the correlative construction. The following examples also show that the entire correlative construction is generally realized in a single utterance.

- (18) It. Queste scritte hanno andamento narrativo; sono precisati per esempio **non solo** l'arrivo al Lazzaretto di navi ed equipaggio e il tempo di permanenza nell'isola, **ma** il luogo di provenienza della nave, le cause della morte degli uomini dell'equipaggio [...] [PACOR_I, doc. 6]

'These writings are narrative-like; they specify, for instance, not only the arrival of boats and crew in Lazzaretto as well as the extent of their stay on the island, but also the boat's place of origin, the cause of the death among members of the crew [...].'

- (19) It. ne risulterebbe un testo **non soltanto** monotono, **ma anche** inconsueto [PACOR_I, doc. 8]

'it would lead to a text that is not only more monotonous, but also more unusual.'

- (20) Sp. [...] y este dato es sumamente significativo **no sólo** por la cantidad, **sino** por la calidad y la fortuna de los términos. [PACOR_S, doc. 19]

'and this information is highly relevant in terms not only of the quantity, but also of the quality and appropriateness of the terms.'

- (21) Sp. Sólo teniendo en cuenta todos estos criterios cualquier diccionario podrá usarse **no solamente** como descodificador, **sino también** como codificador. [PACOR_S, doc. 74]

'Only if all these criteria are considered could a dictionary be used as both a decodifying and a codifying tool.'

The above-mentioned examples also allow us to clarify the main discourse function of the collocation 'Neg. + RFA + DA', which enters in a binary discourse movement presenting a climax (see Andorno / De Cesare 2017: 191-193 for a description of the form and function of the second part of the structure). From the perspective of information structure and argumentation, it introduces the most expected piece of information – coinciding with the content of the adverb's DA – whereas the second part of the correlative construction, introduced by the adversative conjunction, expresses the least expected (possibly also more controversial) piece of information (coinciding with the alternative to the adverb's DA).

There are some instances in the PACOR_IS corpus in which the collocation ‘Neg. + *solo*’ deviates from the prototypical formal and functional patterns described above. These patterns are interesting insofar as they present major intra- and minor cross-linguistic differences. First, Italian *non solo* can occur with an elliptical DA, as in (22), but this construction is not possible with Sp. *no solo*. It is also excluded with It. *non soltanto* and Sp. *no solamente*. This difference points to the fact that It. *soltanto* as well as Sp. *solo* and *solamente* strongly favor the canonical format illustrated in (19), (20) and (21), respectively.

- (22) It. L’italiano è una forte lingua di cultura (e **non solo**) che gode ottima salute. [PACOR_I, doc. 22]

‘Italian is a strong cultural language (and not only) that enjoys excellent health.’

The collocation ‘*no/non solo* + DA’ also occurs as an autonomous piece of discourse, without the second part of the projected construction being expressed. In cases such as (23) and (24), the second part of the correlative construction must be reconstructed in the mind of the reader (e.g. in 23: “proponendo a scuola letture alternative e motivanti non solo di stampo letterario ma anche di stampo non letterario” ‘proposing alternative and motivating school readings not only of literary texts but also of non-literary texts’). Given that it is missing, we perceive that the alternative to the adverb’s DA (“letture alternative e motivanti di stampo non letterario” ‘alternative and motivating school readings of non-literary texts’) is no longer the most relevant piece of information: it is secondary and, as such, relegated to the discourse background.

- (23) It. Già a partire dai dati COMPLINESS era emerso il problema di programmare un intervento didattico specifico sul lessico, calibrato in modo appetibile per i ragazzi, proponendo a scuola letture alternative e motivanti **non solo** di stampo letterario, come ha raccomandato ripetutamente Luca Serianni. [PACOR_I, doc. 31]

‘A problem had already emerged from the COMPLINESS data concerning the programming of a specific didactic intervention on vocabulary, calibrated in an attractive way for children, and proposing alternative and motivating school readings that are not exclusively literary in nature, as Luca Serianni has repeatedly recommended.’

- (24) Sp. [...] es evidente que debemos atender a todas las posibles evoluciones, **no sólo** a las más representativas, pero ello sólo será posible... [PACOR_S, doc. 15]

‘[...] it is clear that we must pay attention to all possible developments, not just the most representative ones, but this will only be possible...’

Again, it seems from the data at hand that the other RFAs tend to occur in the canonical discourse format. *No solamente* always occurs in a correlative construction in PACOR_S, as illustrated in (21). In PACOR_I, a strong tendency for *non soltanto* to occur in correlative constructions, as in (19), is also observable. There are only two cases in which *non soltanto* + DA is not followed by the second part of the correlative construction, which has to be reconstructed.

4.3.2. The coordination of restrictive focus adverbs

In addition to the frequent collocational pattern discussed in § 4.3.1, we identified less frequent patterns that should nonetheless be considered in that they reveal other cross-linguistic differences between It. and Sp. RFAs. One example is the coordination of two RFAs (in the pattern ‘RFA e/y RFA’). As is known from the literature, a RFA cannot be modified by an intensifier such as It. *molto* / Sp. *muy* ‘very’ or It. *estremamente* / Sp. *extremadamente* ‘extremely’ (It. **vogliamo classificare molto/estremamente solo i segnali discorsivi* / Sp. **queremos clasificar muy/extremadamente solo los marcadores discursivos* ‘we want to classify very/extremely only the discourse markers’). The reinforcing of the restrictive feature of a RFA occurs through the coordination of two RFAs.

In the PACOR_I data, It. *solo* occurs in a coordinative construction in which the second coordinate is *esclusivamente* (two occurrences in the same document) or *soltanto* (one occurrence):

- (25) It. Visto che l’attività consiste nell’identificare e nel classificare **solo ed esclusivamente** i SD l’insegnante, chiede ulteriori chiarimenti (turno 19). [PACOR_I, doc. 30]

‘Since the activity consists in identifying and classifying only the SD the teacher asks for further clarification (turn 19)’

- (26) It. Come possiamo noi affermare a priori che un profilo tratti **solo e soltanto** di attualità giornalistica, o di politica, o di questioni aziendali e non presenti espressioni “diaristiche”? [PACOR_I, doc. 13]

‘How can we claim a priori that a profile deals solely and exclusively with journalistic current affairs, or with politics, or with corporate matters and does not present “diaristic” expressions?’

In configurations such as (25) and (26), the meaning of the RFA *solo* is reinforced by the presence of the second restrictive adverb: the second RFA does not add any new semantic content to the one conveyed by *solo*, nor does it cancel a scalar interpretation (which *solo* may license, as seen in Section 2), because the discourse context of both examples does not give rise to such an interpretation. The second RFA in examples such as these merely repeats the restrictive meaning component conveyed by *solo*, thereby reinforcing the restrictive value conveyed by the first adverb. We did not find similar patterns with It. *solamente* and *soltanto* in PACOR_I.

Neither did we find any instances with Sp. *solo*, *solamente* or *tan solo* in PACOR_S. In fact, in Spanish a coordinative construction in which *solo* is reinforced by another RFA seems difficult to accept (*solo y *solamente/*tan solo*) or marginal (*solo y únicamente/exclusivamente*). *Única y exclusivamente* (see example 27, which occurs three times in the corpus) is the only coordinative construction with a reinforcing restrictive function found in PACOR_S. Note that in Spanish it is possible to shorten the first *mente*-derived adverb of the construction by expressing the adverb-suffix marker only once (on this possibility, see Detges 2015).

- (27) Sp. Esto es así porque el foco se proyecta **única y exclusivamente** sobre la alternativa presupuesta, un fondo referencialmente impreciso. [PACOR_S, doc. 5]

‘This is because the focus is solely and exclusively projected onto the presupposed alternative, a referentially imprecise background’

In order to better assess the difference between Italian and Spanish *solo*, let us consider data in larger corpora. Table 13 includes the occurrences based on CORIS and the written part of CORPES XXI.

Collocations	Occurrences	
CORIS (150 million words)	<i>solo ed/e esclusivamente</i>	151
	<i>solo e soltanto</i>	61
	<i>solo e solamente</i>	4
CORPES XXI (77 million words)	<i>solo y exclusivamente</i>	25
	<i>solo y tan solo</i>	0
	<i>solo y solamente</i>	0

Tab. 13. Coordination among restrictive focus adverbs in Italian and Spanish

Bearing in mind that CORPES XXI is half the size of CORIS, we note substantial differences in the distribution of these three coordinative constructions in written Italian and Spanish. The most frequent pattern in Italian is *solo ed esclusivamente*, formally corresponding to Sp. *solo y exclusivamente*, which is nevertheless much less prevalent in CORPES XXI (150 vs. 50 occ.; note that the data has been adjusted to facilitate comparison). It. *solo e soltanto* is also quite frequent; by contrast, Sp. *solo y tan solo* is not documented in CORPES XXI.

The most common construction in Spanish is *única y exclusivamente* (113 occurrences), again including the RFA *exclusivamente* as a second coordinate. This construction does not have a direct equivalent in Italian. The construction *unicamente ed esclusivamente* is not documented at all in CORIS (note that **única ed esclusivamente* is not grammatically correct because Italian does not allow omitting the morpheme *-mente*). The only instances we found with *unicamente* appearing as the first coordinate was *unicamente e solamente* (only two occurrences). The marginality of Italian coordinative constructions with two *mente*-derived adverbs is clearly related to language style. Constructions such as *unicamente ed esclusivamente* and *unicamente e solamente* are highly redundant and are hence perceived as inelegant.

It turns out from the data at hand that the two Romance languages clearly differ in the coordination patterns used: Italian favors the pattern

solo ed esclusivamente, whereas Spanish prefers *única y exclusivamente*. The absence of *solo* in the first part of the coordination could reflect the fact that *solo* in Spanish is also used as an adjective.

5. Conclusion

Contemporary Italian and Spanish show a high degree of polymorphism in the class of adverbs expressing restriction: there are three close cognates in both languages, namely It. *solo*, *solamente* and *soltanto* and Sp. *solo*, *solamente* and *tan solo*. Although these RFAs are equivalent on the semantic level, and morphologically similar, we wanted to know if they differed in usage. No study thus far has given a description of their cross-linguistic similarities and differences in relation to parameters such as their distribution in different language varieties, their discourse frequency, and their occurrence in specific collocations. Our study, which was based on empirical data drawn from various corpora, revealed some interesting cross-linguistic similarities and differences in relation to these parameters.

Amongst the major outcomes of our study we highlight the following findings: (i) there is one main (*solo*) and two marginal (*soltanto/tan solo* and *solamente*) RFAs in both Italian and Spanish; (ii) one of the two marginal forms is clearly more prevalent than the other: *soltanto* in Italian and *solamente* in Spanish. Conversely, It. *solamente* and Sp. *tan solo* are very restricted in their use. A closer look at the distribution of RFAs in written vs. spoken language also shows that the frequency of *solo* is relatively stable across these two language varieties (but more data on spoken Spanish is needed). On the other hand, the frequency of It. *soltanto* and Sp. *solamente* is more dependent upon the language variety. On the basis of these findings we observe that *solo* belongs to the basic vocabulary of Italian and Spanish, whereas the other two RFAs are more specialized lexical items.

In Section 4 of our contribution we described the uses of It. / Sp. *solo*, *solamente*, *soltanto* / *tan solo* in academic texts, based on a self-assembled comparable corpus (PACOR_IS). Our analysis shows that there are both relevant similarities and differences between It. and Sp. *solo* in this genre. As far as their similarities are concerned, in both languages the RFA shows up surprisingly often in the collocation ‘Neg. + RFA’, which generally enters a larger textual movement, involved in the construction of a binary structure expressing a climax. As for the differences, we discovered three

notable facts. First, It. *solo* is more frequent than its Spanish equivalent. Second, the sequence ‘Neg. + RFA’ may occur with an elliptical DA in Italian, but not in Spanish. Third, with regard to the coordination pattern ‘RFA e/y RFA’, allowing reinforcement of the restrictive meaning component of the first RFA, we observe that It. *solo* enters two different patterns (*solo ed esclusivamente* and *solo e soltanto*, the first one being more common), whereas Sp. *solo* enters only one (*solo y exclusivamente*), which is much more marginal than in Italian. The operation consisting of reinforcing the restrictive feature of a RFA in Spanish is achieved by resorting to an entirely different pair of adverbs (*única y exclusivamente*). These results show that It. and Sp. *solo* are far from overlapping when their usage is considered in more detail and in authentic discourses. These results can be extended to the pairs of RFAs ending in *-mente*: Italian favors the use of *esclusivamente* over *unicamente*, while Spanish favors the use of *únicamente* over *exclusivamente*.

In order to further explain some of the results found in this study, such as the cross-linguistic gap between It. / Sp. *solo* in academic texts, we should add an additional step to the research design and analyze data drawn from an Italian-Spanish translation corpus. If we examined the translation equivalents of the most common Italian and Spanish RFAs, we would be able to compile an inventory of restrictive expressions in both Romance languages and determine their relationship.²⁰ More specifically, on the basis of the measure proposed in Johansson (2007), we would be able to calculate the degree of mutual correspondence between It. and Sp. *solo*, and evaluate with greater precision how close these two cognates really are (for a similar study applied to It. *anche* and Fr. *aussi*, see Andorno / De Cesare 2017).

Anna-Maria De Cesare
Universität Basel
Seminar für Italianistik
Maiengasse 51
4056 Basel, Schweiz
anna-maria.decesare@unibas.ch

Begoña Sanromán Vilas
Helsingin yliopisto
Department of Languages
Unioninkatu 40
00014 Helsinki, Finland
begona.sanroman@helsinki.fi

²⁰ See Nicklaus (2015), for a study following a similar procedure, in which the Italian and French translations of the German RFAs *nur* and *allein* ‘only’ occurring in literary texts are investigated.

References

- Alloua, Hugo / Navilli de Scotta, Egle / Pedrotti, Beatriz, 2008, *Estudios de Lingüística Contrastiva: italiano-español*, Córdoba (Argentina), Comunicarte.
- Andorno, Cecilia, 1999, “Avverbi focalizzanti in italiano. Parametri per un’analisi”. *Studi Italiani di Linguistica Teorica e Applicata* 28/1: 43-83.
- Andorno, Cecilia, 2000, *Focalizzatori tra connessione e messa a fuoco. Il punto di vista delle varietà di apprendimento*, Milano, Angeli.
- Andorno, Cecilia Maria / De Cesare, Anna-Maria, 2017, “Mapping additivity through translation: From French *aussi* to Italian *anche* and back in the Europarl-direct corpus”. In: De Cesare, Anna-Maria / Andorno, Cecilia Maria (eds.), *Focus on Additivity. Adverbial Modifiers in Romance, Germanic and Slavic Languages*, Amsterdam-Philadelphia, Benjamins: 157-200.
- Chiari, Isabella / De Mauro, Tullio, 2012, “The new basic vocabulary of Italian: problems and methods”. *Statistica applicata* 22: 21-35.
- Corpas Pastor, Gloria, 2017, “Collocational constructions in translated Spanish: What corpora reveal”. In: Mitkon, Ruslan (ed.), *Computational and Corpus-Based Phraseology: Second International Conference Europhras 2017, London, UK, November 13-14, 2017. Proceedings*, Wolverhampton, Springer: 29-40.
- De Cesare, Anna-Maria, 2015, “Defining *Focusing Modifiers* in a cross-linguistic perspective. A discussion based on English, German, French and Italian”. In: Pittner, Karin / Elsner, Daniela / Barteld, Fabian (eds.), *Adverbs – Functional and Diachronic Aspects*, Amsterdam-Philadelphia, John Benjamins: 47-81.
- De Cesare, Anna-Maria, 2019. CONTRAST-IT e COMPARE-IT. Due nuovi corpora per l’italiano contemporaneo: *CHIMERA. Romance Corpora and Linguistic studies* 6: 43-74.
- De Cesare, Anna-Maria / Garassino, Davide / Agar Marco, Rocío / Albom, Ana / Cimmino, Dorian, 2016, *Sintassi marcata dell’italiano dell’uso medio in prospettiva contrastiva con il francese, lo spagnolo, il tedesco e l’inglese. Uno studio basato sulla scrittura dei quotidiani online*, Frankfurt am Main, Lang.
- Detges, Ulrich, 2015, “The Romance adverbs in *-mente*: a case study in grammaticalization”. In: Müller, Peter O. / Ohnheiser, Ingeborg / Olsen, Susan / Rainer, Franz (eds.), *Word-Formation. An International Handbook of the Languages of Europe*, vol. 3, Berlin, De Gruyter Mouton: 1824-1842.
- Ferrari, Laura / Giammatteo, Mabel / Albano, Hilda, 2011, “Operadores de foco: El caso de *incluso, hasta, solo y aun*”. *Cuadernos de la ALFAL* 3: 30-41.

- Fuentes Rodríguez, Catalina, 2002, “Solamente y la función pragmática de la unicidad”. *Linguística Española Actual* 24/2: 277-303.
- García Pérez, Rafael, 2013, “La evolución de los adverbios de foco en español: adverbios focalizadores de exclusión, inclusión y aproximación”. In: Garcés Gómez, María Pilar (ed.), *Los adverbios con función discursiva. Procesos de formación y evolución*, Madrid/Frankfurt, Iberoamericana/Vervuert: 317-388.
- Geerts, Walter, 1977, “On the semantics of Italian adverbial *solo*”. *Archivum linguisticum* 8/1: 3-12.
- Hummel, Martin 2013, “La dimensión intercultural de la expansión diacrónica de los adverbios en *-mente*”, In: Garcés Gómez, María Pilar (ed.), *Los adverbios con función discursiva. Procesos de formación y evolución*, Madrid/Frankfurt, Iberoamericana/Vervuert: 15-41.
- Johansson, Stig, 2007, *Seeing through Multilingual Corpora. On the Use of Corpora in Contrastive Studies*, Amsterdam-Philadelphia, Benjamins.
- Kilgarriff, Adam / Baisa, Vít / Bušta, Jan / Jakubiček, Miloš / Kovář, Vojtěch / Michelfeit, Jan / Rychlý, Pavel / Suchomel, Vít, 2014, “The Sketch Engine: ten years on”. *Lexicography* 1/1: 7.
- König, Ekkehard, 1981, “The meaning of scalar particles in German”. In: Eikmeyer, Hans-Jürgen / Rieser, Hannes (eds.), *Worlds, Words and Contexts*, Berlin, de Gruyter: 107-132.
- König, Ekkehard, 1991, *The Meaning of Focus Particles: A Comparative Perspective*, London-New York, Routledge.
- Kovacci, Ofelia, 1999, “El adverbio”. In: Bosque, Ignacio / Demonte, Violeta (eds.), *Gramática descriptiva de la lengua española*, vol. 1. *Sintaxis básica de las clases de palabras*, Madrid, Espasa Calpe: 705-786.
- La Forgia, Francesca / Carreras i Goicoechea, Maria, 2008, “Anche solo: riformulazioni e traduzioni in italiano, catalano e spagnolo”. In: Cresti, Emanuela (a cura di), *Prospettive nello studio del lessico italiano, Atti del IX Congresso della Società di linguistica e filologia italiana* (Firenze 14-17 giugno 2008), vol. 2, Firenze, Firenze University Press: 425-431.
- Manzotti, Emilio, 1984, “Costrutti esclusivi e restrittivi in italiano”. *Vox Romanica* 43: 50-80.
- Medina Gómez, Lorena Y. / Alarcón Neve, Luisa Josefina, 2017, “Descriptive and functional analysis of the *solo-solamente* adverbial pair in spoken Mexican Spanish”. In: Hummel, Martin / Valera, Salvador (eds.), *Adjective Adverb Interfaces in Romance*, Amsterdam, Benjamins: 287-303.

- Medina Gómez, Lorena Y. / Fernández García, Francisco, 2018, “Análisis del par adverbial *solo-solamente* como matizador pragmático-discursivo en el español de México”. *Oralia* 21/1: 63-86.
- Montoro del Arco, Esteban T., 2011, “Condicionamientos sociolingüísticos de la variación fraseológica: los operadores de foco de exclusión”. *Lingüística Española Actual* 33/2: 223-263.
- Montoro del Arco, Esteban T., 2012, “¿Son los hablantes andaluces más expresivos? La variación pragmática en el uso de los operadores focales de exclusión en Granada y Valencia”. In: Moya Corral, Juan Antonio / Waluch-de la Torre, Edyta (eds.), *Español hablado. Estudios sobre el corpus PRESEEA-Granada*, Varsovia, Instituto de Estudios Ibéricos e Iberoamericanos de la Universidad de Varsovia: 77-93.
- Nicklaus, Martina, 2015, “Nicht-analoge französische und italienische Übersetzungen von restriktiven Partikeln im Roman *Die Mittagsfrau* von Julia Franck”. *Linguistik online* 71/2: 103-127.
- RAE / ASALE, 2009, *Nueva gramática de la lengua española*, Madrid, Espasa.
- RAE / ASALE, 2010, *Ortografía de la lengua española*, Madrid, Espasa.
- Ricca, Davide, 1999, “Osservazioni preliminari sui focalizzatori in italiano”. In: Dittmar, Norbert / Giacalone Ramat, Anna (a cura di), *Grammatica e discorso. Studi sull'acquisizione dell'italiano e del tedesco*, Stauffenburg, Tübingen: 146-164.
- Sánchez López, Cristina, 1999, “Los cuantificadores: Clases de cuantificadores y estructuras cuantificativas”. In: Bosque, Ignacio / Demonte, Violeta (eds.), *Gramática descriptiva de la lengua española*, vol. 1. *Sintaxis básica de las clases de palabras*, Madrid, Espasa Calpe: 1025-1128.
- Tekin, Özlem, 2012, *Grundlagen der Kontrastiven Linguistik in Theorie und Praxis*. Tübingen, Stauffenburg.

Dictionaries

- CLAVE = Maldonado, Concepción, 2006, *Diccionario de uso del español actual*, Madrid, SM. <http://clave.smdiccionarios.com/app.php>
- Davies, Mark, 2006, *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*, New York/London, Routledge.
- DCOE = Fuentes Rodríguez, Catalina, 2018, *Diccionario de conectores y operadores del español*, Madrid, Arco/Libros-La Muralla.

DEA = Seco, Manuel / Andrés, Olimpia / Ramos, Gabino, 2011, *Diccionario del español actual*, Madrid, Aguilar, 2nd ed.

DISC = Sabatini, Francesco / Coletti, Vittorio, 1997, *Dizionario Italiano Sabatini Coletti – DISC*, Firenze, Giunti.

DLE = RAE / ASALE, 2014, *Diccionario de la lengua española*, Madrid, Espasa, 23st ed.

DP = Santos Ríos, Luis, 2003, *Diccionario de partículas*, Salamanca, Luso-Española de Ediciones.

DSLE = Gutiérrez Cuadrado, Juan (dir.), 2006, *Diccionario Salamanca de la lengua española*, Madrid, Santillana. <http://fenix.cnice.mec.es/diccionario/>

DUE = Moliner, María, 2007, *Diccionario de uso del español*, Madrid, Gredos, 3rd ed.

LEMA = Battaner, Paz (dir.), 2001, *Diccionario de la lengua española. Lema*, Barcelona, Spes Editorial.

LIP = De Mauro, Tullio / Mancini, Federico / Vedovelli, Massimo / Voghera, Miriam, 1993, *Lessico di frequenza dell'italiano parlato*, Milano, Etaslibri.

NVDB = De Mauro, Tullio, 2016, *Nuovo vocabolario di base della lingua italiana*. <https://dizionario.internazionale.it>

TRECCANI, *Vocabolario on line*, <http://www.treccani.it/vocabolario>

Corpora

BADIP = <http://badip.uni-graz.at/en>

CONTRAST-IT = http://philhist-contrast-it-noske.philhist.unibas.ch/cnt/run.cgi/first_form

CORIS = <http://corpora.dslo.unibo.it/TCORIS>

CORPES XXI = <https://www.rae.es/recursos/banco-de-datos/corpes-xxi>

DB-IPIC = <http://lablita.it/app/dbipic/index.php?corpus>

Val.Es.Co. = <http://www.valesco.es/?q=corpus>

LAURA MORI
(Università degli studi internazionali di Roma)

La rappresentazione di scenari deontici e l'espressione della performatività nell'italiano delle leggi: dal diritto europeo alla legislazione nazionale¹

This study focuses on the intra-linguistic variation between three legislative corpora belonging to the Eurolect Observatory Multilingual Corpus: Corpus A (EU directives); Corpus B (Italian laws of implementation) and Corpus C (national laws of Italian origin).

Profiling through Natural Language Processing is combined with quantitative analyses led using corpus-based and corpus-driven methodologies. These data directed qualitative research on relevant phenomena in order to depict the most significant semantic and pragmatic features of legal discourse, namely the deontic encoding and legal performativity. Morphological variants and morphosyntactic patterns featuring the legal framing are detected and exemplified focusing on their double, sometimes overlapped, deontic and performative value. In particular, the role of the present indicative in legislative utterances, the distribution of passive-active diathesis (in relation with agency), the collocational profile of legal verbs as far as the encoding ofthetic-constitutive and prescriptive values are concerned, the distribution of legal performatives across corpora and the relationship between performativity and textuality are analysed.

Cross-corpora results confirm the existence of context-induced linguistic differences and intra-genre variability within the legal domain related to three legislative varieties: Italian Eurolect (Corpus A), EU-derived legislative Italian (Corpus B) and national legislative Italian (Corpus C).

1. Introduzione

Il linguaggio giuridico è un'entità complessa da definire per l'eterogeneità dei testi, scritti e orali, in cui si manifesta tanto da poter essere descritto, sotto il profilo linguistico, in termini di varietà con caratteristiche correlate al registro (Biber / Conrad 2009) e allo stile (Berruto

¹ Per la redazione di questo articolo sono grata ad Annarita Felici e a Giulia Venturi per lo scambio fruttuoso e gli spunti di riflessione utili ad approfondire lo studio. Ringrazio inoltre i revisori anonimi per i suggerimenti che mi hanno consentito di mettere maggiormente a fuoco la tematica.

2011) e, in prospettiva pragmatica (*legal discourse*, Kurzon 1997) e testuale, in relazione allo scopo che si concretizza in un dato genere testuale (Bhatia 1993).

I testi giuridici sono caratterizzati da fenomeni di variazione linguistica che investono vari livelli, in parte rintracciabili nel linguaggio giuridico di diverse tradizioni europee: varianti lessicali marcate in alto in diafasia (tecnicismi specifici, tra cui latinismi giuridici) e tecnicismi collaterali correlati allo stile di un determinato genere, o sotto-genere, testuale; categorie morfologiche e pattern morfosintattici per la codifica della prescrittività e della performatività; combinazioni ricorrenti di parole (*lexical bundles*), strategie di deagentivizzazione (nominalizzazioni e diatesi passiva); complessità sul piano sintattico e una rigida articolazione dell'impianto testuale².

Per quanto riguarda la fisionomia dell'italiano giuridico, essa deriva da una convergenza secolare e dalla stratificazione di diverse tradizioni giuridiche, autoctone e straniere. In particolare, il bilinguismo latino-varietà vernacolari di epoca medievale e il plurilinguismo italiano-latino-francese (con l'influenza dell'apparato concettuale tedesco), durante il XVII e il XVIII secolo, hanno caratterizzato la storia dell'italiano come lingua del diritto per la quale la traduzione da altre lingue ha giocato un ruolo rilevante (Bambi 2012; Caterina / Rossi 2008). Più recentemente, l'italiano giuridico è stato plasmato dal contatto intenso con il francese e l'inglese nel contesto plurilingue dell'Unione europea e dagli scambi all'interno del mondo globalizzato.

La legislazione europea in italiano è frutto un processo di negoziazione politica tra culture linguistiche e giuridiche che si manifesta all'interno di un circuito di redazione-traduzione-revisione plurilingue attivato per rispondere ai vincoli del quadro sovranazionale dell'UE. In tal modo si producono "testi armonizzati" in 24 versioni linguistiche facenti ugualmente fede. Infatti, a livello di Unione europea, la centralità (e non ufficialità) della traduzione impone un nuovo concetto di redazione legislativa, dove "it is the legal language of the institutional framework that is detached from national systems and moulded through translation" (Prieto Ramos 2014: 319). La traduzione giuridica multilin-

² Per una descrizione delle caratteristiche linguistiche dell'italiano giuridico si rimanda a Mortara Garavelli (2001).

gue della legislazione dell'UE è, dunque, frutto di compromessi tra i diversi Paesi membri che mirano al raggiungimento di un'armonizzazione sociale, politica e giuridica. Sul piano linguistico studi condotti su diverse lingue europee all'interno dello *Eurolect Observatory Project*³ hanno già dimostrato empiricamente l'esistenza di varietà legislative europee che si differenziano dalle rispettive varietà legislative nazionali⁴: i cosiddetti euroletti. I risultati delle analisi su corpora per undici lingue mostrano una situazione di plurilinguismo con evidenti analogie e differenze tra varietà legislative europee e peculiarità rispetto alla posizione che ciascuna di esse assume all'interno del proprio spazio sociolinguistico nel contesto nazionale. In particolare si evidenziano dinamiche che incidono sulle caratteristiche distintive degli euroletti, a più livelli linguistici, e riconducibili a tre macro-categorie interpretative: 1. fenomeni condizionati dal contesto UE (*EU-rooted phenomena*), 2. caratteristiche indotte da contatto (*contact-induced features*), 3. variabilità intralinguistica (*intra-linguistic variability*)⁵.

Per quanto riguarda l'euroletto italiano, esso vive principalmente come lingua tradotta da versioni redatte in inglese⁶, da scriventi nativi e non nativi, in quanto dal 2015 l'inglese rappresenta la lingua fonte della legislazione europea nell'81% dei casi. Tuttavia questo scenario di predominio caratterizza solo il periodo più recente mentre il francese, lingua primaria per la redazione del diritto europeo nei primi decenni della storia delle Comunità, rappresenta ancora oggi la lingua fonte in alcuni ambiti, come la pesca e l'agricoltura, e le versioni in francese costituiscono un riferimento nella prassi traduttiva per effettuare riscontri interlinguistici. Come osservato da Mattila (2013: 33), alla fine degli anni Novanta ancora il 40,4 % dei documenti tradotti dalla Direzione Gene-

³ Per maggiori dettagli sul progetto di ricerca internazionale, coordinato da chi scrive, si rimanda alla pagina <http://www.unint.eu/eurolect-observatory/overview>.

⁴ Si rinvergono analogie e differenze tra gli euroletti che sono state descritte e interpretate da Mori / Szmrecsany (in stampa) utilizzando una visualizzazione di tipo *Multidimensional Scaling* basata su distanze linguistiche aggregate di parametri lessicali e grammaticali.

⁵ Cfr. Mori (2018a).

⁶ "Documents are negotiated in a lingua franca which is more and more English but also very little tied to the Anglo-Saxon tradition of legal drafting" (Felici 2013: 42). In effetti, l'inglese giuridico nazionale – sviluppatosi all'interno di un sistema di *common law* – sarebbe la lingua meno adeguata a esprimere i concetti del diritto dell'Unione europea (Pozzo 2012). Pertanto si viene a configurare come una varietà altra (euroletto inglese in Sandrelli 2018), con proprietà che portano a considerarla una lingua franca istituzionale.

rale della Traduzione della Commissione europea erano originariamente redatti in francese rispetto al 45,4% di quelli in inglese. Tale situazione ha subito un cambiamento notevole nel 2008 quando la percentuale di originali in inglese raggiunge il 72,5% e l'uso del francese per la redazione si riduce drasticamente (11,8%)⁷. Tali dati non sono trascurabili e rendono fondamentale la considerazione dell'arco temporale in cui sono stati prodotti i testi quando si analizza la normativa prodotta dai servizi linguistici dell'UE, poiché determinate varianti potrebbero essere dovute al contatto fondante e di lunga data con il francese⁸ e non necessariamente con l'inglese⁹.

Il contesto di plurilinguismo legislativo che si viene a configurare prevede un flusso di lavoro incentrato sulla traduzione interlinguistica della legislazione europea in cui si favorisce l'adozione di soluzioni linguistiche che possano facilitare la comparabilità delle versioni in ventiquattro lingue. Diversamente nella riscrittura di testi derivati da norme europee, ma proiettati e indirizzati alla realtà socio-culturale nazionale, quali le leggi di attuazione di direttive europee, si verifica un processo di "ibridazione", tra due culture giuridiche e due modelli linguistici di riferimento, europeo e nazionale (cfr. Mori 2019d).

Partendo da questo assunto, è interessante osservare la fenomenologia della variazione rispetto a tali contesti (europeo e nazionale) utilizzando le potenzialità delle analisi condotte su corpora contenenti tre diversi sotto-generi: direttive europee (Corpus A), misure italiane di attuazione delle direttive contenute nel Corpus A (Corpus B) e norme nazionali (Corpus C). Nel caso del Corpus A si tratta di testi legislativi originati a livello UE per vincolare gli Stati Membri al raggiungimento di specifici risultati attualizzati dalle leggi nazionali di attuazione (Corpus B) che rimandano alle finalità imposte dalle direttive, da cui derivano e dipendono, pronunciandosi su modalità e mezzi di applicazione. Il confronto con un corpus di norme nazionali senza alcuna re-

⁷ Nel 2008 il 2,7% dei documenti risultano redatti in tedesco mentre il rimanente 13% è originariamente prodotto in altre lingue (cfr. Commissione europea 2009: 7-8).

⁸ Occorre, infatti, considerare come sia stato il diritto francese a influenzare i concetti fondanti del diritto europeo insieme ai principi provenienti dalla tradizione giuridica tedesca (Mattila 2013: 139-140).

⁹ Ad esempio nella descrizione dell'euroletto italiano fornita in Mori (2018c) si evidenziano tracce di contatto con il francese, nella forma di prestiti e calchi.

A questo proposito è opportuno considerare che i servizi linguistici delle istituzioni dell'Unione europea prevedono un coordinamento giurilinguistico per cui giuristi, esperti di qualità, traduttori, giurilinguisti si trovano ad interagire sistematicamente. A livello nazionale, invece, in Italia non esiste a tutt'oggi un'unità centrale impegnata nel processo di attuazione delle direttive sul piano linguistico, al fine di fornire indicazioni coerenti e condivise, diversamente da quanto accade in altri Paesi europei. Infatti, il Dipartimento per gli affari giuridici e legislativi (DAGL) della Presidenza del Consiglio dei Ministri si occupa di esaminare la qualità della regolazione e la sua aderenza al programma di Governo, verificando in particolare l'esame delle fonti giuridiche, il rispetto dei principi costituzionali, la conformità con legislazione dell'UE, con i Trattati e con la giurisprudenza europea (Decreto legislativo 303/1999, art. 6), ma esso non è supportato da un'unità di controllo strettamente linguistico. Ciò, peraltro, è in linea con l'assenza nel panorama italiano – se confrontato con altri paesi europei – di un'azione dall'alto di pianificazione e controllo della qualità redazionale delle leggi per quanto riguarda gli aspetti linguistici in senso ampio (Mori 2019c).

Pertanto, le diverse condizioni extra-linguistiche in cui è redatta la legislazione di origine europea (Corpus A), rispetto a quella italiana di derivazione europea (Corpus B) e di origine nazionale (Corpus C) determinano dinamiche variazionali che meritano di essere approfondite in relazione a una tematica di particolare rilievo per la descrizione linguistica dei testi normativi: la compresenza e l'interazione di dimensione prescrittiva e performativa.

2. Il monitoraggio linguistico di alcune proprietà di semantica e pragmatica legislativa

In ambito giuridico, la performatività degli atti linguistici (Austin 1962; Searle 1969) viene evocata dal filosofo del diritto Carcaterra (1974) là dove l'autore riconosce una corrispondenza tra gli atti dichiarativi e la proprietà costitutiva delle norme. Oltre alla prescrittività, gli studi di filosofia del diritto identificano, infatti, anche la costitutività come proprietà performativa degli enunciati legislativi: “il diritto non soltanto prescrive comportamenti ma forma anche, è appunto costitutivo di modi di essere, proprietà, relazioni degli enti giuridici” (Carcaterra

1979: 68). Secondo Carcaterra, mentre le proposizioni prescrittive tendono a produrre un evento – esercitando una pressione sul comportamento altrui –, gli enunciati costitutivi producono un effetto immediato: “l’oggetto, la situazione, lo stato di cose cui la norma si riferisce viene concretamente a sussistere nel momento in cui la norma entra in vigore” (Roversi 2012: 13).

Da tale prospettiva, in questo studio si propone un’analisi della codifica della modalizzazione deontica e della performatività nel linguaggio legislativo italiano (europeo e nazionale) avvalendoci delle metodologie della linguistica computazionale e della linguistica dei corpora e adottando un approccio quali-quantitativo per l’interpretazione dei dati. Si ipotizza, infatti, che la correlazione con un diverso contesto giuridico si manifesti linguisticamente nelle varianti che codificano la dimensione semantica e pragmatica degli enunciati legislativi. Nello specifico, per quanto riguarda i testi legislativi ancorati alla cornice giuridica europea, si è optato per le direttive in quanto strumenti dinamici fondati su un’interazione tra una fase in cui si definiscono gli obblighi, quindi vincolati a condizioni prescrittive, e un momento, successivo e consequenziale, preposto all’applicazione della norma. A questo proposito Caliendo (2004: 245) identifica l’esistenza di una “recognition phase”, con cui l’autrice descrive la fase di attuazione nel diritto interno di ogni Stato Membro. In tal caso ci si confronta con un messaggio prescrittivo caratterizzato da una proiezione futura della performatività. Tale specificità potrebbe, dunque, tradursi in un diverso bilanciamento tra modalità prescrittiva e modalità performativa nella normativa europea rispetto a quella nazionale. Occorre, infatti, considerare che:

the law is based on experience drawn from real world but it regulates hypothetical future cases. In consequence, the timespan linked to legal rules is often characterised by certain universality, impossible to see from the chronological standpoint. [...] This is evident, for example, in the use of verbs in legislative language. (Mattila 2013: 96)

La ricerca che qui si propone mira a verificare la rappresentazione degli scenari giuridici considerando la sua complementarità con l’espressione della performatività, quali aspetti fondanti del discorso legislativo. L’analisi degli scenari deontici linguisticamente rappresentati

nei diversi testi legislativi (europei, di derivazione europea, e italiani) potrebbe rivelare una differenziazione intralinguistica dovuta alla sovrarappresentazione o sottorappresentazione di determinate varianti nei corpora: A vs. B, A vs. C, ma anche B vs. C. Sebbene si tratti di testi appartenenti al medesimo genere, già altri studi hanno dimostrato come la rappresentazione dell'informazione relativa agli scenari deontici possa essere soggetta a una variabilità in base al corpus in esame¹¹.

In questo caso, le specificità dei testi di diritto europeo potrebbero essere riconducibili all'interferenza delle lingue maggiormente utilizzate in fase di redazione, in particolare il francese e l'inglese. Infatti, la necessità di adottare scelte redazionali che possano facilitare l'allineabilità delle ventiquattro versioni linguistiche incentiva la produzione di versioni tradotte che siano speculari rispetto al testo di partenza.

I corpora legislativi qui considerati appartengono allo *Eurolect Observatory Multilingual Corpus* (EOMC)¹² e coprono l'arco temporale 1999-2008 (Corpus A)¹³ e 1999-2013 (Corpus B e Corpus C). Per procedere al confronto intercorpora si è scelto di concentrare l'analisi sul materiale linguistico proveniente dalle sezioni normative delle leggi, escludendo eventuali allegati¹⁴, e tenendo separate le sezioni di preambolo e articolato. Pertanto, le occorrenze riportate nei paragrafi che seguono sono estratte dalle sezioni di articolato dei tre corpora: sub-corpus A (1.439.069 tokens), sub-corpus B (1.978.795 tokens), sub-corpus C (1.511.738 tokens).

Nei paragrafi che seguono si forniranno alcuni risultati quantitativi provenienti dall'estrazione automatica di caratteristiche morfosintattiche mediante un'analisi linguistico-computazionale di complessità cre-

¹¹ Nello studio di Venturi (2011), sulla relazione tra lingua e diritto in prospettiva linguistico-computazionale, l'autrice ha condotto un'analisi su *frame* semantici tipizzanti corpora di linguaggio giuridico a partire da un'annotazione sintattica a dipendenze realizzata automaticamente.

¹² Per dettagli tecnici su compilazione, architettura e strumenti per l'interrogazione dell'EOMC si rimanda a Tomatis (2018).

¹³ Quasi un terzo delle direttive complessivamente emanate in tale periodo sono indicizzate all'interno del *Repertorio della legislazione dell'UE* (www.eur-lex.europa.eu) come appartenenti ai settori "Politica industriale e mercato interno" (area 13) e "Agricoltura" (area 3), in misura minore al settore "Ambiente, Consumatori e protezione della salute" (area 15) e solo marginalmente coinvolgono gli altri settori (per maggiori dettagli si veda Mori 2018b: 16).

¹⁴ Negli allegati dei testi legislativi la presenza di lessico specialistico proveniente da altri settori (a seconda della politica su cui la norma legifera) è molto più consistente, e sia la sintassi che la testualità presentano differenze rilevanti rispetto alle sezioni normative (preambolo e articolato).

scente (cfr. Montemagni 2013)¹⁵: dalla segmentazione del testo in parole ortografiche (*tokens*), all’analisi morfo-sintattica e lemmatizzazione del testo a seguito del processo di “tokenizzazione”, fino all’analisi della struttura sintattica della frase in termini di relazioni di dipendenza.

Contestualmente si procederà all’applicazione delle metodologie della linguistica dei corpora (cfr. Tognini Bonelli 2001), adottando un approccio *corpus-based* finalizzato a isolare i fenomeni di variazione più rilevanti in termini di frequenza e un approccio *corpus-driven* per identificare i profili collocazionali degli *item* linguistici ritenuti più rilevanti ai fini dell’analisi. In tal modo sarà possibile indirizzare la successiva fase di analisi qualitativa di verbi e pattern che, all’interno delle leggi, possono avere sia valenza prescrittiva che performativa. L’esemplificazione dai tre corpora consentirà di evidenziare empiricamente gli aspetti semantici e pragmatici che investono il discorso legislativo, esplorando anche eventuali correlazioni con la testualità che caratterizza i tre corpora.

Nella discussione, là dove le questioni di semantica e pragmatica giuridica lo consentano, si mirerà a proporre una riflessione in un’ottica contrastiva in particolare rispetto alla descrizione dell’euroletto italiano per il quale le tracce di contatto con l’inglese, ma anche con il francese, sono all’origine di molte sue specificità.

3. *Gli scenari deontici per la rappresentazione della semantica giuridica*

Tra le aree soggette a variabilità intralinguistica merita di essere approfondita la rappresentazione della semantica giuridica negli enunciati che esprimono atti direttivi, funzionali alla natura primariamente prescrittiva delle leggi: “[...] laws are written in language the function of which is not just to express or convey knowledge and information, but also to direct, influence or modify people’s behaviour” (Cao 2007: 13). Nelle leggi, in quanto testi di tipo regolativo secondo la nota tipologia di Werlich (1976), le strategie predicazionali assumono un ruolo centrale nella pratica discorsiva; pertanto si presuppone che, mediante l’analisi

¹⁵ Si ringrazia Giulia Venturi per aver estratto i dati morfosintattici dell’EOMC utilizzando le tecnologie sviluppate dal laboratorio ItaliaNLP dell’Istituto di Linguistica Computazionale (ILC) “A. Zampolli” di Pisa.

si delle forme verbali e dei costrutti morfosintattici più ricorrenti, sarà possibile isolare alcune peculiarità del linguaggio legislativo italiano in riferimento al contesto giuridico: europeo o nazionale.

Innanzitutto occorre muovere dalla considerazione che i testi appartenenti al Corpus A, come già evidenziato (§ 1), sono esito di un processo di traduzione per cui nella maggior parte dei casi le forme che esprimono la prescrittività in euroletto italiano traducono le varianti delle versioni originali in inglese o in francese. Si partirà, dunque, dalle indicazioni contenute nelle linee guida nazionali (inglese, francese e italiana) per la redazione dei testi legislativi rispetto alla codifica della prescrittività:

Drafting guidance. Office of the Parliamentary Counsel (2008: 14)

There are various alternatives to “shall” which can be used, depending on context:

- “must” in the context of obligations (although “is to be” and “it is the duty of” may also be appropriate alternatives in certain contexts);
- “there is to be” in the context of the establishment of new statutory bodies etc.;
- use of the present tense in provisions about application, effect, extent or commencement;
- “is amended as follows” in provisions introducing a series of amendments;
- “is repealed” in the context of free-standing repeals;
- “is to be” in the context of provisions relating to statutory instruments (and, if appropriate, “may not” as an alternative to “shall not”).

Guide de légistique (2005: 165):

En règle générale, les verbes sont à conjuguer au présent de l’indicatif et non au futur. Ce présent a valeur impérative.

Regole e raccomandazioni per la formulazione tecnica dei testi legislativi. Circolare del Presidente del Senato (2001: 8):

Nella formulazione dei precetti è adottata la massima uniformità nell’uso dei modi verbali, la regola essendo costituita dall’indicativo presente, escludendo sia il modo congiuntivo sia il tempo futuro¹⁶.

¹⁶ A questo proposito è interessante notare come la stessa *Circolare* si pronuncia sull’uso del verbo *dovere*: “è evitato l’uso del verbo servile diretto a sottolineare la imperatività della norma (“deve”; “ha l’obbligo di”; “è tenuto a”). Si consideri, tuttavia, che l’analisi empirica dimostra come l’uso del verbo *dovere* risulta essere maggiormente utilizzato nelle leggi nazionali italiane, in particolare in quelle di attuazione, rispetto alla legislazione europea (cfr. Mori 2020).

Nonostante l'indicazione a preferire forme alternative a *shall*, in inglese giuridico sono numerosi gli studi che evidenziano la sua sovrarappresentazione per l'espressione di un comando positivo nel diritto europeo (ad es. Caliendo 2004, Caliendo *et al.* 2005; Felici 2012; Garzone 2008, 2013; Sandrelli 2018, 2019; Williams 2005, 2011). Nella legislazione dell'Unione europea *shall* rappresenta la soluzione preferita per esprimere un dovere,¹⁷ diversamente da quanto si osserva nell'inglese giuridico nazionale, colpito dalla cosiddetta *modal revolution*¹⁸ (Garzone, 2013: 69). Infatti, l'uso di *shall*, come variante controversa nella redazione di testi prescrittivi, è motivata dalla constatazione che essa può esprimere sia la modalità deontica, per definire obblighi e necessità, che avere un valore performativo, in quanto consente di attuare contestualmente lo stato di cose enunciato (Garzone, 1996, 1999). Inoltre è interessante osservare come la valenza (prescrittiva o performativa) di *shall* può essere associata a proprietà morfologiche e semantiche del soggetto: “it has a performative value when it is used with an impersonal or inanimate subject and refers to a state of things which becomes effective only for the fact of being predicated in the text” (Garzone / Salvi 2007: 40).

Dal confronto tra euroletto inglese e inglese legislativo nazionale, le evidenze di una redazione *shall-free*¹⁹ emergono maggiormente nelle leggi nazionali di attuazione dove *shall* è circoscritto nel rispetto delle *Plain language techniques* che invocano una sua sostituzione con il modale *must* o con il semi-modale *is/are to*. I risultati emersi dal recente studio di Sandrelli (2018) hanno evidenziato una significativa sovrarap-

¹⁷ Caliendo (2004) riporta una diversa distribuzione delle forme *shall/should/must* nel diritto europeo in relazione al genere testuale considerato. L'autrice si sofferma, in particolare, sulle differenze tra legislazione secondaria direttamente applicabile (regolamenti e decisioni) e, dall'altro, decisioni-quadro e direttive soggette al processo di attuazione negli ordinamenti giuridici nazionali per la loro applicazione. L'uso della forma *shall* è più frequente in queste ultime, mentre le forme *should* e *must* sono equamente distribuite tra i quattro sotto-generi da lei considerati.

¹⁸ Tra il 2005 e il 2010 il gruppo che si occupa delle tecniche redazionali (*Drafting Techniques Group*) dell'Office of Parliamentary Counsel ha pubblicato delle linee guida finalizzate alla modernizzazione del linguaggio giuridico in prospettiva di *Plain language* che prevedono, tra le altre, l'eliminazione di espressioni arcaiche e latine, la riduzione della nominalizzazione e del passivo, l'uso di soluzioni neutrali in base al genere (cfr. Williams 2011).

¹⁹ La critica nei confronti dell'uso di questo modale dipende dalla sua ambiguità: in *Old English*, *shall*, insieme a *should*, indicava degli obblighi fisici e morali; in *Middle English* era usato (alla prima persona) per esprimere promesse e intenzioni, arrivando in seguito a codificare il tempo futuro ed essendo oggi percepito come formale e datato (Felici 2012: 53).

presentazione di *shall* nel Corpus A dell'EOMC rispetto al Corpus B. A questo proposito è da notare che nella redazione in inglese delle direttive europee tale forma è usata per imporre un forte obbligo sul destinatario, in quanto il suo impiego è immediatamente compreso da traduttori e traduttrici che lavorano nelle altre lingue dell'UE: "*shall* has value in a multilingual contest as its function is well understood as normative and it is a clear and unambiguous sign, whereas the present indicative already has a factual signification and there could be ambiguity" (Robertson 2010: 156).

Per quanto riguarda la codifica della prescrittività nei testi legislativi in francese si considerino le seguenti considerazioni:

“L’indicatif présent suffit non seulement à exprimer le droit, mais, plus spécifiquement, à marquer l’obligation. Le verbe devoir – le devoir – est sous-entendu. [...] L’indicatif remplace l’impératif grammatical. L’indicatif vaut l’impératif. C’est une particularité de l’énoncé législatif”. (Cornu 2005: 268)

Date queste premesse, può essere interessante verificare quali siano in italiano le varianti selezionate per la resa delle forme con valenza deontica adottate nel testo primario della direttiva, originariamente redatto in lingua inglese o in francese²⁰.

Per quanto riguarda il confronto intercorpora (A-B-C) sulla distribuzione dei verbi modali che codificano l’obbligo, il divieto e il permesso così come di strutture impersonali prescrittive, esso è già stato oggetto di uno studio sulle dinamiche di variazione connesse all’uso dei modali deontici *dovere* e *potere* all’interno dei tre sotto-generi testuali, quale espressione di particolari esigenze extra-linguistiche nella diversa cornice giuridica di riferimento (Mori 2020).

²⁰ In questa prospettiva, lo studio contrastivo (francese-inglese-italiano-tedesco) condotto da Felici (2013: 8) sulle rese di *shall* in regolamenti dell’UE, emanati tra 2001 e 2004, ha dimostrato la sua sovrarappresentazione nel diritto europeo rispetto a quello nazionale. Dal confronto è emerso che nell’80% dei casi i traducenti utilizzati sono forme flesse al presente indicativo mentre solo il 5% delle occorrenze di *shall* è reso mediante modali d’obbligo (*devoir; dovere, müssen/sollen*). L’analisi delle concordanze da lei condotta ha inoltre suggerito usi più “costitutivi”, ovvero performativi, di *shall* al fine di sancire diritti e garantire poteri (*avoir le droit; avere il diritto, avere il potere; das Recht/ den Anspruch haben*).

Per le finalità di questo studio, nel § 3.1 sono considerati i dati percentuali provenienti dall'annotazione morfosintattica²¹ ottenuti dal monitoraggio automatico del profilo linguistico dei tre corpora con metodiche NLP (*Natural Language Processing*) in modo tale da evincere informazioni quantitative legate alla rappresentazione di scenari deontici (Tabelle 1-3). I risultati discussi nei successivi paragrafi (da § 3.2 a § 5.2) sono estratti con l'uso di metodiche *corpus-based* e *corpus-driven* utilizzando il programma di analisi lessico-sintattica WordsSmith Tools 6.0. e presentati come valori normalizzati a un milione (Tabelle 4-10).

3.1. *Le forme del verbo*

Come già detto, per loro natura, le direttive mirano a costruire uno scenario deontico che definisce i confini entro i quali ciascun Stato Membro è chiamato a raggiungere obblighi generali, ovvero gli obiettivi prefissati a livello europeo, lasciando aperte le possibilità per una loro applicazione. La codifica della modalità deontica risulta, infatti, garantita dall'uso del cosiddetto presente prescrittivo e, in minima parte, dal futuro iussivo (Tabella 1). Ciò motiva l'alta percentuale d'uso, in tutti e tre i corpora, del modo indicativo, seppur in misura superiore nelle direttive europee del Corpus A come si vede nella Tabella 1 a seguire.

In generale, il confronto rivela come nei corpora nazionali, sia di derivazione europea (Corpus B) che di origine nazionale (Corpus C), il ricorso a forme participiali nell'articolato sia superiore, probabilmente in ragione della maggiore complessità sintattica che caratterizza l'italiano legislativo nazionale rispetto all'euroletto italiano²².

²¹ Il modulo di annotazione automatica morfo-sintattica descritto da Dell'Orletta (2009) dimostra un'accuratezza del 96,34% nell'identificazione simultanea della categoria grammaticale e dei tratti morfologici associati.

²² In precedenza si è evidenziata la correlazione tra complessità sintattica dei testi contenuti nei tre corpora legislativi e loro indice di leggibilità sintattica: "i risultati ottenuti mostrano una chiara correlazione tra i vari parametri disposti lungo una direzionalità che va da [- complesso], per il Corpus A, a [+ complesso] per i corpora B e C. [...] Questo si evince anche dai valori di leggibilità del modulo base di READ-IT che identificano il corpus dei dispositivi nazionali di attuazione come significativamente meno leggibile rispetto alle direttive e anche – seppur in misura marginale – alle leggi e ai decreti di derivazione nazionale" (Mori 2019c: 650-651).

Modo	Corpus A	Corpus B	Corpus C
Condizionale	0,04	0,02	0,00
Congiuntivo	3,30	4,90	4,30
Gerundio	0,90	1,60	2,10
Infinito	15,00	16,60	17,40
Indicativo	33,00	22,20	20,20
Imperativo	0,17	0,50	1,00
Participio	26,20	34,80	35,50

Tab. 1. Distribuzione percentuale dei modi verbali

Dalla Tabella 2 qui di seguito si può evincere come, per innescare scenari deontici all'interno dei quali non si attua né si modifica un nuovo stato di cose, ma si definisce un obbligo, siano quasi esclusivamente utilizzate le forme al presente in tutti e tre i corpora. Tuttavia, nei corpora nazionali (B e C) si rileva anche il ricorso, seppur circoscritto, al futuro iussivo che emerge, in particolare, nelle norme di origine italiana del Corpus C:

Tempo	Corpus A	Corpus B	Corpus C
Presente	98,69	96,91	92,54
Futuro	0,85	2,51	6,83

Tab. 2. Distribuzione percentuale del tempo presente e futuro

Nei pochi casi in cui nelle direttive del Corpus A si riscontra la presenza del futuro (es. 1), questo co-occorre con un esplicito riferimento indessicale, che permette di collocarle nel tempo, o con l'esplicitazione di un rimando intertestuale a evidenziare la cornice giuridica di riferimento e la chiara funzione iussiva, come si può notare dagli esempi che seguono:

- (1) Le modalità di applicazione del presente articolo **saranno stabilite entro il 1° gennaio 2002** secondo la procedura di cui all'articolo 11. (Direttiva 1999/74/CE)

Le disposizioni di cui all'articolo 1 **saranno rivedute** per la prima volta entro il 31 dicembre 2004 alla luce di nuovi dati sulla presenza di diossina e di PCB diossina-simili, in particolare al fine di includere i PCB diossina-simili nei livelli da fissare. (Direttiva 2001/102/CE)

La portata dell'esenzione **sarà definita facendo riferimento alle pertinenti disposizioni**. La durata di un esperimento non può essere superiore a sette anni. (Direttiva 2002/55/CE)

Nelle norme nazionali, sia di derivazione europea che di origine nazionale, si osserva un uso del futuro iussivo quantitativamente più rilevante rispetto al Corpus A e senza un esplicito ancoraggio temporale. Negli esempi che seguono, gli enunciati estratti dal Corpus B (es. 2) e dal Corpus C (es. 3) codificano atti direttivi in grado di proiettare le aspettative in termini di comportamenti futuri in capo ai destinatari della norma, mediante il ricorso alla diatesi passiva.

- (2) Il presente decreto, munito del sigillo dello Stato, **sarà inserito** nella Raccolta ufficiale degli atti normativi della Repubblica italiana. (Decreto Legislativo 4 marzo 2010 n. 28)

Con provvedimento del Ministero delle politiche agricole alimentari e forestali, da adottarsi entro sei mesi dalla data di entrata in vigore del presente decreto, **saranno emanate** disposizioni applicative per stabilire le modalità per l'ammissione al Registro nazionale delle varietà da conservazione. (Decreto legislativo 29 ottobre 2009 n. 149)

- (3) Nel caso di progetti per i quali la competenza allo svolgimento della valutazione ambientale spetta alle regioni, **si provvederà** con la pubblicazione su un quotidiano a diffusione regionale o provinciale. (Decreto legislativo 3 aprile 2006 n. 152)

Il Tribunale Arbitrale **delibererà** con voto di maggioranza e la sua decisione **sarà** vincolante. Le due Parti Contraenti **sosterranno** le spese relative al proprio arbitrato e al proprio rappresentante per le

udienze. Le spese relative al Presidente e tutte le altre spese **saranno** equamente **suddivise** fra le Parti Contraenti. Il Tribunale Arbitrale **stabilirà** le proprie procedure. (Legge 19 novembre 2010 n. 209)

Per quanto riguarda le differenze intercorpora legate alla codifica del numero alla terza persona, il confronto in Tabella 3 qui di seguito evidenzia una preferenza del singolare nei testi normativi in italiano legislativo nazionale, mentre la terza persona plurale è maggiormente attestata in euroletto italiano.

Numero	Corpus A	Corpus B	Corpus C
3a pers. sing.	24,93	39,63	38,49
3a pers. plur.	41,77	22,80	18,30

Tab. 3. Distribuzione percentuale del numero

Tale dato è facilmente riconducibile alla sovrarappresentazione nel Corpus A di europeismi sintagmatici con cui si designano referenti plurali (*Stati membri, autorità competenti, Paesi terzi*) rispetto ai soggetti singolari che caratterizzano la legislazione nazionale. Infatti, nei corpora B e C, sono particolarmente frequenti referenti testuali quali *decreto legislativo, Camera dei deputati, Presidente della Repubblica, Ministro, Senato della Repubblica*.

3.2. La distribuzione del presente indicativo

Le analisi *corpus-driven* sull'EOMC hanno evidenziato la ricorrenza di forme al presente prescrittivo all'interno di *lexical bundles*²³ che contribuiscono “alla costruzione di una testualità altamente codificata e formulaica, agendo sia come marche di registro che di dominio così evidenziando la loro rilevanza nei linguaggi specialistici” (Mori 2019b: 86). In particolare si può notare come tali combinazioni ricorrenti di parole si configurino come *content bundles*, composti da cinque o sei ele-

²³ L'estrazione dei *lexical bundles* è avvenuta mediante l'uso della funzione N-Grams del programma di analisi lessico-sintattica WordSmith Tools 6.0.

menti, dove le forme al presente prescrittivo sono utilizzate per riferirsi a comportamenti che gli Stati membri sono chiamati a porre in essere e, pertanto, sono spesso corredate da avverbi temporali o da deittici di tempo assoluti (data) o relativi per il riferimento al campo indicale, del tipo “il giorno successivo a [...]”.

L’analisi quantitativa *corpus-based* che segue è stata condotta partendo dall’assunto che in italiano le forme al presente indicativo possono assumere sia un valore deontico che performativo, senza alcuna variazione superficiale. A questo proposito Garzone (2008: 75) specifica che:

quando la norma viene esercitata su un determinato soggetto a cui è affidata la sua realizzazione, l’enunciato ha per lo più valore prescrittivo, mentre nei casi in cui la sua realizzazione è immediata, nominando essa uno stato di cose che viene attuato contestualmente al proferimento della norma stessa, essa assume valore performativo.

I verbi al presente qui selezionati sono tra quelli tipicamente utilizzati in enunciati legislativi direttivi o dichiarativi per la costruzione della cornice giuridica di riferimento, il cosiddetto *legal framing*. Ai fini dello studio sarà interessante considerare anche la distribuzione delle forme alla diatesi attiva (Tabella 4) e alla diatesi passiva (Tabella 5) dei seguenti verbi alla terza persona singolare e plurale: adottare, applicare, approvare, autorizzare, comunicare, consentire, disciplinare, disporre, garantire, informare, inserire, istituire, modificare, permettere, proibire, provvedere, punire²⁴, stabilire, trasmettere.

A questa prima fase *corpus-based* seguirà l’analisi qualitativa del profilo collocazionale dei verbi in esame, che ci consentirà di osservare la fisionomia degli enunciati legislativi: direttivi, con valenza deontica, o dichiarativi, con valenza performativa (§ 3.3).

Nelle Tabelle che seguono si presentano i risultati estratti dai tre corpora dell’EOMC per procedere a un confronto tra le varietà legislative d’italiano nella distribuzione dei verbi selezionati. I valori riportati sono stati normalizzati a 1 milione in modo tale da rendere comparabili i dati provenienti da corpora di diversa grandezza (§ 2).

²⁴ Del verbo “punire” non si registrano occorrenze alla forma attiva del presente (3 pers. sing. e pl.).

Forma verbale	Corpus A	Corpus B	Corpus C
<i>adotta</i>	323	158	82
<i>adottano</i>	1023	72	48
<i>applica</i>	0	43	20
<i>applicano</i>	103	25	13
<i>approva</i>	8	27	44
<i>approvano</i>	8	6	7
<i>autorizza</i>	26	32	21
<i>autorizzano</i>	49	7	1
<i>comunica</i>	119	149	42
<i>comunicano</i>	0	66	26
<i>consente</i>	58	45	36
<i>consentono</i>	71	51	33
<i>disciplina</i>	30	283	358
<i>disciplinano</i>	70	55	21
<i>dispone</i>	57	136	132
<i>dispongono</i>	103	36	19
<i>garantisce</i>	65	60	20
<i>garantiscono</i>	141	43	16
<i>informa</i>	322	163	36
<i>informano</i>	412	36	7
<i>inserisce</i>	2	4	1
<i>inseriscono</i>	2	3	1
<i>istituisce</i>	78	33	14
<i>istituiscono</i>	40	12	7
<i>modifica</i>	231	58	10
<i>modificano</i>	24	17	5
<i>permette</i>	19	10	0
<i>permettono</i>	16	11	2
<i>proibisce</i>	0	0	0
<i>proibiscono</i>	1	0	0
<i>provvede</i>	89	451	523
<i>provvedono</i>	602	175	102
<i>stabilisce</i>	148	152	80
<i>stabiliscono</i>	112	20	15
<i>trasmette</i>	97	139	77
<i>trasmettono</i>	76	64	24
Totale	4.525	2.642	1.843

Tab. 4. Distribuzione forme attive al presente indicativo
(frequenze normalizzate a 1 milione)

Dall'analisi *corpus-based* sul campione di verbi selezionati, il Corpus A si riconferma per la sovrarappresentazione del presente indicativo (Mori 2018c), in particolare ciò risulta evidente per alcune forme verbali (*adotta/adottano, applicano, dispongono, garantiscono, informal/informano, modifica/modificano, provvedono, stabiliscono, trasmettono*). La concentrazione di alte frequenze di *tokens* su un numero più circoscritto di *types* mostra una minor variabilità lessicale nelle direttive, dovuta anche alla presenza di porzioni di frasi che si ripetono identiche o con una minima variazione (Mori 2019c). Negli altri due corpora si rileva una diversa distribuzione delle forme verbali maggiormente rappresentate nelle misure nazionali di attuazione (*adotta, comunica, dispone, stabilisce, trasmette*) e nelle leggi italiane di origine nazionale (*approva, disciplina, provvede*). Dal confronto si delinea una differenza di valori frequenziali tra i corpora A e B, da un lato, e il Corpus C, dall'altro per quanto riguarda tali verbi utilizzati per designare il compimento di un'azione che si deve/dovrà compiere al fine di garantire e rendere operativo quanto richiesto.

Nella Tabella 5 a seguire si forniscono i dati relativi all'incidenza della diatesi passiva per i medesimi verbi rispetto all'uso del presente indicativo. La scelta della forma attiva o passiva dipende dalla diversa funzione comunicativa che le due costruzioni permettono di realizzare, dunque, da considerazioni pragmatiche piuttosto che sintattiche. È noto, infatti, come nel discorso giuridico, l'uso del passivo, in cui tra l'altro l'agente viene spesso ommesso, è un espediente che permette al legislatore di codificare permessi e comandi con la massima oggettività.

<i>È/sono</i>	Corpus A	Corpus B	Corpus C
<i>adottato/a/i/e</i>	297	127	106
<i>applicato/a/i/e</i>	44	39	32
<i>approvato/a/i/e</i>	13	30	75
<i>autorizzato/a/i/e</i>	117	166	585
<i>comunicato/a/i/e</i>	47	83	83
<i>consentito/a/i/e</i>	21	116	80

<i>disciplinato/a/i/e</i>	31	86	192
<i>disposto/a/i/e</i>	10	45	114
<i>garantito/a/i/e</i>	10	16	13
<i>informato/a/i/e</i>	40	17	5
<i>inserito/a/i/e</i>	188	518	468
<i>istituito/a/i/e</i>	15	96	131
<i>modificato/a/i/e</i>	549	39	16
<i>permesso/a/i/e</i>	3	2	1
<i>proibito/i/a/e</i>	4	3	0
<i>punito/a/e/i</i>	0	59	213
<i>sostituito/a/i/e</i>	982	1.172	1.500
<i>stabilito/a/i/e</i>	156	183	290
<i>trasmesso/a/i/e</i>	49	90	79
Totale	2.576	2.887	3.983

Tab. 5. Distribuzione forme passive al presente indicativo
(frequenze normalizzate a 1 milione)

Per quanto riguarda il rapporto tra diatesi attiva e passiva nei tre corpora, le frequenze delle forme verbali hanno chiaramente mostrato come ciascun verbo selezionato si distribuisce rispetto a questo tratto con cui si rappresenta la pragmatica giuridica. Ciò ha consentito anche di individuare una direzionalità variabile lungo la quale si posizionano le varie forme verbali, mettendo in evidenza, talvolta, la vicinanza tra l'italiano legislativo delle direttive (A) e delle leggi nazionali di attuazione (B), talaltra una maggior affinità tra leggi nazionali di attuazione di derivazione europea (B) e norme italiane (C). La correlazione, rispetto alla diatesi, evidenzia una netta preferenza delle forme attive al presente indicativo per il Corpus A, là dove per il Corpus C si registrano frequenze superiori al passivo per i verbi qui considerati. In particolare l'analisi *corpus-based* consente di

definire una geometria dei verbi più rappresentati al presente passivo (terza persona singolare o plurale) nel confronto tra i tre corpora: Corpus A (*adottare, applicare, modificare*), Corpus B (*consentire, inserire*), Corpus C (*approvare, autorizzare, disporre, punire, sostituire, stabilire*).

In generale, i risultati totali confermano frequenze superiori delle forme passive, e una diversa distribuzione dei verbi in correlazione con la diatesi come si evince confrontando i valori riportati nelle Tabelle 4 e 5. Da notare come in nessuno dei tre corpora si registrino occorrenze di *punire* alla forma attiva dell'indicativo presente (terza persona singolare e plurale), mentre per tutti i verbi alla forma passiva (terza persona singolare e plurale) si hanno attestazioni in almeno uno dei tre corpora. Inoltre, è sicuramente da segnalare la sovrarappresentazione delle forme passive nel Corpus C e una convergenza tra euroletto italiano e tra varietà legislativa di derivazione europea rispetto alle scelte lessicali per la creazione di scenari giuridici realizzati da agenti che non sono necessariamente esplicitati o che risultano vagamente identificati. All'interno di una cornice giuridica, di modalità deontica o performativa, la diatesi passiva contribuisce chiaramente ad attenuare il gradiente di agentività e di responsabilità nella formulazione del testo della legge, così come si osserva frequentemente nell'italiano giuridico-amministrativo²⁵.

4. *L'enunciazione legislativa al presente indicativo: tra prescrittivo e performativo*

L'analisi *corpus-based* condotta (§ 3.2) ha consentito di identificare la distribuzione di alcuni verbi che designano obblighi ma che, in diversi casi, possono (anche) avere la funzione di performativi eseguendo un'azione nel momento stesso in cui la evocano, dando così forma ad enunciati che si autorealizzano (Caffi 2009: 114). Questa classe di verbi ha un ruolo fondamentale soprattutto nei testi normativi dove lo stato del mondo generato da un performativo giuridico ha massima forza vincolante poiché il suo effetto ricade *erga omnes* (Fiorito 2006: 104).

Il confine tra il valore performativo e quello prescrittivo del presente indicativo richiede uno sforzo interpretativo quando si considera la varietà giuridica italiana, ma anche quella francese, soprattutto in una si-

²⁵ Cfr. Cortelazzo (2010).

tuazione di traduzione. A proposito della resa di *shall* nella traduzione in francese così si pronunciava Gémard (1981: 346-347): “Trop souvent shall est pris pour un futur quand il n’a que valeur emphatique. [...] la difficulté consiste à distinguer la forme comminatoire.”

Analogamente, per l’inglese *shall* (così come per il tedesco *sollen*), la sua valenza funzionale non è di immediata categorizzazione (Garzone 1999; Felici 2012; Sandrelli 2019). A questo proposito Garzone (1999: 73) sottolinea che:

la differenza tra i due tipi di discorso non è identificabile attraverso nessuna 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 per l’appunto con il presente indicativo.

La riflessione congiunta sulla codifica morfologica della prescrittività e della performatività in testi legislativi prodotti in diversi contesti giuridici di riferimento (europeo e nazionale) è particolarmente interessante proprio in ragione di una sovrapposizione tra forme linguistiche che in italiano, ma anche in inglese e in francese, possono esprimere entrambe le valenze senza alcuna variazione formale. Infatti, a seconda del contesto di enunciazione, i verbi considerati (Tabella 4 e 5) potrebbero esprimere tanto la valenza prescrittiva quanto quella performativa. Nel caso del Corpus A, occorre considerare che nelle direttive la semantica giuridica è relazionata alle proprietà di tale strumento, finalizzato all’imposizione di obblighi, che saranno oggetto di attuazione e, solo successivamente, resi applicativi.

Per poter discriminare la valenza funzionale si è andati innanzitutto alla ricerca di spie presenti nel contesto in cui tali verbi si inseriscono e si è rintracciata la presenza di riferimenti temporali espliciti che guidano verso un’interpretazione deontica. Infatti, la co-occorrenza dell’espressione deittica *entro* + deittico temporale, assoluto o relativo, proietta nel futuro l’azione obbligatoria da raggiungere. In questo senso *entro* può essere considerata una spia contestuale atta ad agevolare la disambiguazione di forme verbali potenzialmente di difficile categorizzazione²⁶ così come l’uso del modificatore *presente* o *tale* indirizza verso l’identificazione di una valenza performativa.

Al fine di distinguere gli usi prevalenti nei tre corpora, prescrittivi e/o performativi, si è deciso di partire dai risultati sulla distribuzione quantitativa (Tabelle 4 e 5) per procedere a una verifica delle concordanze associate alle forme verbali specifiche di ciascun corpus, così da desumere elementi utili a un'identificazione funzionale del presente indicativo all'interno degli enunciati legislativi. Tale fase di analisi di tipo *corpus-driven*, guidata dall'evidenza dei dati (Tognini Bonelli 2001: 84), è stata mirata alla raccolta di informazioni sul profilo collocazionale dell'elemento linguistico analizzato, ovvero una lista di concordanza con tutte le occorrenze dell'item nel formato KWIC (Key Word in Context).

4.1. *Analisi qualitativa delle forme attive al presente indicativo*

L'analisi che segue conferma la maggiore fissità formulaica che caratterizza, in generale, le direttive rispetto a una maggiore variabilità dei corpora legislativi nazionali, sia come gamma di azioni obbligatorie in riferimento alle diverse autorità competenti di volta in volta coinvolte, che per la necessità di adottare distinti strumenti legislativi. A seguire gli esempi mostrano l'uso delle forme più rappresentate per ognuna delle tre varietà di italiano legislativo: Corpus A (es. 4-5); Corpus B (es. 6-7); Corpus C (es. 8-9). Nel caso delle direttive, l'analisi del profilo collocazionale dei verbi quantitativamente più frequenti ha mostrato come *Stati membri* sia il soggetto sintattico più ricorrente, ovvero i primi destinatari dell'atto, chiamati ad adempiere agli obblighi imposti dalla direttiva stessa dove il presente indicativo assume una valenza deontica (es. 4). Quando il soggetto dell'enunciato è, invece, la *direttiva* o la *Commissione* al presente indicativo è attribuita una valenza performativa (es. 5).

- (4) Per conformarsi agli obiettivi fissati dalla presente direttiva, gli *Stati membri* **adottano** le misure necessarie per realizzare i seguenti obiettivi su tutto il loro territorio. (Direttiva 2004/10/CE)

Gli *Stati membri* ne **informano** immediatamente la Commissione. (Direttiva 2003/114/CE)

²⁶ Dal confronto delle occorrenze di *entro* nei tre corpora non si evincono differenze rilevanti in termini di frequenze normalizzate: Corpus A (1.368); Corpus B (1.512); Corpus C (1.474).

Gli *Stati membri* **provvedono** affinché tali autorità esercitino le loro funzioni in modo imparziale ed indipendente da tutte le parti dell'offerta. (Direttiva 2004/10/CE)

- (5) La *Commissione* **adotta**, secondo la procedura prevista dall'articolo 21, le disposizioni necessarie affinché si effettuino le prove e le analisi comparative. (Direttiva 2003/61/CE)

La presente *direttiva* **stabilisce** diritti ed obblighi per gli operatori e per le imprese che intendono interconnettersi e/o avere accesso alle loro reti o a risorse correlate. (Direttiva 2002/19/CE)

Nel Corpus B, le forme verbali più frequenti sono flesse alla terza persona singolare con valenza deontica in co-occorrenza di un soggetto animato, seppur non prototipico, come il *Ministero* o l'*autorità competente*, e il vincolo implicito a una proiezione futura (es. 6). Là dove il soggetto è inanimato, e in presenza della spia cotestuale *presente*, si configurano enunciati di tipo dichiarativo dove il presente acquisisce una valenza performativa (es. 7).

- (6) Il *Ministero* **comunica** inoltre annualmente tali informazioni alla Commissione europea, su apposita richiesta. (Decreto legislativo 27 marzo 2006 n. 161)
- (7) Il presente *decreto* **disciplina** il coinvolgimento dei lavoratori nelle attività delle società per azioni, di seguito denominata [...] (Decreto legislativo 19 agosto 2005 n. 188)

Il *presente decreto* legislativo **stabilisce** le condizioni per l'esercizio del diritto al ricongiungimento familiare dei cittadini di Paesi terzi, legalmente soggiornanti nel territorio dello Stato italiano, in applicazione della direttiva 2003/86/CE del Consiglio, del 22 settembre 2003. (Decreto legislativo 8 gennaio 2007 n.5)

Nel Corpus C, i verbi con le più alte frequenze di occorrenza si attestano in enunciati direttivi (es. 8) dove è esplicitata l'autorità deputata all'adempimento di un determinato dovere imposto dalla legge. Diversamente, negli esempi successivi (es. 9) si configurano enunciati legislativi di tipo dichiarativo, con un soggetto impersonale o inanimato, con cui avviene un'autorealizzazione di quanto detto.

- (8) A tale fine, entro il 30 novembre di ciascun anno precedente, il *presidente* dell'ente **trasmette** la proposta di accordo al Ministro dell'economia e delle finanze. (Legge 4 agosto 2008 n. 132)

Se la conciliazione non riesce, il *collegio* **provvede**, ove occorra, a interrogare le parti e ad ammettere e assumere le prove, altrimenti invita all'immediata discussione orale. (Legge 15 dicembre 1998 n. 465)

- (9) All'istituzione e al funzionamento della Commissione e della Conferenza di cui agli articoli 4 e 5 *si* **provvede** con le risorse umane, strumentali e finanziarie disponibili a legislazione vigente. [...] il presente decreto **disciplina** il riordino della struttura organizzativa e funzionale dell'Arma. (Decreto legislativo 5 ottobre 2000 n. 297)

4.2. *Analisi qualitativa delle forme passive al presente indicativo*

L'analisi *corpus-based*, condotta anche sulle forme passive del presente indicativo dei verbi in esame, ha rivelato differenze nella distribuzione quantitativa nei tre corpora, così come un rapporto variabile tra attivo e passivo per i medesimi verbi (Tabelle 4 e 5).

Andando ad analizzare il profilo collocazionale, nelle direttive i sostantivi che co-occorrono con *è adottata*, anche in posizione post-verbale, sono *misura o decisione*; *direttiva* è il collocato più frequente di *è modificata* anche se si registrano collocazioni in cui *è modificato* ha come collocato *articolo* o *allegato*. Degna di nota è la presenza di spie di performatività (*tale* e *come segue*) che creano un vincolo all'*hic et nunc* dell'enunciazione, come si evince dagli esempi (es. 10).

- (10) Tale misura, intesa a modificare elementi non essenziali della presente direttiva completandola, **è adottata** secondo la procedura di regolamentazione con controllo di cui all'articolo 24, paragrafo 3." (Direttiva 2008/12/CE)

La *direttiva* 91/447/CEE **è modificata** come segue [...] (Direttiva 2008/51/CE)

Nei corpora legislativi nazionali, le forme passive dei verbi portatori di semantica giuridica qui considerati prevedono il ricorso ad ancoraggi intertestuali o intratestuali che contribuiscono all'espressione della performatività in enunciati dichiarativi, come si può osservare negli esempi estratti sia dal Corpus B (es. 11), che dal Corpus C (es. 12).

- (11) **È consentito** l'ingresso e il soggiorno per motivi di studio, secondo le modalità stabilite nel regolamento di attuazione, dei cittadini stranieri. (Decreto legislativo 10 agosto 2007 n. 154)

All'articolo 11 del decreto legislativo 24 febbraio 1997, n. 46, e successive modificazioni, dopo il comma 12 **è inserito** il segunte [comma...]. (Decreto legislativo 31 ottobre 2002 n. 271)

- (12) Il *Ministro dell'economia e delle finanze* **è autorizzato** ad adottare, con propri decreti, le necessarie variazioni di bilancio, nel rispetto delle quote come annualmente determinate. (Decreto legislativo 15 gennaio 2002 n. 9)

L'articolo 17 del decreto legislativo numero 58 del 1998 **è sostituito** dal segunte [articolo] (Decreto legislativo 6 febbraio 2004 n. 37)

Le *norme* disciplinanti l'ordinamento del Comando operativo di vertice interforze **sono stabilite** nel regolamento. (Decreto legislativo 28 gennaio 2008 n. 35)

Occorre segnalare che, dal confronto tra diatesi attiva e passiva, il verbo *punire* si registra solo in costrutti passivi, con cui si determina un abbassamento del grado di agentività, ed esclusivamente nei corpora B e C. Nello specifico, *è punito* caratterizza le leggi di origine nazionale dove occorre con un soggetto sintattico indefinito come *chiunque* e con un ancoraggio contestuale che presuppone un certo grado di futurità:

- (13) Salvo che il fatto costituisca più grave reato, *chiunque* compie atti violenti diretti e idonei a sottoporre il territorio dello Stato o una parte di esso alla sovranità di uno Stato straniero, ovvero a menomare l'indipendenza o l'unità dello Stato, **è punito** con la reclusione non inferiore a dodici anni. (Legge 13 marzo 2006 n.85)

Inoltre, si evidenzia un uso del *si* passivante (*si applica*)²⁷, in particolare con il sostantivo *disposizione*, *pena*, *sanzione*, che mostra un'incidenza maggiore nelle leggi nazionali di attuazione (es. 14). Per quanto riguarda la forma al plurale *si applicano*²⁸, concordata con sostantivi quali *disposizioni*, *modalità*, *norme*, questa si registra maggiormente nelle leggi italiane di origine nazionale (es. 15).

²⁷ Frequenze normalizzate a 1 milione: Corpus A (383), Corpus B (416), Corpus C (334).

²⁸ Frequenze normalizzate a 1 milione: Corpus A (473), Corpus B (647), Corpus C (699).

- (14) La stessa *sanzione si applica* a chi con più azioni od omissioni, esecutive di un medesimo disegno, commette anche in tempi diversi più violazioni della stessa o di diverse disposizioni di cui al presente articolo. (Decreto legislativo 30 gennaio 2001 n. 94)
- (15) Le *disposizioni* di cui al presente decreto **si applicano** ai procedimenti disciplinari promossi a decorrere dalla data della sua entrata in vigore. (Legge 24 ottobre 2006 n. 269)

In questi casi il ricorso alla diatesi passiva “spersonalizza e, quando si ometta l’agente, consente una maggiore astrattezza” (Mantovani 2008: 41-42), così da realizzare una proprietà che contraddistingue il linguaggio delle leggi al fine di garantirne l’efficacia e la durata nel tempo, a prescindere dall’evoluzione di costumi, conoscenze o valori.

5. *La performatività come proprietà costitutiva delle norme*

La predominante funzione conativa delle leggi non si manifesta solo attraverso la codifica morfologica e morfo-sintattica della modalità deontica, ma anche mediante l’espressione della performatività. Con la nozione di performatività si rimanda a enunciati legislativi che attuano contestualmente lo stato di cose che esprimono, ponendo in atto una nuova situazione, instaurando nuovi rapporti (nel nostro caso giuridici²⁹) o alterando quelli esistenti.

Rispetto al già citato effetto costitutivo delle norme (§ 2), considerato come una forma particolarmente forte di performatività, Conte (1994: 249) propone di procedere a una differenziazione tra enunciati legislativi deontici ed enunciati legislativi thetici (dal greco “atto di posizione”) che attuano uno stato di cose e, per questo, definiti thetico-costitutivi³⁰. In tale studio Conte elabora una distinzione all’interno della categoria dei performativi, tra thetici e athetici: è thetica la performatività delle enunciazioni che attuano uno stato di cose, mentre è athetica (o non-thetica) la performatività delle enunciazioni che realizzano un atto. Per gli enunciati performativi (in senso giuridico), o thetico-costitutivi, l’effetto giuridico è mediato dal-

²⁹ Per una trattazione dettagliata della nozione di performatività in riferimento al linguaggio giuridico inglese e italiano si rimanda a Garzone (1996).

³⁰ Per un approfondimento si può consultare Roversi (2012).

l'esercizio del potere, della norma stessa, di "produrre immediatamente modificazioni, di creare realtà giuridiche, quale ne sia la specie, in virtù di una norma che tali modificazioni e realtà costituisca in dipendenza dal potere stesso" (Carcaterra 1974: 82). Pertanto si può osservare come sia la dimensione pragmatica a consentire di attribuire la valenza comunicativa di un determinato segno linguistico all'interno di un testo giuridico: "as part of a social context, relations among signs are given by 'agents', by the type of 'norm-action' they are meant to state or perform, and by the 'situational context' in which they take place" (Felici 2012: 64).

5.1. *La dimensione performativa degli enunciati legislativi*

Al fine di approfondire la riflessione sulla costitutività delle norme, in questa sezione si considera la distribuzione di tre verbi con una valenza performativa giuridica (*abrogare, emanare, vietare*) per esprimere l'esercizio di un'autorità che si pronuncia su uno stato del mondo, con lo scopo di produrre un effetto sul destinatario. A questo proposito è da notare come in italiano non sia possibile operare una netta distinzione formale tra verbi performativi giuridici utilizzati in enunciati direttivi, né in correlazione al genere normativo (sia esso una legge, un decreto legislativo o un decreto legge), né rispetto all'autorità, al ruolo e alla funzione del soggetto enunciante e firmatario (cfr. Zuanelli 1998; 2000).

Nelle Tabelle (6 e 7) che seguono sono presentate le frequenze delle forme attive e passive di tali verbi nei tre corpora così da evidenziare analogie e differenze tra le varietà legislative di italiano.

Forma verbale	Corpus A	Corpus B	Corpus C
<i>abroga</i>	4	20	1
<i>abrogano</i>	3	0	0
<i>emana</i>	1	143	22
<i>emanano</i>	2	1	3
<i>vieta</i>	8	17	1
<i>vietano</i>	44	4	0
Totale	62	185	27

Tab. 6. Distribuzione verbi performativi giuridici alla diatesi attiva (frequenze normalizzate a 1 milione)

<i>È/sono</i>	Corpus A	Corpus B	Corpus C
<i>abrogato/a/i/e</i>	136	198	214
<i>emanato/a/i/e</i>	1	30	37
<i>vietato/a/i/e</i>	41	112	32
Totale	178	340	283

Tab. 7. Distribuzione verbi performativi giuridici alla diatesi passiva (frequenze normalizzate a 1 milione)

Dal confronto intercorpora si nota una sovrarappresentazione delle forme attive nelle leggi nazionali di attuazione, rispetto agli altri due corpora (Tabella 6). Per quanto riguarda la diatesi passiva (Tabella 7), i valori di frequenza d'uso ne delineano una preferenza nelle leggi nazionali, in particolare in quelle di attuazione (cfr. Tabella 4 e 5). In generale, la distribuzione delle forme attive risulta sottorappresentata rispetto al passivo in tutti e tre i corpora; tuttavia la maggior differenza si nota per il Corpus C dove il valore delle forme al passivo per gli stessi verbi risulta decuplicato rispetto a quello alla diatesi attiva. Nelle direttive europee e nelle leggi di attuazione il rapporto, tra diatesi attiva e passiva, è dello stesso segno per i performativi giuridici qui considerati: le forme passive sono circa il triplo per il Corpus A e quasi il doppio nel caso del Corpus B.

A partire da tali risultati si propongono alcuni esempi contenenti le forme, attive e passive, più ricorrenti in ognuno dei tre corpora: Corpus A (es. 16), Corpus B (es. 17) e Corpus C (es. 18).

- (16) Gli Stati membri **vietano** gli scambi di prodotti non conformi alla direttiva 2000/13/CE, come modificata dalla presente direttiva, con decorrenza 1° gennaio 2003. (Direttiva 2001/101/CE)

Il regolamento (CE) n. 925/1999 è **abrogato** con effetto dalla data di entrata in vigore della presente direttiva. (Direttiva 2002/30/CE)

- (17) Il Presidente della Repubblica **emana** il seguente decreto legislativo [...] (Decreto legislativo 24 giugno 2003 n. 211)

Il decreto del Presidente della Repubblica 21 dicembre 1996, n. 698, è **abrogato**. (Decreto legislativo 7 luglio 2011 n. 124)

È **vietato** utilizzare detti nomi per indicare prodotti che non siano conformi alla relativa definizione. (Decreto legislativo 12 giugno 2003 n. 178)

- (18) Entro il termine del periodo sperimentale di cui al comma 2, il Ministro delle attività produttive **emana** un regolamento per il funzionamento del sistema telematico delle borse merci italiane. (Decreto legislativo 28 agosto 2000 n. 274)

All'articolo 4 del decreto legislativo 29 gennaio 1998, n. 19, il comma 3 è **abrogato**. (Decreto legislativo 8 gennaio 2004 n. 1)

Inoltre, è interessante rilevare la presenza di costrutti morfosintattici del tipo “verbo+sostantivo” con cui ci si pronuncia sullo stato di cose e si sanciscono diritti, come si può notare dalla Tabella 8 che segue:

Frame	Collocazione giuridica	Corpus A	Corpus B	Corpus C
verbo+sostantivo	<i>costituisce (grave) causa/danno/discriminazione/ titolo preferenziale/violazione</i>	20	236	292
	<i>ha/hanno diritto</i>	29	81	99
	<i>ha/hanno facoltà</i>	22	41	66
Totale		71	358	457

Tab. 8. Distribuzione di collocazioni costitutive (normalizzazione a 1 milione)

Dal confronto tra i corpora si evidenzia come tali collocazioni costitutive, riportate tra le caratteristiche del linguaggio giuridico italiano (cfr. Rovere 2005), siano specifiche dell'italiano legislativo nazionale e, in misura più ridotta, anche delle leggi italiane di attuazione ma non siano distintive dell'euroletto italiano. Tale differenza, nella distribuzione di sequenze morfosintattiche con cui si esprime l'effetto costitutivo della

norma, conferma come la dimensione performativa sia più ridotta nelle direttive dove si manifesta maggiormente la modalità deontica per ragioni di natura contestuale che definiscono tale sotto-genere legislativo.

5.2. *Tra costitutività thetica e modalità deontica*

Dall'analisi quantitativa e qualitativa condotta sui verbi con rilevanza giuridica maggiormente rappresentati nelle tre varietà legislative dell'EOMC (§ 3.2 e § 4) sono emersi casi in cui gli enunciati legislativi possono essere contestualmente categorizzati come performativi, o meglio thetico-costitutivi, e deontici. Si può, infatti, osservare come le forme al presente indicativo possano imporre obblighi che si realizzano nel momento dell'enunciazione purché ciò avvenga nel rispetto di determinate condizioni contestuali necessarie alla realizzazione dell'atto esercitativo. Come osserva Felici (2013: 5):

This component of actualisation establishes a direct link between the utterance and the act that is to be performed and also explains why present indicative is to be regarded as the tense of law for its features of being constantly speaking.

Gli esempi che seguono mirano a identificare empiricamente la presenza di circostanze in cui la forza costitutiva dell'enunciato legislativo viene ad essere vincolata o soggetta a restrizioni. Come si può notare dall'esempio che segue, l'enunciato legislativo estratto dalla direttiva diventa costitutivo nel rispetto di condizioni che sono esplicitamente evocate mediante una subordinata condizionale:

- (19) La **presente** direttiva **si applica** a tutti i cittadini di paesi terzi ed agli apolidi che presentano domanda di asilo alla frontiera o nel territorio di uno Stato membro, purché siano autorizzati a soggiornare in tale territorio in qualità di richiedenti asilo, nonché ai familiari già definiti all'articolo 2, lettera d), se inclusi nella domanda di asilo a norma del diritto internazionale. (Direttiva 2003/9/CE)

Un confronto *corpus-based* ha mostrato come il connettivo *purché* si attesti prevalentemente nelle direttive e, in misura inferiore, nelle leggi di attuazione mentre non è una variante significativa nelle leggi di origi-

ne nazionale. Ciò si evince chiaramente dalle seguenti frequenze normalizzate: Corpus A (335); Corpus B (195); Corpus C (93). Tale dato sembrerebbe delineare una maggiore presenza di subordinate condizionali esplicite, attivate con l'uso di *purché*, nell'euroletto italiano delle direttive, là dove la valenza performativa dell'enunciato legislativo dipende dal rispetto di vincoli contestuali che ne limitano l'efficacia immediata. Andando a confrontare la frequenza d'uso di collocazioni restrittive che esercitano, analogamente, la funzione di ancorare la singola norma all'interno dell'universo del discorso giuridico, nel Corpus A l'uso di *legal reasoning bundles* (Mori 2019b) è molto più ridotto rispetto ai corpora B e C. L'analisi *corpus-driven* condotta sui *lexical bundles* più frequenti nei tre corpora rivela, infatti, come tali combinazioni ricorrenti di parole sono sovrarappresentati nei corpora nazionali. La Tabella 9 a seguire mostra una variabilità distribuzionale anche di natura qualitativa tra le tre varietà legislative: l'uso preponderante di *fatto salvo* (*quanto*) nelle direttive (es. 20), *fermo restando* nelle leggi nazionali di attuazione (es. 21) e del costrutto *previo* + N nelle leggi di origine italiana (es. 22).

Legal lexical bundles	Corpus A	Corpus B	Corpus C
<i>fatto/a/e/i salvo</i>	229	166	81
<i>fermo restando</i>	19	192	164
<i>ove non diversamente previsto</i>	0	2	42
<i>previo accertamento/assenso /esame/concorso/giudizio</i>	38	116	180
Totale	286	476	467

Tab. 9. Distribuzione di collocazioni restrittive
(frequenze normalizzate a 1 milione)

- (20) **Fatta salva** la direttiva 98/34/CE, gli Stati membri notificano tali misure speciali alla Commissione e agli altri Stati membri. (Direttiva 2004/108/CE)

- (21) **Fermo restando** quanto disposto dal comma 5, l'amministratore e il liquidatore esercitano in Italia gli stessi poteri che hanno il diritto di esercitare nel proprio Stato membro d'origine. (Decreto legislativo 9 aprile 2003 n. 93)
- (22) Il verbale di accordo, il cui contenuto non è contrario all'ordine pubblico o a norme imperative, è omologato, su istanza di parte e **previo accertamento** anche della regolarità formale, con decreto del presidente del tribunale nel cui circondario ha sede l'organismo. (Decreto legislativo 4 marzo 2010 n. 28)

Nello specifico la resa eurolettale *fatto salvo* (es. 20) è attestata nella legislazione europea come traduce del francese *sans préjudice* e/o dell'inglese *without prejudice* che si rintracciano nelle corrispondenti versioni delle direttive. Nella riformulazione intralinguistica, da euroletto italiano a italiano legislativo in uso nella legge nazionale di attuazione, il connettivo *fatto salvo*³¹ viene generalmente alternato con l'uso di *fermo restando* che, tra le due, risulta essere la variante preferita nelle leggi di origine nazionale.

Una sovrapposizione tra rappresentazione di scenari deontici ed espressione della performatività si può osservare anche in alcune sequenze di verbi, sostantivi e preposizioni “attraverso i quali si esprime sul piano linguistico la funzione prescrittiva delle norme, che qualificano giuridicamente comportamenti o attribuiscono posizioni giuridiche” (Belvedere 1994: 23). Come dimostrato da Venturi (2011), nei testi normativi italiani si evidenzia la centralità dei termini per l'espressione della realtà giuridica e si può notare la loro “forza associativa” all'interno di costrutti in grado di evocare una determinata situazione conoscitiva. In particolare, attraverso l'individuazione di polirematiche costruite intorno a determinati sostantivi (come *obbligo*) con verbi supporto diversi, che attivano il frame³² BEING OBLIGATED all'interno dello scenario deontico generale. Nelle costruzioni a verbo supporto utilizzate in costrutti

³¹ Secondo Gualdo *fatto salvo* rappresenta un esempio di connettivo pesante che rientra tra i tecnicismi collaterali microsintattici che producono “ben noti effetti di straniamento e di oscurità” (2011: 427).

³² Il metodo di annotazione semantica utilizzato da Venturi faceva riferimento al progetto *Framenet* (cfr. Venturi 2012). In Venturi (2011) l'informazione relativa al tipo di *frame* semantico è stata aggiunta manualmente sulla struttura sintattica ad albero generata automaticamente dall'annotazione sintattica a dipendenze.

tecnici, come quelle identificate da Venturi, il verbo subisce un depotenziamento semantico (cfr. Rovere 2005) per cui esso “perde il proprio potere semantico e la capacità evocativa passa al sostantivo” (Venturi 2011: 244).

Da questa angolazione, si è proceduto isolando sequenze morfosintattiche dotate di una struttura predicativa in grado di evocare una situazione-tipo mediante un’articolazione verbo+aggettivo+preposizione, verbo+sostantivo+preposizione o verbo+preposizione+sostantivo; verbo+(articolo)+sostantivo. In tali configurazioni l’informazione semantica è veicolata non dalla testa sintattica dell’unità lessicale, ossia dal sostantivo, ma dall’intero sintagma o clausola verbale o preposizionale. Si considerino i seguenti dati ottenuti mediante un’analisi *corpus-based* da cui si evince come alcune collocazioni si distribuiscono variabilmente nei tre corpora qui considerati nella Tabella 10:

Frame	Collocazione	Corpus A	Corpus B	Corpus C
verbo+aggettivo+preposizione	<i>è/sono soggetto/a/i/e a</i>	21	190	77
	<i>è/sono assoggettato/a/i/e a</i>	8	37	21
verbo+preposizione+sostantivo	<i>entra/entrano in vigore³³</i>	448	40	90
verbo+articolo+sostantivo	<i>ha/hanno l’obbligo</i>	0	3	28
verbo+sostantivo	<i>è fatto obbligo</i>	4	166	196
Totale		481	436	412

Tab. 10. Distribuzione di collocazioni giuridiche
(normalizzazione a 1 milione)

³³ Nel Corpus A occorre anche la variante *mettere in vigore* che non si attesta negli altri due corpora, con 375 occorrenze alla terza persona plurale dell’indicativo presente.

La Tabella 10 mostra la distribuzione intercorpora di alcune collocazioni giuridiche potenzialmente con duplice valenza, prescrittiva e performativa, che sono variabilmente utilizzate nelle tre varietà legislative: *entrare/mettere in vigore* è significativamente sovrarappresentato in euroletto italiano; *essere soggetto/i a* contraddistingue l'italiano delle misure nazionali di attuazione; il costruito *è fatto obbligo* risulta invece caratterizzare l'italiano legislativo di origine nazionale del Corpus C e, in misura inferiore, il Corpus B, nazionale ma di derivazione europea, mentre è significativamente sottorappresentato nel Corpus A. In particolare è da notare come esso occorra quasi esclusivamente nella formula conclusiva con rimando anaforico al decreto legislativo: “è fatto obbligo a chiunque spetti di osservarlo e di farlo osservare”³⁴. Inoltre, a corollario di questa formula, nelle norme nazionali vi è il riferimento al sigillo di Stato e le firme del Presidente della Repubblica e dei vari Ministri competenti ne assicurano validità ed efficacia giuridica. A seguire alcuni esempi estratti dai tre corpora: Corpus A (es. 23), Corpus B (es. 24), Corpus C (es. 25).

- (23) La presente direttiva **entra in vigore** il giorno della pubblicazione nella Gazzetta ufficiale delle Comunità europee. (Direttiva 1999/22/CE)

Gli Stati membri **mettono in vigore** le disposizioni legislative, regolamentari ed amministrative necessarie per conformarsi alla presente direttiva entro due anni dalla sua entrata in vigore. (Direttiva 1999/31/CE)

- (24) Chiunque omette di indicare il prezzo per unità di misura o non lo indica secondo quanto previsto dal presente capo **è soggetto alla** sanzione di cui all'articolo 22, comma 3, del decreto legislativo 31 marzo 1998, n. 114, da irrogare con le modalità ivi previste. (Decreto legislativo 6 settembre 2005 n. 206)

Alla stessa sanzione **è assoggettato** chiunque apporta modifiche agli apparecchi dotati della prescritta marcatura che comportano mancata conformità ai requisiti essenziali. (Decreto legislativo 9 maggio 2001 n. 269)

³⁴ Nel 91% dei casi in cui tale costruito compare nelle misure nazionali di attuazione del Corpus B esso è inserito in tale enunciato legislativo e, allo stesso modo, nel 97% degli esempi estratti dai decreti legislativi nazionali del Corpus C.

- (25) Il presente decreto, munito del sigillo dello Stato, sarà inserito nella Raccolta ufficiale degli atti normativi della Repubblica italiana. **È fatto obbligo** a chiunque spetti di osservarlo e di farlo osservare. (Decreto legislativo 28 agosto 2000 n. 274)

Analogamente, per il costrutto *è fatto divieto*, che contestualmente costruisce uno scenario deontico ed esercita un effetto costitutivo, si registra una diversa rappresentazione delle occorrenze, normalizzate a un milione: Corpus A (3), Corpus B (17) e Corpus C (15). Di seguito alcuni esempi estratti dalle leggi nazionali di attuazione (es. 26) e dalla legislazione italiana (es. 27) dove tale struttura risulta più utilizzata:

- (26) **È fatto divieto di** offrire, elargire o richiedere incentivi o benefici finanziari per la partecipazione dei soggetti alla sperimentazione clinica, ad eccezione delle eventuali indennità per il volontario sano. (Decreto legislativo 24 giugno 2003 n. 211)
- (27) [...] **è fatto divieto** alla pubblica amministrazione, alle stazioni appaltanti, agli enti aggiudicatori e ai soggetti aggiudicatori di richiedere alle imprese che concorrono alle procedure di cui al comma 1 requisiti finanziari sproporzionati rispetto al valore dei beni e dei servizi oggetto dei contratti medesimi. (Legge 11 novembre 2011 n. 180)

5.3. *Forza costitutiva*³⁵ e *testualità*

La natura regolativa del genere “legge” si rapporta all’esigenza da parte del legislatore di esercitare un rigido controllo “sulla libertà d’interpretazione” (Sabatini 1999: 150) e di rendere massimamente esplicito il contenuto, in ragione della primaria funzione conativa del testo legislativo. La specificità di essere testi fortemente vincolanti si evidenzia nelle leggi anche a livello testuale: si pensi alla rigida organizzazione macrostrutturale in sezioni e sotto-sezioni, spesso corredate da espedienti grafici (come l’uso del maiuscolo) ed elementi paratestuali (testo centrato), vincolate anche a una fissità formulaica che contribuisce alla costruzione della testualità.

³⁵ Per la nozione di costitutività si rimanda alla trattazione di Carcaterra (1979) e al commento critico in Roversi (2012).

In questa prospettiva si intende verificare se, rispetto all'ambiguità della valenza (prescrittiva o performativa) del presente indicativo italiano si possano rintracciare delle associazioni con specifiche sezioni testuali, quali preambolo e articolato della legge.

L'analisi della macrostruttura testuale delle norme italiane (sia europee che nazionali) rivela un'articolazione interna in due, talvolta tre, sezioni qualora risulti presente anche un allegato. A tale impianto sul piano della testualità corrisponde anche una diversa strutturazione pragmatica. Nella parte iniziale del testo, il cosiddetto preambolo, si fonda l'agentività che codifica il valore performativo del *dictum*: all'interno del preambolo si esplicita la cornice giuridica di riferimento atta a garantire la validità dell'atto, si menziona il ruolo istituzionale di chi lo emette e si inseriscono le fonti giuridiche che ne illustrano la motivazione mediante rimandi intertestuali. Il preambolo contribuisce a situare il testo normativo attraverso l'uso di "textual mapping devices", finalizzati a "indicate textual relations between various (sub)sections of the same or related provisions" (Bhatia 1987: 2). In esso sono contenute motivazioni e intenti che ne rendono necessaria l'emanazione ma, al tempo stesso, si distinguono quelle che Felici (2013: 35) definisce "marks of performativity": i nomi dei soggetti, i *visto/considerato/considerando*³⁶, il rispetto delle condizioni di felicità legato alla legittimità delle autorità competenti, e la proclamazione dello strumento adottato attraverso un enunciato dichiarativo. In tal modo, nel preambolo, si viene ad attualizzare un nuovo stato di cose con un'esplicita forza performativa veicolata da uno schema fisso del tipo [Autorità + verbo performativo giuridico + strumento legislativo] che caratterizza così ogni direttiva del Corpus A: [*Il Parlamento europeo/ il Consiglio/la Commissione delle Comunità europee HA/HANNO ADOTTATO la presente direttiva*]. Nelle misure giuridiche dei corpora B e C: [*Il Presidente della Repubblica EMANA il presente decreto legislativo*].

Nella sezione normativa, ovvero nel cosiddetto articolato o dispositivo, sono contenuti gli enunciati legislativi con valenza prescrittiva o performativa, ivi compresi i casi in cui è evidente una sovrapposizione

³⁶ It. *considerando* (ingl. *having regard to*; fr. *considérant*) può essere considerato una marca eurolettale in quanto caratterizza esclusivamente la varietà legislativa di italiano della legislazione dell'Unione europea (cfr. Mori 2018c).

tra le due modalità (§4). Secondo Garzone (2008: 66) la formula introduttiva (*enacting formula*) costituisce un prefisso performativo identificabile in inglese giuridico dalla presenza di *hereby*,³⁷ un arcaismo utilizzato in particolare nella normativa nazionale (rispetto all'euroletto inglese). La sua forza pragmatica si estende grazie alla "proprietà transitiva degli enunciati tetici" (Garzone 1999: 132-134) con cui si spiega il motivo per cui, all'interno di testi normativi, può non essere esplicitamente espressa la principale a cui il verbo dichiarativo-ingiuntivo fa riferimento, poiché il *dictum* "ha forza sufficiente a instaurare un nuovo stato di cose". Ciò consentirebbe di traslare la valenza performativa associata con quanto contenuto nel preambolo agli articoli, così producendo, secondo l'autrice, "i loro effetti nello stesso momento di emanazione del provvedimento e pur non presentando al loro interno alcun prefisso performativo".

Negli esempi che seguono, estratti da A (es. 28), B (es. 29) e C (es. 30), è da segnalare come in italiano l'uso del modificatore *presente* può contribuire a creare un forte ancoraggio al momento dell'enunciazione così incrementando la performatività thetica ad esso associata, analogamente a quanto si osserva con l'uso di *hereby* in inglese.

- (28) La *presente* direttiva³⁸ **disciplina** l'accesso alle attività svolte dagli enti pensionistici aziendali o professionali, nonché l'esercizio di tali attività. (Direttiva 2003/41/CE)

La *presente* direttiva **stabilisce** i diritti degli utenti finali e i corrispondenti obblighi delle imprese che forniscono reti e servizi di comunicazione elettronica accessibili al pubblico. (Direttiva 2002/22/CE)

- (29) Il *presente* decreto **detta** norme in materia di giardini zoologici finalizzate a potenziarne il ruolo nella conservazione della biodiversità, allo scopo di proteggere la fauna selvatica e di salvaguardare la stessa diversità biologica. (Decreto legislativo 21 marzo 2005 n. 73)

³⁷ La possibilità di trovare la forma avverbiale *hereby* insieme ai performativi pare costituire un vero e proprio test in base al quale è possibile accertare la natura performativa del verbo (Benveniste 1963).

³⁸ Nel Corpus A le occorrenze del sintagma "la presente direttiva" corrispondono alle versioni in francese (*la présente directive*), là dove in euroletto inglese si opta per l'uso del dimostrativo: *this directive*.

- (30) Le disposizioni del *presente* decreto **disciplinano** gli aspetti organizzativi e procedurali della attività di gestione amministrativa e contabile delle rappresentanze diplomatiche e degli uffici consolari di I categoria, di seguito denominati: “uffici all’estero”. (Decreto legislativo 15 dicembre 2006 n. 307)

La corrispondenza tra forza costitutiva dell’enunciato e sezione testuale si manifesta con una maggiore fissità nell’impianto dell’articolato delle direttive in cui è presente uno specifico articolo (frequentemente art. 1) deputato a definire l’oggetto della stessa e il suo campo/ambito di applicazione o scopo/obiettivo, come esplicitato nel titolo stesso dell’articolo in cui tali enunciati performativi sono contenuti. La presenza di enunciati performativi *thetic* – che contestualmente codificano la prescrittività – si rintraccia anche negli articoli dedicati alle definizioni che caratterizzano in particolare la legislazione di origine europea e le leggi nazionali di attuazione delle direttive europee. Occorre, infatti, considerare che l’articolo intitolato “Definizioni” rappresenta una caratteristica specifica del sotto-genere *direttiva* e, in misura minore, delle norme di attuazione della legislazione europea, mentre è solo incidentalmente inserito nella normativa nazionale. A conferma di ciò si considerino i seguenti dati normalizzati relativi alla presenza di tali articoli: Corpus A (220), Corpus B (156), Corpus C (39). Tali frequenze dimostrano chiaramente come la prassi redazionale che prevede un articolo (di frequente il numero 2), in cui sono spiegati molti dei termini di riferimento per la norma in questione, sia una prerogativa di origine unionale.

Nella fase di attuazione delle direttive, per la resa dell’articolo “Definizioni” spesso si adotta la tecnica della trasposizione copia-incolla, poiché qualsiasi discrepanza concettuale o semantica potrebbe inficiare l’efficacia della norma. Nello specifico, con la formula “ai fini della presente direttiva s’intende” si attiva un vincolo interpretativo che crea un nuovo stato di cose legato a un determinato *definendum* all’interno del quadro giuridico unionale, così da evitare ogni ambiguità rispetto all’interpretazione delle nozioni in essa contenute³⁹. Infatti, la presenza

³⁹ Di frequente il linguaggio unionale adopera terminologia già presente negli ordinamenti nazionali, risemantizzandoli sulla base delle specifiche esigenze comunicative degli atti normativi europei (cfr. Megale 2008: 79-82).

di un apparato definitorio, come quello che si rintraccia abbondantemente nei corpora A e B, è fondamentale per evitare “problemi di sovrapposizione di concetti comunitari a concetti già in uso nei diversi ordinamenti giuridici” (Ioriatti Ferrari 2010: 300). A questo proposito è interessante segnalare l’uso del verbo *intendere*, in particolare (*si intende*/(*si intendono*)/*si deve intendere* nei tre corpora, che rispecchia la scelta delle corrispondenti versioni francesi con il verbo *entendre* (*on entend*) per esprimere la valenza costitutiva dell’enunciato finalizzato a fornire un’istruzione ai destinatari⁴⁰. Nella Tabella 11 a seguire sono riportate le clausole introduttive più frequentemente utilizzate nei corpora di origine e di derivazione europea:

Corpus A	<p>Ai fini della presente direttiva (e dei suoi allegati) valgono le seguenti definizioni (e classificazioni) [...]</p> <p>Ai sensi della presente direttiva, si intende per [...]</p> <p>Ai fini della presente direttiva si applicano le seguenti (specifiche) definizioni [...]/di cui all’articolo [...]</p> <p>Ai fini della presente direttiva si intende per [...]</p> <p>Ai fini della presente direttiva, valgono le seguenti definizioni [...]</p>
Corpus B	<p>Ai fini del presente decreto (e dei suoi allegati) si intende per [...]</p> <p>Ai fini del presente decreto, valgono le seguenti definizioni [...]</p> <p>Ai fini del presente capo si intende per [...]</p> <p>Ai fini del presente sezione si intende per [...]</p> <p>Ai fini del presente decreto legislativo si applicano le seguenti definizioni [...]</p> <p>Agli effetti delle disposizioni di cui al presente titolo si intende per [...]</p> <p>Ai fini delle disposizioni di cui al presente decreto si intendono per [...]</p>

Tab. 11. Clausole introduttive delle definizioni

⁴⁰ Sulle corrispondenze tra le versioni francese-inglese-italiana-tedesca in merito alle espressioni definitorie in testi legislativi dell’Unione europea si veda Felici (2012: 58-59).

Nei limitati casi in cui articoli dedicati alle definizioni sono presenti nelle norme nazionali, questi si aprono con la clausola introduttiva *Ai fini del presente decreto devono intendersi* oppure mediante inserti definitivi dislocati nell'articolato, senza che questi siano ben identificabili in un articolo dedicato, come nell'esempio che segue, estratto dal Corpus C, in cui compare anche il modale di obbligo e numerosi rimandi intertestuali:

- (31) La terminologia tecnica utilizzata nel presente decreto legislativo **deve intendersi** nel significato suo proprio desumibile dalla normativa di riferimento ed, in particolare, dal decreto del Presidente della Repubblica 29 marzo 1973, n.156, dal decreto del Presidente della Repubblica 19 settembre 1997, n. 318, dalla direttiva del Presidente del Consiglio dei Ministri 3 marzo 1999, pubblicata nella Gazzetta Ufficiale n. 58 dell'11 marzo 1999, nonché dalla legge 22 febbraio 2001, n. 36, e successive modificazioni. (Decreto legislativo 4 settembre 2002 n. 198)

L'analisi qualitativa, in prospettiva pragmatica, degli articolati dei testi contenuti nei tre corpora ha consentito, inoltre, di rintracciare casi in cui la struttura sintattica è quella di "verbo dichiarativo ingiuntivo più dictum" (Garzone 1996: 31) mediante i quali la funzione di indicatore di forza performativa è estesa a una dipendente, che esprime il contenuto proposizionale dell'enunciato. Come si può evincere a seguire (es. 32-34) si tratta di enunciati dichiarativi dove la forza illocutiva risiede nell'autorità esterna e nel ruolo da essa esercitato che le conferisce il potere di provocare un cambiamento sullo stato di cose. In particolare "la forza performativa non si limita al verbo della principale, ma investe anche la proposizione contenuta nel *dictum*."

Negli esempi estratti dai corpora A (es. 32), B (es. 33) e C (es. 34) a seguire, la valenza performativa dell'enunciato è altresì avvalorata dalla presenza di rimandi intratestuali alla cornice giuridica e da sintagmi nominali che designano referenti portatori di un'autorità giuridica (Stati membri, organi o autorità competenti, tribunale, etc.).

- (32) Ciascuno Stato membro dispone che, nei casi specifici e nelle condizioni e limiti stabiliti dalla legge nazionale, l'organo di vigilanza o di amministrazione della SE o della società partecipante situato nel

proprio territorio non sia obbligato a comunicare informazioni che, secondo criteri obiettivi, siano di natura tale da creare notevoli difficoltà al funzionamento della SE (o, eventualmente, della società partecipante) o delle sue affiliate e dipendenze, o da arrecar loro danno. (Direttiva 2001/86/CE)

- (33) Nel caso di mancanza di licenza o del registro o del certificato, di cui ai commi 1, 2, e 4, l'organo competente, nel pronunciare il provvedimento definitivo di accertamento delle infrazioni, **dispone che** il provvedimento venga comunicato all'organismo ufficiale competente e tramite questo, al Ministero, il quale provvede a pubblicare su sito internet l'elenco di tali provvedimenti e dei trasgressori, e a renderlo disponibile a chi ne fa richiesta. (Decreto legislativo 10 novembre 2003 n. 386)
- (34) Il tribunale, con l'applicazione della misura di prevenzione, **dispone che** la persona sottoposta a tale misura versi presso la cassa delle ammende una somma, a titolo di cauzione, di entità che, tenuto conto anche delle sue condizioni economiche e dei provvedimenti adottati a norma dell'articolo 22, costituisca un'efficace remora alla violazione delle prescrizioni imposte. (Decreto legislativo 6 settembre, 2011, n. 159)

Per quanto riguarda la porzione finale dell'articolato, in essa si sancisce quanto espresso in premessa e si legittima la validità del testo normativo mediante la firma apposta dalla persona che gode di autorità giuridica. In alternativa, può essere il quadro istituzionale di riferimento a garantire legittimità ad alcuni soggetti così attribuendo un valore performativo all'enunciazione. In questi casi è l'ufficialità stessa dell'atto, a firma dell'autorità competente, a instaurare una condizione contestuale e dei vincoli cotestuali (deittici di tempo, luogo e persona) che determinano la concretizzazione del contenuto nel momento stesso della sua enunciazione. Così Garzone (1999: 134):

[...] all'interno del testo normativo il prefisso performativo non è espresso in ogni singola parte o sezione, ma è dislocato, estromesso in ambito per così dire paratestuale: la soggettività da cui deriva la natura performativa dell'enunciato, oltre ad essere garantita dalla firma della persona in cui tale autorità è investita, viene fondata nelle premesse del

testo stesso. Anche nei casi in cui si registri l'ellissi totale del prefisso performativo, che non compare neanche in posizione dislocata, è il quadro istituzionale o convenzionale che fonda la soggettività a determinare in ultima analisi la forza tetica del dictum.

I testi legislativi si concludono con gli articoli di entrata in vigore che precedono la firma e i riferimenti al campo indicale (*fatto/dato* + deittico di luogo, *addi* + deittico di tempo), che si ripetono con minime variazioni nel Corpus A (es. 35) mentre la formula cristallizzata utilizzata nel Corpus B (es. 36) corrisponde esattamente a quanto previsto dalla legislazione di origine nazionale del Corpus C (es. 37).

(35) *Articolo 16*

La presente direttiva entra in vigore il settimo giorno successivo alla pubblicazione nella Gazzetta ufficiale delle Comunità europee.

Articolo 17

Gli Stati membri sono destinatari della presente direttiva. (Direttiva 1999/2/CE)

(36) Il presente decreto entra in vigore il giorno successivo alla pubblicazione del medesimo nella Gazzetta Ufficiale della Repubblica italiana.

Il presente decreto, munito del sigillo dello Stato, sarà inserito nella Raccolta ufficiale degli atti normativi della Repubblica italiana. È fatto obbligo a chiunque spetti di osservarlo. (Decreto legislativo 21 marzo 2005 n. 73)

(37) Il presente decreto entra in vigore dal 1° giugno 2004.

Le disposizioni di cui agli articoli 2 e 4 entrano in vigore il quindicesimo giorno successivo alla data di pubblicazione nella Gazzetta Ufficiale.

Il presente decreto, munito del sigillo dello Stato, sarà inserito nella Raccolta ufficiale degli atti normativi della Repubblica italiana. È fatto obbligo a chiunque spetti di osservarlo e di farlo osservare. (Decreto legislativo 5 aprile 2002 n. 77)

6. Conclusioni

Nello studio qui proposto ci si è concentrati su aspetti di semantica e pragmatica giuridica che si manifestano nella codifica morfosintattica degli enunciati legislativi che caratterizzano le norme in lingua italiana. L'uso di corpora ha consentito di corroborare la tesi sull'esistenza di varietà legislative correlate con variabili situazionali quali il contesto giuridico di riferimento (europeo o nazionale), la prassi redazionale (plurilingue o monolingue) e il sotto-genere legislativo finalizzato al raggiungimento di un specifico intento comunicativo. Per le direttive si tratta di testi legislativi che mirano a stabilire obiettivi da raggiungere necessariamente; per le misure italiane di attuazione, di definire modalità e mezzi con cui raggiungerli; per gli strumenti legislativi di origine italiana, della determinazione congiunta di obblighi e di indicazioni per il loro raggiungimento. A livello di prassi redazionale, ci si confronta con un processo plurilingue su cui si fonda il diritto europeo (Corpus A) rispetto a una redazione orientata da una traduzione intralinguistica (Corpus B) e dall'adesione alle caratteristiche intertestuali del genere legislativo nella tradizione italiana (Corpus C).

L'ipotesi di partenza era che si venisse a configurare un continuum orientato, da un lato, verso il polo del diritto dell'Unione europea, e il diritto nazionale dall'altro. Le misure di attuazione collocate in una posizione intermedia: tradotte intralinguisticamente per rispondere alle norme redazionali italiane ma con proprie specificità, spesso dovute a un processo di ibridazione con le norme europee da cui si originano,⁴¹ tali da confermare per il Corpus B l'esistenza di un sotto-genere a sé, rispetto al Corpus A, anche nella rappresentazione di scenari deontici e nell'espressione della performatività, analogamente a quanto osservato per la distribuzione dei verbi modali (Mori 2020), a livello di organizzazione sintattica (Mori 2019b) e sul piano lessico-sintattico (Mori 2019b).

Le analisi condotte sono state rivolte all'individuazione delle forme con cui modalità deontica e performatività giuridica si manifestano negli enunciati legislativi (rispettivamente direttivi o dichiarativi) nei testi legislativi europei e nazionali, al fine di evidenziare analogie e differenze nella loro distribuzione e qualificazione.

⁴¹ A questo proposito Ferraresi e Bernardini (2019) propongono di parlare di un "mediation effect" per interpretare alcuni risultati emersi dal confronto interlinguistico (inglese-italiano) e intralinguistico sull'EOMC.

Il criterio seguito nell'identificazione di caratteristiche morfosintattiche ha previsto la selezione di forme e strutture rilevanti sul piano della semantica e della pragmatica giuridica in relazione all'espressione della modalità deontica e della performatività giuridica in termini di costitutività.

La metodologia utilizzata ha previsto la valutazione comparativa di risultati provenienti dal monitoraggio automatico del profilo linguistico (§ 3.1) e dalla combinazione di approcci quantitativi, *corpus-driven* e *corpus-based*, che hanno consentito l'individuazione di fenomeni di variazione caratterizzanti i tre corpora legislativi così da orientare un'analisi qualitativa dei dati nel loro contesto d'uso fornendone esemplificazioni.

Per quanto riguarda la codifica morfologica, in tutti e tre i corpora legislativi, l'uso del modo indicativo è predominante, sebbene B e C siano allineati rispetto a una maggior frequenza dei participi, che risulta marginale in A (Tabella 1). A livello di tempo verbale, si rinviene una presenza quasi esclusiva del presente nel Corpus A, mentre in C (e in misura inferiore in B) emerge anche il ricorso al futuro (Tabella 2). Per quanto riguarda il numero, nelle direttive sono sovrarappresentate le forme flesse alla terza persona plurale, mentre nei dispositivi nazionali di attuazione delle direttive e nella normativa di origine italiana si registra una maggiore frequenza di occorrenze alla terza persona singolare (Tabella 3). Nel confronto intercorpora, volto a evidenziare la rappresentazione di scenari deontici, è stata anche considerata la distribuzione frequenziale delle forme attive e passive per valutarne la produttività nei tre corpora (Tabelle 4-5). La riflessione sul presente indicativo, che nel discorso giuridico non ha una valenza referenziale ma può codificare tanto la modalità deontica quanto quella performativa, è stata orientata dalla netta tendenza che emerge nella distribuzione del campione di verbi, tipicamente presenti in enunciati legislativi: le direttive sono caratterizzate dall'uso delle forme attive del verbo (con un valore circa doppio rispetto alle corrispondenti forme al passivo) mentre nelle norme di origine italiana è sovrarappresentato il ricorso al passivo, con un valore del doppio superiore rispetto ai medesimi verbi alla diatesi attiva. Le forme attive e passive del presente indicativo (terza persona singolare e plurale) più ricorrenti sono state oggetto di un approfondimento qualitativo finalizzato ad evidenziarne la duplice valenza, prescrittiva e performativa andando ad esaminare l'uso ambivalente dell'indicativo presente, finalizzato a realizzare l'effetto costitutivo della norma oltre che a rappre-

sentare scenari deontici (§ 4). L'analisi qualitativa condotta su una serie di esempi ha evidenziato una correlazione tra criteri semantici e parametri pragmatici che consentono di interpretare la valenza pragmalinguistica degli enunciati legislativi più rappresentati. In particolare quando in presenza di soggetti animati, seppur non prototipici (ad es. *Stati membri* o *autorità competente*), il presente indicativo occorre in enunciati che si configurano come direttivi mentre con soggetti inanimati e determinate spie cotestuali (*presente, tale, come segue*, etc.) l'indicativo presente assume una valenza performativa, all'interno di enunciati dichiarativi. L'analisi del profilo collocazionale dei verbi maggiormente attestati ha evidenziato differenze quantitative e un diverso rapporto tra i verbi giuridici alla diatesi attiva o passiva. La distribuzione delle forme passive è sovrarappresentata in tutti e tre i corpora rispetto alle forme attive dei verbi considerati, in misura maggiore nella legislazione di origine nazionale del Corpus C. Inoltre gli esempi estratti dai corpora legislativi nazionali (B e C) mostrano il ricorso ad ancoraggi intertestuali o intratestuali che contribuiscono all'espressione della performatività all'interno di enunciati dichiarativi.

Nel § 5 l'attenzione si è concentrata sull'espressione della performatività a livello giuridico, andando a estrarre dai corpora i dati quantitativi relativi ad alcuni verbi performativi giuridici (Tabelle 6-7), per poi analizzarne l'uso all'interno di enunciati legislativi produttivi di effetti linguistici in termini di costitutività. Gli esempi discussi in § 5.1 mostrano come l'espressione della performatività giuridica si manifesta variabilmente nei tre corpora, mediante forme verbali o collocazioni (Tabella 8) dotate di una "forza costitutiva" riconducibile a un'autorità esterna (più o meno esplicitamente presente nel cotesto) che ha il potere di provocare un cambiamento nello stato di cose. Pertanto, nel § 5.2, ci si è soffermati su casi in cui il discrimine risulta di difficile identificazione o, a volte, ci si trova di fronte a una sovrapposizione delle due valenze o a casi di limitazione della forza illocutiva di alcuni enunciati legislativi che, pur realizzando atti esercitivi, risultano sottoposti a vincoli giuridici o a condizioni extra-linguistiche. Nel Corpus A questo si manifesta mediante il frequente uso del connettivo condizionale *purché* o di *fatto salvo* come legal lexical bundle preferito rispetto a *fermo restando* e *previo* + N, rispettivamente sovrarappresentati in B e in C (Tabella 9). È stato inoltre interessante

considerare alcuni costrutti morfosintattici, dotati di struttura predicativa, che sono attestati nel discorso giuridico dove assumono tanto valenza prescrittiva quanto performativa in relazione al cotesto e al contesto (Tabella 10).

Infine, nel tentativo di isolare delle tendenze, è stato utile considerare la correlazione tra costitutività delle norme e testualità rispetto all'articolazione dell'impianto macro-strutturale (§ 5.3). Una peculiarità, che interessa i casi in cui valore deontico e performativo convergono, riguarda gli enunciati deputati a fornire definizioni, contenute o meno, in sezioni testuali preposte. Infatti, diversamente da quanto accade a livello nazionale, nelle direttive così come nelle misure di attuazione, si rinvencono articoli di "Definizioni" in cui andare a reperire chiarimenti terminologici. All'interno di questi, sia in A che in B, si ripetono clausole introduttive (Tabella 11) che confermano la necessità, per le norme inserite nella cornice europea (direttive) o da essa dipendenti (leggi e decreti legislativi di attuazione delle direttive), di rendere inequivocabile il significato per il sistema giuridico di riferimento.

Complessivamente i risultati ottenuti e discussi hanno consentito di individuare diverse soluzioni finalizzate a rappresentare linguisticamente gli scenari deontici all'interno dei quali si situano le norme, dimostrando traiettorie di variazione intra-genere tra direttive europee (Corpus A), leggi italiane di attuazione (Corpus B) e leggi redatte a livello nazionale, senza condizionamenti da parte del diritto dell'UE (Corpus C). L'osservazione di due aspetti complementari per lo studio dei testi legislativi, modalità deontica e performatività, mostra come nel sottogenere direttiva si manifestino maggiormente le norme prescrittive con un maggior ricorso alla costruzione di scenari deontici che troveranno una loro successiva attualizzazione. Nella legislazione nazionale è più rilevante l'incidenza delle norme costitutive a causa della loro applicabilità diretta che si rinviene in un maggiore ricorso all'espressione della performatività.

Inoltre le riflessioni qui presentate si inseriscono nella discussione al centro del dibattito scientifico internazionale sulle dinamiche variazionali correlate al diverso contesto legislativo e al loro polimorfismo redazionale, più o meno orientato verso il plurilinguismo, confermando le intuizioni di Gémar (1981: 349): "au rythme des relations internationales peu à peu il se dégage par osmose des principes «communautaires»

sur lesquels finiront par se rejoindre les grandes traditions et les grands systèmes juridiques pour former l'union symbiotique déjà réalisée dans quelques domaines”.

In conclusione, l'approccio metodologico utilizzato conferma i vantaggi che derivano dall'interpretazione di dati ottenuti con una triangolazione di metodi: NLP integrato con un approccio misto quantitativo (di tipo *corpus-based* e *corpus-driven*) e qualitativo nella prospettiva di una sociolinguistica dei corpora (Mori 2019a).

Laura Mori
Università degli Studi Internazionali di Roma - UNINT
Via Cristoforo Colombo, 200 - 00147 Roma
laura.mori@unint.eu

Bibliografia

- Austin, John Langshaw, 1962, *How to do things with words*, Oxford, Oxford University Press.
- Bambi, Federigo, 2012, “Le ragioni della storia. Tra due bilinguismi”. In: Pozzo, Barbara / Bambi, Federigo (a cura di), *L'italiano giuridico che cambia*, Firenze, Accademia della Crusca: 15–29.
- Belvedere, Andrea, 1994, “Linguaggio giuridico”. *Digesto delle discipline privatistiche XI*: 21–31.
- Benveniste Émile, 1963, “La philosophie analytique et le langage”. *Les Études philosophiques* 1: 3-11.
- Berruto, Gaetano, 2011, “Registri, generi e stili: Alcune considerazioni su categorie mal definite”. In: Cerruti, Massimo / Corino, Elisa, Onesti, Cristina (a cura di), *Formale e informale. La variazione di registro nella comunicazione elettronica*, Roma, Carocci Editore: 15-35.
- Biber, Douglas / Conrad, Susan, 2009, *Register, genre, style*, Cambridge, Cambridge University Press.
- Bhatia, Vijay Kumar, 1987, “Textual-mapping in British Legislative Writing”. *World Englishes* 6/1: 1-10.
- Bhatia, Vijay Kumar, 1993, *Analysing genre: language use in professional settings*, London, Longman.
- Caffi, Claudia, 2009, *Pragmatica. Sei lezioni*, Roma, Carocci.

- Caliendo, Giuditta, 2004, “Modality and Communicative Interaction in EU Law.” In: Candlin, Christopher, N. / Gotti, Maurizio (eds.) *Intercultural Aspects of Specialized Communication*, Bern, Peter Lang: 241–259.
- Caliendo, Giuditta / Di Martino, Gabriella / Venuti, Marco, 2005, “Language and Discourse Features of EU Secondary Legislation”. In: Cortese, Giuseppina / Duszak, Anna (eds.) *Identity, Communication, Discourse. English in Intercultural settings*, Bern, Peter Lang: 381-404.
- Cao, Deborah, 2007, *Translating law*, Clevedon, Multilingual Matters.
- Carcattera, Gaetano, 1974, *Le norme costitutive*, Milano, Giuffrè.
- Carcattera, Gaetano, 1979, *La forza costitutiva delle norme*, Roma, Bulzoni.
- Caterina, Raffaele, / Rossi, Piercarlo, 2008, “L’italiano giuridico”. In Pozzo, Barbara / Timoteo, Marina (a cura di), *Europa e linguaggi giuridici*, Milano, Giuffrè Editore: 185–208.
- Commissione europea. Direzione Generale della Traduzione, 2009, *Tradurre per una comunità multilingue*, Lussemburgo, Ufficio per le pubblicazioni ufficiali delle Comunità europee.
- Conte, Amedeo, 1994, “Fenomenologia del linguaggio deontico”. In: Scarpelli, Ugo / Di Lucia, Paolo (a cura di), *Il linguaggio del diritto*, Milano, LED Edizioni Universitarie.
- Cornu, Gérard , 2005, *Linguistique juridique*, 3ème édition, Paris, Montchrestien.
- Cortelazzo, Michele, 2010, “Linguaggio giuridico-amministrativo”. In: Simone, Raffaele (a cura di), *Enciclopedia dell’italiano*, Roma, Treccani.
- Dell’Orletta, Felice, 2009, “Ensemble system for Part-of-Speech tagging”. In: *Proceedings of Evalita’09* (Evaluation of NLP and Speech Tools for Italian), http://www.evalita.it/sites/evalita.fbk.eu/files/proceedings2009/PoSTagging/POS_ILC.pdf.
- Felici, Annarita, 2012, ”Shall ambiguities in EU legislative texts”. *Comparative Legilinguistics. International Journal of Legal Communication* 10: 51-66.
- Felici, Annarita, 2013, “Human Rights across time and space: a cross-linguistic analysis of performatives in English and Italian”. *Journal of Applied Linguistics* 23/1: 31-43.
- Ferraresi, Adriano / Bernardini, Silvia, 2019, “Lexical simplification in English and Italian Eurolects: (con|di)vergences”. Convegno dell’Associazione Italiana di Anglistica *Thinking out of the Box in Language, Literature, Cultural and Translation Studies: Questioning Assumptions, Debunking Myths, Trespassing Boundaries*, Università degli Studi di Padova, 5-7 Settembre 2019.

- Fiorito, Lorenzo, 2006, "On Performatives in Legal Discourse". *Metalogicon* 19/2: 101-112.
- Garzone, Giuliana, 1996, *Performatività e linguaggio giuridico. Una proposta di classificazione. Con osservazioni contrastive inglese-italiano*, Milano, Centro Linguistico Università Bocconi.
- Garzone, Giuliana, 1999, "Espressione della performatività nel testo giuridico: italiano e inglese a confronto". In: Azzaro, Gabriele / Ulrych, Margherita (a cura di). *Anglistica e....: metodi e percorsi comparatistici nelle lingue*, Trieste, Edizioni dell'Università di Trieste: 127-144.
- Garzone, Giuliana, 2008, "Gli enunciati performativi nel testo giuridico inglese: la prospettiva linguistica". In: Garzone, Giuliana / Santulli, Francesca (a cura di). *Il linguaggio giuridico. Prospettive interdisciplinari*, Milano, Giuffrè: 57-87.
- Garzone, Giuliana, 2013, "Variation in the use of modality in legislative texts: focus on shall", *Journal of Pragmatics* 57: 68-81.
- Garzone, Giuliana / Salvi, Rita, 2007, *Legal English*, Milano, Egea - Università Bocconi Press.
- Gémar, Jean-Claude, 1981, "Réflexions sur le langage du droit: problèmes de langue et de style". *Meta. Journal des traducteurs* 26: 338-349.
- Gualdo, Riccardo, (2011), "Il linguaggio del diritto". In: Gualdo, Riccardo / Telve, Stefano, *Linguaggi specialistici dell'italiano*, Roma, Carocci editore: 411- 461.
- Ioriatti Ferrari, Elena, 2010, "Linguismo eurunionico e redazione della norma comunitaria scritta". In: Visconti, Jaqueline (a cura di) *Lingua e diritto: livelli di analisi*, Milano, LED Edizioni Universitarie: 261-312.
- Kurzton, Dennis, 1997, "Legal language: varieties, genres, registers, discourses". *International Journal of Applied Linguistics* 9/2: 119-139.
- Mattila, Heikki, 2013, *Comparative legal linguistics: Language of law, Latin and modern lingua francas* (2nd ed.), Farnham, Ashgate.
- Mantovani, Dario, 2008, "Lingua e diritto. Prospettive di ricerca fra sociolinguistica e pragmatica". In: Garzone, Giuliana / Santulli, Francesca (a cura di), *Il linguaggio giuridico. Prospettive interdisciplinari*, Milano, Giuffrè: 17-56.
- Megale, Fabrizio, 2008, *Teorie della traduzione giuridica fra diritto comparato e "translation studies"*, Napoli, Editoriale scientifica.
- Montemagni, Simonetta, 2013, "Tecnologie linguistico-computazionali e monitoraggio della lingua italiana", in *Studi Italiani di Linguistica Teorica e Applicata* 1: 145-172.

- Mori, Laura, 2018a (ed.) *Observing Eurolects. Corpus analysis of linguistic variation in EU law*, Amsterdam, John Benjamins.
- Mori, Laura, 2018b, "Introduction. The Eurolect Observatory Project". In: Mori, Laura (ed.), *Observing Eurolects. Corpus analysis of linguistic variation in EU law*, Amsterdam, John Benjamins: 2-26.
- Mori, Laura, 2018c, "Observing Eurolects: the case of Italian". In: Mori, Laura (ed.), *Observing Eurolects. Corpus analysis of linguistic variation in EU law*, Amsterdam, John Benjamins: 200-242.
- Mori, Laura, 2019a, "La sociolinguistica dei corpora per lo studio della lingua inclusiva di genere nelle varietà legislative dell' Eurolect Observatory Multilingual Corpus (francese, inglese, italiano, spagnolo, tedesco)". In: Cavagnoli, Stefania / Mori, Laura (eds.) *Gender in legislative languages. From EU to national law in English, French, German, Italian and Spanish*, Berlin, Frank & Timme: 39-65.
- Mori, Laura, 2019b, "Configurazioni strutturali e funzioni discorsive dei *lexical bundles* nella costruzione testuale delle leggi italiane nazionali e di derivazione europea". In: Bombi, Raffaella (a cura di), *Saggi interlinguistici e metalinguistici*, Roma, Il Calamo, 79-88.
- Mori, Laura, 2019c, "Complessità sintattica e leggibilità. Un monitoraggio linguistico per la valutazione dell'accessibilità dei testi legislativi europei e italiani". *Studi Italiani di Linguistica Teorica e Applicata* 48: 627-657.
- Mori, Laura, 2019d, "Dall'armonizzazione all'ibridazione nei testi legislativi: evidenze linguistiche e manifestazioni interculturali nell'italiano tradotto". *EntreCulturas. Revista de Traducción y Comunicación Intercultural* 10: 377-392.
- Mori, Laura, 2020, "La distribuzione dei verbi modali in testi legislativi europei e italiani. Uno studio corpus-based sulla variazione intralinguistica di *dovere* e *potere*". *AIQN-L. Annali del Dipartimento di Studi Letterari, linguistici e comparati. Sezione Linguistica*, 9: 141-163.
- Mori, Laura, Szmrecsanyi, Benedikt, 2020, "Mapping Eurolects: an aggregate perspective on similarities between legislative varieties". *Languages in Contrast. International Journal for Contrastive Linguistics*, Online First.
- Mortara Garavelli, Bice, 2001, *Le parole e la giustizia*, Torino, Einaudi.
- Pozzo, Barbara, 2012, "Le ragioni del Convegno. L'italiano e la sfida del multilinguismo europeo". In: Pozzo, Barbara / Bambi, Federico. *L'italiano giuridico che cambia*, Firenze, Accademia della Crusca: 7-14.

- Prieto Ramos, Fernando, 2014, "International and supranational law in translation: from multilingual lawmaking to adjudication". *The Translator* 20/3: 313-331.
- Robertson, Colin, 2010, "Legislative drafting in English for non-native speakers". *ESP Across Cultures* 5: 147-163.
- Rovere, Giovanni, 2005, *Capitoli di linguistica giuridica. Ricerche su corpora elettronici*, Torino, Edizioni dell'Orso.
- Roversi, Claudio, 2012, *Costituire: uno studio di ontologia giuridica*, Torino, Giappichelli Editore.
- Sabatini, Francesco, 1999, "'Rigidità-esplicitezza' vs 'elasticità-implicitezza': possibili parametri massimi per una tipologia dei testi". In: Skytte, Gunter / Sabatini, Francesco (a cura di), *Linguistica testuale comparativa*, København, Museum Tusulanum Press: 141-172.
- Sandrelli, Annalisa, 2018, "Observing Eurolects: the case of English". In Mori, Laura (ed.), *Observing Eurolects. Corpus analysis of linguistic variation in EU law*, Amsterdam, John Benjamins: 64-92.
- Sandrelli, Annalisa, 2019, "A corpus-based study of deontic modality in English Eurolect". Convegno dell'Associazione Italiana di Anglistica *Thinking out of the Box in Language, Literature, Cultural and Translation Studies: Questioning Assumptions, Debunking Myths, Trespassing Boundaries*, Università degli Studi di Padova, 5-7 Settembre 2019.
- Searle, John R., 1969, *Speech Acts: an essay in the philosophy of language*, Cambridge, Cambridge University Press.
- Steunenbergh, Bernard / Voermans, Wim, 2006, *The transposition of EC directives: A comparative study of instruments, techniques and processes in six member states*, Nijmegen, Wolf Legal Publisher.
- Tognini Bonelli, Elena, 2001, *Corpus linguistics at work*, Amsterdam, John Benjamins.
- Tomatis, Marco Stefano, 2018, "The Eurolect Observatory Multilingual Corpus. Construction and query tools". In: Mori, Laura (ed.) *Observing Eurolects. Corpus analysis of linguistic variation in EU*, Amsterdam, John Benjamins: 28-45.
- Venturi, Giulia, 2011, *Lingua e diritto: una prospettiva linguistico-computazionale*, Tesi di dottorato non pubblicata, Torino, Università degli Studi di Torino.
- Venturi, Giulia, 2012, "*Semantic annotation of Italian legal texts: A FrameNet-based approach*". In Fried, Mirjam / Nikiforidou Kiki (eds.) *Advances in Frame Semantics*, Amsterdam, John Benjamins: 51-84.

- Zuanelli, Elisabetta, 1998, “La dimensione pragmatica del testo normativo”. *Iter legis* 2/1: 260-274.
- Zuanelli, Elisabetta, 2000, “Macro-struttura pragmatica de modelli di interazione nel testo normativo”. In: Veronesi, Daniela (a cura di) *Linguistica giuridica italiana e tedesca/Rechtslinguistik des Deutschen und Italienischen*, Padova, Unipress: 85-99.
- Werlich, Egon, 1976, *A Text Grammar of English*, Quelle-Meyer, Heidelberg.
- Williams, Christopher, 2005, “Vagueness in legal texts: is there a future for shall?”. In: Bhatia, Vijay K. / Engberg, Jan / Gotti, Maurizio, Heller, Dorothee (eds.), *Vagueness in Normative Texts*, Bern, Peter Lang: 201-224.
- Williams, Christopher, 2011, “Legal English and Plain language: an update”. *ESP Across Cultures* 8: 139-151.

DANIEL RUSSO, ANGELA ANDREANI
(Università degli Studi dell'Insubria, Università degli Studi di Milano)

Ogden's Basic English and its roots in the Early Modern English search for language simplicity¹

The idea of English as a lingua franca was anticipated in the early 1920s by Charles Kay Ogden, who conceived his Basic English as a form of controlled language that could be widely understood and used with proficiency by both native and non-native speakers. Basic English arguably represents a crucial moment in the development of the ideas of language simplicity and it remains a milestone which has influenced later controlled versions of English (e.g. Simplified Technical English, Special English, Simple English Wikipedia). In Ogden's codification of simplified language (which was deeply influenced by his previous semiotic studies) a few simplification criteria can be distinguished: controlled vocabulary, lexical isomorphism, standardisation (intended as establishing a discrete number of shared rules for the use of language), morphosyntax as an outgrowth of lexicon, the superiority of analytic structures over synthetic structures, and universality of a shared language as a key factor in the resolution of human conflicts. Language simplicity is therefore essential for better communication and improved social interactions. This essay argues that Ogden's idea of language simplicity is deeply rooted in the history of the English language and can be tracked back to the Reformation period, in which "plain speech" was seen as a source of truth and virtue as opposed to the ambiguity of a language rich in figures of speech and rhetorical artifice, and to the 17th century, in which we can distinguish a supposedly scientific approach in language simplification.

Nei primi anni Venti Charles Kay Ogden intuì che l'inglese poteva ricoprire il ruolo di lingua franca e concepì Basic English come una lingua controllata che potesse essere compresa su larga scala e impiegata con competenza da parlanti madrelingua e non. Si può affermare che il Basic di Ogden rappresenti un momento chiave nello sviluppo delle idee di semplicità linguistica; esso rimane una pietra miliare cui si sono ispirate versioni recensori di inglese controllato (ad esempio Simplified Technical English, Special English, Simple English Wikipedia). Nell'idea di semplificazione linguistica di Ogden (che era stata ampiamente influenzata dai suoi studi in campo semiotico) si possono distinguere alcuni principi essenziali: lessico controllato,

¹ Both authors are responsible for the overall planning and research for this paper. In particular, Daniel Russo is responsible for sections 1 and 2 while Angela Andreani for section 3 and subsections. Section 4 was written jointly by the two authors.

isomorfismo lessicale, standardizzazione (intesa come codifica di un numero preciso di regole condivise nell'uso della lingua), morfosintassi come prodotto del lessico, superiorità di strutture analitiche rispetto a quelle sintetiche e universalità di una lingua comune come fattore chiave nella risoluzione dei conflitti umani. La semplicità della lingua è pertanto ritenuta funzionale al miglioramento della comunicazione e delle interazioni sociali. In questo studio sosteniamo che l'idea di semplicità linguistica di Ogden affonda le radici nella storia della lingua inglese, radici che possono essere rintracciate nel XVI secolo, epoca in cui al "parlar chiaro" (plain speech) erano ricondotti valori di verità e virtù in contrapposizione alle ambiguità di un linguaggio figurato ricco di artifici retorici, e nel XVII secolo, quando emerse l'applicazione di criteri di scientificità all'ambito della semplificazione linguistica.

1. Introduction

Amongst the plethora of international constructed languages created since the 19th century – e.g. Volapük, (Schleyer 1882 [1880]), Esperanto (Zamenhof 1889 [1887]; 1905); Latino sine flexione (Peano 1903), Novial (Jespersen 1928), Sona (Searight 1935), Interlingua (Gode and Blair 1951), Frater (Phạm 1957), Neo (Alfandari 1961) – Charles Kay Ogden's Basic English is based upon a peculiar approach which sets it apart from other contemporary projects aiming at establishing a simple and comprehensive language for the purposes of international communication. Unlike Esperanto, which is arguably the most successful constructed auxiliary language (at least in terms of its general popularity), Basic English has a relatively short history; nevertheless, at the peak of its prominence in the 1930s and 1940s, its success was inextricably intertwined with the spread of the English language worldwide, to such an extent that Basic English would eventually transition from a proposed lingua franca *per se* to teaching material for the beginner to elementary levels of (standard) English as a lingua franca². Short-lived as it may have been, Basic found many advocates in Britain and the US; in particular, two influential statesmen of the time initially promoted its use and

² Cf. the introduction of Ogden (1930: 3) where Basic is presented as an international auxiliary language and "more than a mere educational experiment" and the introduction of Ogden (1934: 4) where Basic has an explicit two-fold purpose: "to serve as an international auxiliary language" and "to provide a rational introduction to normal English". Furthermore, the educational aspect is also prevalent in the works of Richards (1935) and Myers (1938), who were active in spreading Basic as a teaching method respectively in China and India.

divulgation: these are none other than British Prime Minister Winston Churchill and US President Franklin D. Roosevelt (McArthur et al. 2018 [1997]).

One feature seems to be shared by all international constructed languages in the 19th and 20th centuries: they all claim to be ‘simple’ and ‘easy’ to learn, but the actual practical implications of this simplicity vary considerably across the auxiliary language projects of the time. For these reasons, this argumentative essay will discuss the role of Ogden’s Basic English in the history of the English language by concentrating on the following aspects: in Section 2 Ogden’s approach is introduced in relation to the notion of “linguistic simplicity” to explain that this can be better understood in contrast with other contemporary approaches; Section 3 provides an overview of the notion of linguistic simplicity in various authors from the 16th through the 17th centuries to trace the history of simplicity in English. In this section, corpus linguistics tools are used exclusively for the purposes of lexical reference and are not meant to produce a corpus-based analysis. The Conclusion sets out to highlight the evolution of the ideas and principles linking the Early Modern debates to Ogden’s Basic English.

2. *Simplicity in Basic English*

Before analysing the main features of Ogden’s Basic English and its underlying idea of simplicity, it is necessary to place it in the context of the auxiliary constructed languages between the 19th and the 20th century, as Ogden was no stranger to the international language movement: he was even invited as a consultant to the International Auxiliary Language Association in 1951 (Esterhill 2000: 5). The international language movement began with enthusiastic amateurs in the mid-19th century and reached maturity in the beginning of the 20th century, when it came to be increasingly dominated by academic linguists. These linguists styled themselves as the heirs of the Enlightenment tradition of searching the philosophical basis of language in order to shape a new neutral science-based language for the purposes of an improved international and intercultural communication, which will eventually end all conflict and support a peaceful and progress-driven society. These principles were particularly dominant in the early projects of the movement – Schleyer’s Volapük (literally “the world’s speech”, see Schleyer 1982 [1880]) and

Zamenhof's Esperanto (originally called *lingvo internacia*, see Zamenhof 1889 [1887]; 1905) – but they were still present, albeit as a form of resistance against linguistic and cultural imperialism, in the auxiliary languages of the mid-20th century, especially in the so-called “Esperantidos” (the constructed languages derived from Esperanto). Ogden clearly shared the same values and purposes:

The absence of a common medium of communication is the chief obstacle to international understanding, and therefore the chief underlying cause of War. It is also the most formidable obstacle to the progress of international Science, and to the development of international Commerce. (Ogden 1931: 13)

However, he opposed the assumed neutrality of Esperanto and its reformed versions (such as Ido, Mundolinco, Universal) as they were still based mainly on Indo-European languages and thus as difficult for learners from other continents as any already existing European language.

From the standpoint of an Eastern learner, [...] it is just one more European dialect, closely akin to Spanish or French, however much it may be simplified and systematised. If China were to offer world a form of Cantonese, unintelligible as such either in China or Japan, though similar in many points both to Chinese and Japanese, the rest of the world would hardly regard it as neutral; and in the same way the claim of Esperanto that it provides a neutral solution has no justification. (Ogden 1934: 10)

Moreover, he maintained that the adherents of constructed auxiliary languages did not address “the problem of simplification systematically” (1968: 7) unlike Basic; this is why Ogden's notions of “simplicity” and “simplification” are key to understand the principles underlying his approach. His distance from other linguistic projects is argued on scientific grounds, for instance while welcoming the US Bureau of Standards support of the movement for an international auxiliary language, he states the direction of his linguistic endeavours:

The achievements of science in the right direction are also usefully summarized, with respect to: the system of numbers; the metric system; the measurement of latitude and longitude; mathematical symbols; chemical formulae; time and the calendar; notation in music. (Ogden 1968: 8)

The “achievements of science” listed above are standards in the realm of mathematics, symbols and measurements; this is what Ogden praised of the scientific method and also how he conceived a global language, i.e. a matter of criteria established by an authority as a model. Based on the same principles, he therefore concluded that there were “statistical considerations” supporting English as the basis of a universal standardised language:

It is often stated that English is the language of 200,000,000 people, and this figure is then compared with the figures for French, German, Spanish, etc., with the implication that it would be invidious to be influenced by so small a lead, when the tide of national prejudice is running so high. Actually, however, English is the expanding administrative (or auxiliary) language of over 600,000,000 people, and financial reasons alone should convince even those who take statistics seriously that it is bound to expand more rapidly in the near future. (Ogden 1968: 7)

Here Ogden anticipated the international success of English as a lingua franca not only on numerical grounds in relation to the extension of the British Empire, but also in the crucial domains of finance and business due to the growing influence of the US economy on a global scale. This is also the reason why he advertised BASIC as the acronym for “British American Scientific International Commercial” with purpose of promoting it as a way to standardise the English dialects in the vital fields of economy and science as well as to improve international communication.

In addition to the statistical considerations supporting English as the basis of a universal standardised language, Ogden believed the English language possessed an inherent characteristic that made it special amongst the most widely spoken languages of the time: its natural analytic tendency that sets it apart from the highly inflected Romance languages, from the even more inflected Classical languages, but even from other Germanic languages, such as German:

the fact that English is the only major language in which the analytic tendency has gone far enough for purposes of simplification. Inflected systems are highly resistant to simplification, and their Latin origin is still only too evident in all the Romance languages of today. In the course of

centuries, however, most of the European speech systems have progressed considerably in the right direction, and the analytic tendency, [...]. English, both in its Anglo-Saxon and its Latin derivatives, has carried the process of simplification to a point where the final step was possible; and by the selection of its vocabulary from the word groups most adapted for universal purposes, irregularities of form and idiom in the Basic nucleus have been reduced to negligible dimensions. [...] The memorization of these irregularities is fortunately only a matter of days, or even hours; but since we have to admit them temporarily into Basic (i.e., until such time as Standard English, with its growing tendency to simplification, shall have progressed far enough to allow us to dispense with them if we so desire), [...]. Provided the exceptions are not too numerous, and have a significant historical background, they may even assist the memory. This does not mean that every anomaly which Basic includes can be excused or justified – many will doubtless pass away gradually, as linguistic analogy completes its inevitable work – but it serves to emphasize the negligible character of those irregularities which need give rise to real regret. (Ogden 1968: 14-15)

This is one of the most substantial differences between Ogden's morphosyntactic approach and the morphological typology of the major constructed auxiliary languages of the time. Whilst Schleyer's Volapük and Zamenhof's Esperanto were based on the grammatical models of synthetic agglutinative languages (such as Turkish, Finnish, Japanese, Tagalog, Nahuatl and Bantu languages) with a complex system of affixes (which can also be lexicalised into free semanteme in Esperanto) with neither exceptions nor irregularities, Ogden stressed the isolating analytic tendency of the English language – despite its intrinsic exceptions and irregular forms – as a natural propensity towards simplification; he even predicted that, once the usage of Basic would be sufficiently widespread, English would increasingly and naturally dispose of its imperfections. According to him, the evidence of the superiority of analytic languages lies in the Romance languages of nowadays, which progressively evolved “in the right direction”, from the intricacies of Latin's synthetic fusional morphology into more analytic structures. Consequently, another significant difference between Basic and the other constructed languages was Ogden's lexical approach: Ogden lamented that the artificial languages paid little attention to the simplification – i.e. reduction – of vocabulary (Ogden 1935: 8), which was his utmost concern. Through the

process of paraphrase, he aimed at reducing the English lexicon to its most essential lexemes: for instance, he affirmed that the word “Southerner” could be replaced with the more analytic phrase “a man from the South” and the word “octogenarian” with “an 80-year-old man” (1930: 14); for the purposes of comparison, both these expressions can be rendered in Esperanto with one word, the former being *sudaniĉo*, concatenating the following morphemes: *sud(o)* “south”, *-an-* “inhabitant”, *-iĉ-* “male”, *-o* “singular noun”, and the latter being *okdekjaraĝuliĉo*, which can be broken down to *okdek* “eighty” (more literally *ok* “eight” and *dek* “ten”), *jar(o)* “year”, *aĝ(o)* “age”, *-ul-* “person characterised by the quality expressed in the root”, *-iĉ-* “male”, *-o* “singular noun”.

The issue of Ogden’s lexical approach leads to possibly the most distinctive trait of Basic English: word selection. Whilst most of the earlier auxiliary language projects dealt equally with morphosyntax and lexicon, Ogden focused exclusively on vocabulary. Given that a constructed language with a rich vocabulary would require too great an effort on the part of learners, he maintained that the crucial element for a successful international language consisted in selecting a lexicon with a relatively small number of items that were to be “scientifically selected” (Ogden 1933: 1) according to their semiotic potential. Ogden’s views about semantics were deeply influenced by Charles Sanders Peirce’s semiotics (McElvenny 2015), which had informed his previous academic career culminating in his essay *The Meaning of Meaning* (1923) written with his collaborator (and later Basic enthusiast) Ivor Armstrong Richards³. With Basic, Ogden intended to pursue the theoretical assumptions of his semiotic work to language, by stripping English of the “word magic”, i.e. the power of words to obscure the real meaning of our utterances, with the intention of fostering an open and straightforward communication amongst people. On this basis, Basic offered a minimal core vocabulary of what he claimed to be the most useful 850 words:

It is clear that the problem of a universal language would have been solved if it were possible to say all that we normally desire to say with no more words than can be made easily legible to the naked eye, in column form, on the back of a sheet of notepaper. The fact, therefore, that it is possible

³ For a discussion on the role of Richards in the history of Basic English, see Russo (1989).

to say almost everything we normally desire to say with the 850 words on the endpapers, which occupy about three-quarters of the space on the back of an ordinary sheet of business notepaper, makes Basic English something more than a mere educational experiment. (Ogden 1933: 4)

The key factor for a successful universal language meeting his criteria of simplification is therefore a controlled and limited vocabulary to pursue a “panoptic” language, i.e. a language that can be “seen at a glance”. The term “panoptic” is a key word in Ogden’s texts about Basic English; “panoptic conjugation” is the process that he developed to shrink the lexicon to 850 words. Starting from a “root word”, he traced the “radial definition routes” from which different semantic “conjugates” (as he called them) departed, with each branch adding an extra meaning to the previous node; using the examples above, the node “man” can have two branches, one denoting origin (“Southerner”) and one denoting age (“octogenarian”) (Ogden 1930: 13-14). He called these semantic relations “conjugates” as he imagined each root word as having an inflectional paradigm with increasing additional meanings⁴. Here comes the importance of paraphrasing in the lexical selection whereupon Basic English is based: only the root words extracted though these criteria – which he called “elimination formula” (Ogden 1930: 14) – are admitted in Basic’s core vocabulary, whereas all peripheral lexemes are replaced with paraphrases through the 850 words, whereby “Southerner” becomes “a man from the South”. The only limit to this process was the extension of the paraphrase: if it required more than nine words or it sounded “awkward” (a notion that he never clarified) the original word was supposed to be retained.

Ogden’s word list also provides insight into the author’s idea of simplicity when analysing not only the selection criteria, but also the nature of the lexemes resulting from the selection. The list includes (in Ogden’s terminology) 100 “operation-words”, 600 “things” (divided into 400 “general things” and 200 “pictured things”, i.e. which can be taught through the use of visual aids) and 150 “qualifiers” (100 in the category “general qualifiers” and 50 in “opposites”). Whereas the categories of “things” and “qualifiers” can respectively be associated with the

⁴ For a more detailed description of panoptic conjugates and their relationship with Ogden’s theory of meaning, see McElvenny (2018: 77-92).

grammatical classes of nouns and adjectives, the “operation-words” are more complex to define, as it covers a blurry range of classes: verbs, prepositions, determiners, conjunctions, adverbs, cardinal directions (!), etc. By rejecting the conventional and established terms of grammar, Ogden deliberately undermines the role of morphosyntax in language learning and assumes that grammar can be acquired from vocabulary (he only mentions “five simple rules covering the formation of plurals, compounds, derivatives, comparatives, and adverbs” in Ogden 1932: 7). A glance at the numbers makes it immediately apparent that one part of speech in particular is predominant, as nouns constitute more than 70% of the lexemes listed; conversely, another grammatical class – i.e. verbs – seems to appear in a surprisingly reduced number. Only 18 of them (including two modal verbs, *may* and *will*) are listed in the section “Operation words”, which constitutes a mere 2% of the total.

Another difference between Ogden and the other creators of international constructed languages lies in that he was not so focused on phonology and spelling as he was on lexicon. One of the main concerns of the creators of international auxiliary languages – e.g. the above-mentioned Volapük, Esperanto, all the Esperantidos, Latino sine flexione (Peano 1903) Novial (Jespersen 1928), Sona (Searight 1935), Interlingua (Gode and Blair 1951), Frater (Phạm 1957), Neo (Alfandari 1961), etc. – was to establish a finite phonemic repertoire with phonemes that could be “easily” distinguished and uttered by at least the majority of speakers of European languages (with more or less successful outcomes, one may argue) as well as to create a “simple” regular spelling system mostly based on the correspondence between phoneme and grapheme, following the models of the mainly phonemic orthographies of Classical Latin, Ancient Greek, Spanish, Italian and most Slavic languages (both in Latin and Cyrillic scripts). By contrast, Ogden overlooked this aspect and justified this by adopting a completely different approach to phonology and spelling.

A chief obstacle to the spread of English has hitherto been its phonetic irregularity, the frequency with which the same symbols are used to represent different sounds, and the uncertainties of stress. There is the fact that the word *fish*, as Sir Richard Paget has noted, might appear as *ghoti* (*gh* as in *enough*, etc.); and if dealt with in the same way *foolish* might be spelled in 613,975 different ways. To master such details in a vocabulary

of 20,000 words, or even 2,000, necessitates an amount of drudgery which has given phoneticians and advocates of synthetic languages their opportunity. With the Basic vocabulary, however, such irregularities are reduced to a minimum in which, by treating each word as an individual, the learner can even profit by its peculiar appearance in written form as an aid to memory, and historical continuity can thus be preserved. The 850 sounds being fixed by the gramophone records, their written forms can be memorized as individual entities, with no special emphasis on any principle but that of stress. (Ogden 1934: 57)

Even though he claimed that Basic's vocabulary was free from irregularities, one can but notice that in the 850-word list there are homophones such as "I" and "eye". Essentially, Ogden asserts that the lack of correspondence between spelling and pronunciation does not constitute an obstacle for Basic English learners due to the limited number of lexical entries that they are required to memorise; he therefore shifts the focus from spelling to lexicon by suggesting that learners should consider the written form of each word as an individual graphic segment like symbols or logograms such as Chinese characters. However, Ogden eventually acknowledged the inherent limits of English spelling and of this approach, he thus opened up to the possibility of a forthcoming spelling reform:

Phonetic (spelling) reform can thus be left to pursue its separate path. It may find Basic a useful ally, and Basic may later profit by its progress. Hence the importance of Basic for educational work which cannot allow itself to be involved in controversies such as any violent departure from the habits of centuries must always engender. (Ogden 1968: 12)

The ultimate issue to address in the argument for the adoption of Basic English as an international standard was the nationalistic aspect. English, even in its Basic form, was still a national language. Ogden opposed this view on practical grounds, as he believed that English simply happened to be the best option in those historical circumstances:

The objection that many Frenchmen and Indians would not be in favour of the adoption of any form of English as an auxiliary language is not more serious than the objection that many diplomats and military men are not in favour of peace – as an argument against international arbitration. (Ogden 1931: 107)

In the aftermath of the First World War, the adoption of an auxiliary, albeit originally national, language had to be the priority over conflict and, finally, war. Ogden did not consider language exclusively as a means of communication, but also as the main factor influencing a person's mentality and perception of the world (the connection with Sapir is shown in Joseph 1996). In this respect he supported a mild (intended as patronising) form of cultural colonialism, which was not meant to replace and erase foreign cultures, but to enrich them for the sake of those people who were not lucky enough to be born as English speakers and, therefore, with an advanced English mindset. This view is clearly expressed by Ogden's life-long friend and collaborator Richards, who worked as a Basic English teacher in China for a decade:

[the Chinese] need an understanding of an enormous number of ideas, feelings, desires and attitudes that they can only gain through some form of Western Language. In practice this means some form of English. (Richards 1935: 45)

By forcing people to speak in the same language, Ogden and Richards believed that they could use the same words and, thus, share the same understanding and thoughts; and this would eventually end all human conflicts.

In conclusion, the analysis of Ogden's writings provides an insight into the simplification criteria of his linguistic approach, which markedly diverged from the idea of simplicity underlying the other contemporary international constructed languages. First and foremost, there was no need to create a new language *ex novo* if an existing language already displayed certain characteristics, despite the potential problems of cultural colonialism that the adoption of a national language might have caused. That language existed and it was English, a language that was supported (in his view) by a successful commercial and scientific mentality that would be beneficial for the rest of the world. Whereas all the other constructed languages struggled to gain a wider base of speakers and general popularity, English was already spoken by millions of people around the globe and was (and still is) a leading language in science and business. In addition to this, English met certain criteria of simplicity: a mostly analytic isolating language (definitely less inflectional than other Indo-European languages) with short words and a supposed limited number of irregular forms. The only final step towards "the right

direction” was to work on the vocabulary of English so as to reduce the potential misunderstanding caused by an excessive lexicon with ambiguous meanings; in this way the entire lexicon could be represented as a finite collection of symbols (perfectly in line with Pearce’s semiotics) which could also help Basic learners by decreasing the amount of lexical material to memorise. Therefore, Basic English can be viewed as a controlled language, i.e. a restricted form of the grammar and vocabulary of an existing language aimed at reducing ambiguity and complexity.

If we leave aside for a moment his patronising tone and the questionable intrinsic superiority that he attributed to his own language, Ogden actually exposed the fact that most of the other auxiliary constructed languages, which boasted cultural neutrality and a supranational status, were in fact mostly Euro-centric in all aspects of phonology, morphology, syntax and lexicon; therefore the issue of language nationality was just shifted from one nation to a group of Western nations. Furthermore, it should be acknowledged that, in a period when there seemed to be room for competition, Ogden had already realised that the commercial, scientific, political and military power of the English-speaking world had already turned English into the *de facto* lingua franca worldwide.

3. *A short history of simplicity in English*

James McElvenny posited that Ogden’s Basic must have been inspired, if not influenced, at least to some extent, by the constructed language projects of the previous centuries (2018: 23). According to McElvenny, Ogden “was almost certainly familiar” with the work of John Wilkins, George Dalgarno and Gottfried Wilhelm Leibniz (2018: 100). Ogden would have known these antecedents through his readings, editorial work and translations, which made him a member of a virtual community of knowledge stretching back into the 17th century. Ogden’s position towards his native language was naturally unthinkable in earlier centuries, when English was only gradually shifting from an insular minority language to the main language of a united nation with a strong and influential literary and scientific culture. English was still far from becoming the language of a colonial and commercial empire, and processes of standardisation were underway. Nonetheless, Ogden, his contemporaries, and their Renaissance and Enlightenment predecessors

all wrote as language reformers, and if we look beneath the individual schemes and projects, proposals for reform seem to display striking continuities, most significantly an enthusiasm for simplicity.

The universal language schemes of the 17th and 18th centuries have received extensive scholarly attention (Eco 1993; Knowlson 1975; Maat 2004; Stillman 1995 among others). More recently, Kristen Poole began to uncover the extent to which “the concerns of biblical hermeneutics made their way” into constructed language projects (2019: 123). Informed by recent historiography, Poole contended that “a long view” of the Reformation can help scholars gain insights into processes of inheritance, re-purposing and adoption of themes and terminology across the 16th and 17th century. Similarly, whilst taking a long view of the concept of simplicity in the history of linguistic ideas, we wish to gain insights into what appears to be an established tradition with its implicit and explicit sources. After examining Ogden’s idea of simplicity we will now explore the Enlightenment and the Reformation language debates as the backdrop to the development of ideas about simplicity in English.

The term “plain” has an intriguing historical dimension. Plainness was a crucial concept in Renaissance hermeneutics and it was set as an ideal in such influential treatises on language as Thomas Wilson’s *The arte of retorique* (1553). The nexus between language and religion has been defined with special acuity by Kristen Poole and Judith H. Anderson. Anderson pointed out the particular role of figurative language and investigated the functioning of metaphor in the early reforms of the established Church in Tudor England (2001 and 2005), whilst Poole explored how “efforts to define the ‘literal’ – and related or synonymous terms, such as ‘plain’ and ‘simple’” within 16th century religious treatises produced “extended meditations on the nature and structure of language” (2019: 124).

The notions of simplicity and plainness overlap very interestingly, if partially, at the level of lexis. If through word reduction Ogden aimed to target what he saw as an excessive lexicon with ambiguous meanings, the lexicographical activity of Early Modern authors reflected anxieties concerning the lexical expansion and the influx of foreign and obscure terms. Dictionaries and collections of the “hard words” in English proliferated from the early 17th century; the earliest, Robert Cawdrey’s *Table Alphabeticall* (1604), aimed to explain “hard vsuall English wordes,

borrowed from the Hebrew, Greeke, Latine, or French, &c. [...] by plaine English words” (cit. in De Witt and Noyes 1991: 13). Plain words, as opposed to “strange” words borrowed from foreign languages, supposedly granted a fairer access to knowledge and to the correct understanding of meaning, not least to the meaning of the Scripture (Blank 2006: 233-4; Brown 2001: 140; Nevalainen 2000: 341-2).

In order to explore the semantic contour of the word “plain” in a large number of texts in various centuries, we have employed the online corpus query tool Sketch Engine (Kilgarriff *et al.* 2004 and 2014), which can be used to carry out lexical and morphosyntactic analyses with a number of functions, such as word sketches (the grammatical and collocational behaviour of a word). Sketch Engine includes a large number of inbuilt corpora, such as The Historical Book Collection, which comprises texts published between 1473 and 1820 from the collections Early English Books Online (EEBO) Phase I, Eighteenth Century Collections Online (ECCO) and Readex’s Evans projects. We have focused on the adjective “plain” in the English Historical Book Collection corpus and run a word sketch in order to understand the behaviour of the term, the contexts wherein it typically occurs, and the words wherewith it is frequently paired. The partial overlap between “plainness” and “simplicity” is perhaps best illustrated by the adjectives associated with “plain” in our corpus: these are “simple”, as in the phrase “plain and simple”, “evident”, “obvious”, “clear” and “easy”. The results point then to a connection between the semantics of “plainness”, religion and language. The single most frequent term occurring in conjunction with “plain” is in fact “Scripture”, which appears as the top collocate in nominal sentences like “the Scripture is plain” and amongst the top three collocates in prepositional phrases such as “plain from the Scripture”. The top five collocates in adjective phrases are “term”, “English”, “truth”, “meaning” and “dealing”, with “Scripture” in sixth position, and amongst the top five collocates in nominal and prepositional phrases we find “word”, “text” and “context”. As an example, the noun phrase “plain words” (and variants) gives 2,327 results (2.36 per million) dating from 1485 to 1800, with the vast majority of occurrences in texts published between 1500 and 1699, and in works of Biblical exegesis, catechesis and controversial literature. Finally, the most frequent collocates in terms of verbs are the *verba dicendi* (speak, tell) alongside the verbs “seem”, “appear” and “be”.

From this simplified overview of the behaviour of “plain” in historical texts there emerges how a preoccupation with “plain speaking” has been shared amongst English writers across the centuries.

3.1. *Plainness: The 16th century*

Examples of usage of “plain” in 16th century texts highlight the strong positive connotations of the term. Plain speech was the language of truth and virtue, and somewhat a commonplace stretching back to classical Antiquity: “The communication of trueth, is simple and plain”, recited an adaptation of Cato’s *Distichs* annotated by Erasmus published in English in 1553 (retrieved from the Sketch Engine). Some might be familiar with the adage in the epistle to the reader in John Stow’s first edition of the *Chronicles*, “Of smoothe and flatteryng speache remember to take hede: For Trouthe in playn wordes may be tolde, of craft a lye hath nede” (1565: [a4]). In his famous *Epistle of Comfort* the Jesuit Robert Southwell urged to honour exemplar Catholic lives: “Let all historyes witnesse their sincere dealing, playne wordes, simple attire, frugal tables, vnfained promises, assured loue, & amity, and most intier & friendlye conuersation one with an other” ([1587]: 90).

According to rhetoricians and grammarians, plain speech meant clear and explicit language:

In speakeyng and wrytyng nothyng is more folyshe than to affecte or fondly to laboure to speake darkely for the nonce, sith the proper vse of speach is to vtter the meaning of our mynd with as playne wordes as maye be (Sherry 1550: [A8])

“Playne wordes” were thus understood to be antithetical to “dark”, obscure speaking. In John Sleidan’s *Commentaries concerning the state of religion and common wealth*, published in a translation by John Daus in 1560, it is apparent that obscurity may be caused by ambiguous words, as opposed to plain ones: “It ought to be vttered with playner wordes, to take awaye all ambiguitie” (OED s.v. ‘plain’).

Another sense of plainness was the bare language void of figures or rhetorical artifice, as underlined by John Day, who, describing the structure of epistles in *The English secretorie*, would oppose “plain terms” to rhetorical construction:

The first place is Exordium, a beginning or induction to the matter to be written of, Narratio or Propositio, each serving to one effect, wherein is declared or proponed, in the one by playne termes, in the other by inference, or comparisons, the very substance of the matter whatsoever to be handled (Day 1586: 22).

The 16th century was a period of heightened interest in language, doubtless favoured by the religious and cultural revolutions that were taking place, through the popularisation of the printing press, the vernacularisation of the Bible, the discovery of the New World and increased language contacts. As is well known, this was an age of extraordinary expansion of the English vocabulary, a phenomenon that elicited different responses from Tudor intellectuals. The pace of lexical borrowing from the classical languages gave rise to the Inkhorn Controversy, which opposed supporters of lexical innovations and those who heavily criticised what was seen as an overuse of superfluous, magniloquent terms that were hard to understand. Whilst a number of English writers readily adopted foreign influences as an opportunity for language enrichment, others advocated for a rationalisation of the spelling and lexicon for the advancement of English (Blank 2006: 222; Nevalainen 2000: 346, 359). In a famous letter to Thomas Hoby, the translator of Baldassarre Castiglione's *Courtier*, Sir John Cheke claimed to be of the opinion that "our own tung should be written cleane and pure, vnmixt and unmangeled with borowing of other tungen" (Castiglione 1561: no sig.). Sir Thomas Elyot believed that a nobleman must speak "none englisshe but that which is cleane, polite, perfectly and articulately pronounced" (1537: 18v). Even more significantly, Thomas Wilson promoted "plainnesse" in "elocution":

Doeth wit reste in straunge wordes, or els standeth it in wholsome matter, and apt declaryng of a mannes mynd? Do we not speake, because we would haue other to vnderstande vs, or is not the tongue geuen for this ende, that one might know what another meaneth? [...] Therefore to auoyde suche folie, we maie learne of that most excellent Orator Tullie, who in his thirde booke, where he speaketh of a perfect Oratoure, de lareth vnder the name of Crassus, that for the choise of wordes, foure thinges should chiefly be obserued. First, that suche wordes as we vse, shuld bee proper vnto the tongue, wherein wee speake, again, that thei be plain for all men

to perceiue: thirdly, that thei be apt and mete, moste properly to sette out the matter. Fourthly, that woordes translated from one significacion to another, (called of the Grecians, Tropes) bee vsed to beautifie the sentence, as precious stones are set in a ryng, to commende the golde. (Wilson 1553: 87-8.)

The language of the Scripture was typically described as “plain”, and any intentional or unintentional failure to understand the sacred sources was a heinous error for theologians. Allegations of tampering with the word of God and of the Apostles, of “corrupt” interpretation and deceitful translation were part of the stock vocabulary of religious controversies in the 16th century. The question was central, for instance, to the polemical battle between William Tyndale and Thomas More in the early 1530s:

And thus good chrysten reders ye maye clerely se, that all Tyndals proper processe of kynge Dauyd, concernyng the order of hys eleccyon, that he was therby preserved for euer frome all dedely synne, ys clerely comen to noughte / and all hys wordes reproued by the very playne wordes of scrypture (More 1533: cxlix).

Further,

All these textes lo do Luther and Tyndale say, that the catholyke chyrch iugle from theyr trew sense, bycause they teche them as god and the holy goost hath spoken them & verely ment and entended by them. And all these textes do these holy sectes so restore agayn to theyr ryght sense and vnderstandyng, that they clene destroye them, and constrewe them clene contrary bothe to the playne wordes and menyng (More 1533: cclxxxix)

In particular, the relationship between plainness and simplicity becomes explicit in the work of the Scottish exegete Robert Rollock, who advocated the plain meaning of Scripture. Rollock maintained that the Scripture does not have “any ambiguitie or doubtfullnesse in it”, it is “simple in sense and signification” (cit. in Poole 2019: 125). Rollock glossed the term “plaine” with sets of synonyms (i.e. “simplicitie”, “pure”, “single”) and antonyms (i.e. “ambiguity”, “doubfulness”, “amphibology”)

demonstrating “that the word [...] signifies a meaning that cannot be broken into parts but is single” or that a plain meaning “is one that has no multiplicity, no polysemy” (Poole 2019: 125). As is well known, the univocality of the Biblical vocabulary was a matter of contention between reformers and conservatives. The moot point was the authentic meaning of the Scriptures. Multiple readings were not possible in that they brought ambiguity to the Scripture; yet the Scripture was “plain”, so there must be only one pure authentic meaning.

The question of meaning, and in particular the contest between the figurative and the literal, was central to the debate on the Eucharist. In defending his *Apology of the Church of England*, the bishop of Salisbury John Jewel asked, “For what can be saide more plainely, then that, whiche Ambrose saithe, Breade, and Wine remaine stil the same they were before, and yet are changed into an other thinge?” (1567: no sig.); the controversialist William Fulke refuted an opponent’s literal reading of the words of Christ at the Last Supper:

And where he saith, that none of ye fathers teacheth, yt these words: This is my body, &c. be words figuratiue, it shal suffice, to oppose Augustine, who in plaine termes saith these words: Except ye eat ye flesh of the sonne of man, &c. are a figuratiue speach. (Fulke 1581: 458)

The notion of plainness had thus a deeply intertwined linguistic and religious valence. On the opposite side, ambiguity was a source of uncertainty, doubt and hermeneutical error.

3.2. *Simplification: The 17th century*

McElvenny pointed out that the need to establish “correct” associations between name and object had a long tradition in the history of English and Continental linguistic thought. It concerned John Locke and several of his contemporaries, “chief among these” John Wilkins and Gottfried Willhelm Leibniz, “whose constructed language projects aimed in part at creating philosophically sound linguistic expressions” and “may have provided inspiration to Ogden in designing Basic English” (2018: 23). “Some of the most prominent features of Basic”, McElvenny further argued,

the treatment of metaphor, the notion of “universal grammar” with a campaign against verbs, the emphasis on the visible – have quite probable antecedents in the Enlightenment projects of Dalgarno and Wilkins, with which Ogden was almost certainly familiar (2018: 100).

John Wilkins, a Dean and a fellow of the Royal Society, published his constructed language project in 1668, the *Essay Towards a Real Character and a Philosophical Language*. Around the same time, other attempts were undertaken to construct a language free from ambiguity. Ideas had been “in the air” for several decades and collaborative discussions started at Oxford in the mid-1650s (Cram 1985: 36). There was a whole circle of members of the Royal Academy around and behind Wilkin’s *Essay*, and other treatises appeared. One was published by the schoolmaster and preacher George Dalgarno, the *Ars Signorum* in 1661. Dalgarno’s character and language were influenced by his comparative studies of shorthand systems and his belief in the superiority of affixation, deriving from his Hebrew studies. His character had a visual component in that it relied on the representation of things, rather than sounds (Maat 2004: 40). Outside scholarly milieus, the works by the merchant Francis Lodwick deserve mention, *A Common Writing* and *The Ground-work*, published in 1647 and 1652 respectively, preceding both Wilkins and Dalgarno. In spite of their theological differences, these authors all situated their ideas of the origins and development of language within a biblical context in which the loss of the Adamic language resulting from the fall of man and the *confusio linguarum* at Babel represented the key sites (Poole 2003). They were also connected by the implicit or openly declared practical implications of their projects. It is not a chance that one of the earliest schemes was conceived by a professional engaged in trading like Lodwick; Wilkins himself would state that “that most obvious advantage which would ensue” from adopting his character and language was “of facilitating mutual Commerce, amongst the several Nations of the World, and the improving of all Natural knowledge” (*Ep. Ded.*). The need for an auxiliary language in 17th century Europe was widely felt for pragmatic reasons that included trade, international exchange and business. We will now focus on Wilkins’ *Essay* in closer detail as a test case for our longer view of the connections posited by McElvenny (2018) and Poole (2019).

Wilkins believed that natural languages have defects that can be avoided. Describing the development of his “real Character” he “found it necessary” to abstract “from those many unnecessary rules belonging to instituted Language” (1668: *Ep. Ded.*). Beginning with the letters of the alphabet, these are often “in several respects both Redundant, and Deficient” (15). Vowels “have several sounds” (15); “Some words are distinguished in writing, and not in pronunciation” and *vice versa* (16); and besides the defects of the alphabet or letters “there are several others likewise in the Words of Language”, especially “Equivocals” or “the ambiguity of words” (17). Furthermore, natural languages abound with “Anomalies and Irregularities in Grammatical construction” (18). Wilkins aimed for a regular and perfect system capable to “refer directly to what knowledge and thought are about, rather than using the imperfect medium of ordinary language” (Aarsleff 1992: 24). Wilkins’ universal language would therefore work based on a number of corrections, compared to natural languages:

1. The words of it should be *brief*, not exceeding two or three Syllables; the Particles consisting but of one Syllable.
2. They should be *plain* and *facil* to be taught and learnt.
3. They should be *sufficiently distinguishable* from one another, to prevent mistake and equivocalness; and withal *significant* and *copious*, answerable to the conceipts of our mind.
4. They should be *Euphonical*, of a pleasant and graceful sound.
5. They should be *Methodical*; those of an agreeable or opposite sense, having somewhat correspondent in the sounds of them. [...] (1668: 414, Wilkins’ italic)

Wilkins’ simplification entailed a work on phonology, semantics and vocabulary; in particular, fewer words would suffice, since “the multiplying of words, about things that are plain enough of themselves, doth but contribute to the making of them more obscure” (1668: 24) and “the most facil and natural order to be observed in this, will be, to begin with the 40 common Heads or Genus’s” identified in part II of his *Essay* (1668: 441).

3.3. *Difference and ambiguity*

In a passage that echoes the views of 16th century rhetoricians, Wilkins expressed scepticism towards ambiguity and ornaments:

And though the varieties of Phrases in Language may seem to contribute to the elegance and ornament of Speech; yet, like other affected ornaments, they prejudice the native simplicity of it, and contribute to the disguising of it with false appearances (1668: 18).

Encapsulated in this evaluation there is the idea that simplicity is original, “native”, altered by addition. Addition is neither an ornament nor a form of improvement, but rather a kind of falsification. 17th century language theory viewed the history of languages as a process of decay and degeneration, of a breaking down into dialects of an original perfection and unity. This deterioration of the universal language led to the loss of an effective means of communication. Wilkins understood then words as accidental, without any intrinsic connection with the objects and entities they signified. However, he believed that they maintained a relationship with the natural world, as underscored through the very structure of the *Essay*, in which the description of the “real character” follows Wilkins’ mapping of the entities and phenomena of the world. If words are figure of the world, differences in language are figure of different views of the world. However, Wilkins was not a pluralist: difference must be the product of some misunderstanding of the real nature of things, leading to inconsistency, contradiction and error. Aarsleff described Wilkins’ *Essay* as “the largest and most complete work in a long tradition of speculation and effort to create an artificial language that would [...] ‘repair the ruins of Babel’” (1992: 24). Wilkins seemed to believe that a common language would favour natural religion, so that disorder, strife and sectarianism would vanish.

A key passage in Wilkins’ presentation of his plan allows us to understand his faith in the potential for a perfect language to overcome division. In Wilkins’ worldview, linguistic misunderstanding was to be blamed as the source of groundless strife; his design seems to betray the conviction that language uniformity would lead to conformity in religion and politics. It ought to be remembered that Wilkins lived at a time of

turmoil, marred by an internecine religious and political conflict that had culminated in the devastating civil wars of 1642-51. Rather than discussing whether Wilkins' project had any religious motive, however, we wish to pay attention to his use of language in this extract:

this design will likewise contribute much to the clearing of some of our Modern differences in *Religion*, by unmasking many wild errors, that shelter themselves under the disguise of affected phrases; which being Philosophically unfolded, and rendered according to the genuine and natural importance of Words, will appear to be inconsistencies and contradictions. And several of those pretended, mysterious, profound notions, expressed in great swelling words, whereby some men set up for reputation, being this way examined, will appear to be, either nonsense, or very flat and jejune. (Wilkins 1668: *Ep. Ded.*)

The idea of “unmasking errors”, the reference to “false appearances” as opposed to “native simplicity” and the language of deceit and “disguise” takes us back to the Reformation controversies and to the particular concern of the English reformers with Catholic dissimulation. One of the tropes of anti-Catholic polemic was hypocrisy, which allegedly led to an uncontrollable compulsion to “disguise” the truth with false appearances. This involved them personally, Catholics were the devil in sheep's clothing, and textually in the interpretation of the Scripture. Protestants would accuse Catholics “to loade and disguise the sentence” of the Scripture “with many waste woordes, that you may in so doing hide the error” (Charke 1581: F1). On the Catholic side, failure to understand the complex meaning of the Scripture and allegations of an imperfect knowledge of classical languages and doctrines were recurring weapons against English reformers, who due to their ignorance “understand not soe cleere doctrine but huddle up matter” (Persons 1581: [B5v]).

Kristen Poole pointed out that well in the 17th century Bishop Samuel Parker's *Discourse of Ecclesiastical Politie* (1671) demonstrates the continuity of these concerns. Aligning conformists with literalism and sectaries with allegory, Parker identified “exegetical differences as ‘the most material difference’ between Church of England conformists and radical sectarians” (2019: 126-7). The bishop claimed how “we express the Duties and the Precepts of the Gospel in plain and intelligible Terms, whilst they trifle them away by childish Metaphors and Allegories” (cit. in

Poole 2019: 126). The work of Wilkins, also a man of the Church, further shows how linguistic thought was organised around the polarity between simplicity and clarity on the one hand, and complexity and obscurity on the other. In the second part of his *Essay* Wilkins presents the catalogue “of all those things and notions to which names are to be assigned” (1668: 22); names must be assigned to “universal” or “transcendental” notions, which he defines based on kinds, causes, modes and differences (1668: 25). Everything is structured into sets of binary oppositions. Amongst differences that define transcendentals we find the following opposition:

SIMPLICITY, *mere, sheer, clear, fine, plain, right, pure, unmixed, Ingredient, single, uncompounded.*
MIXEDNESS, *mingle, compound, blend, shuffle, Medly, Miscellany, promiscuous, temper; Commixtion, complex, complicate, confound, intermingle, Hodg-podge, Gallimaufry, Rhapsody, Centon, dash, brew.* (1668: 28, all Wilkins' italic)

Plainness pertains to “simplicity”, and in Wilkins' binary world, the “plain and facil” imperative of his constructed language existed in opposition to a number of concepts. As laid out further on in the *Essay*, plain is that which is “Evident, Perspicuous, clear, express, obvious, easie, facil” connected with the verbs “explain, explicate, unfold, illustrate, open, make out” and antithetical to “obscure”, that which is “Dark, abstruse, riddle, aenigmatical, deep, profound, hard, difficult, mysterious” and to “intrigue” (1668: 47). According to Poole, the poles of the conflict as identified by Wilkins (“plain” vs “obscure” and “simplicity” vs “mixedness”), “are coded exegetic terms that continue a long association of Catholics with allegory and Protestants with perspicuity” (2019: 130). The direct correlation of signifier and signified sought after by the language reformers of the Enlightenment, in other words, inherited and repurposed the early reformers' defence of plain and simple language against Catholic obscurity (2019: 127).

Although he conceded that the Bible “provided an unquestionable framework for any linguistic theory” on language origin and diversity, Maat argued that in universal language projects religious motives should not be exaggerated (2004: 9-10). We wish to stress that we do not intend to discuss motives, but to trace the connections behind what we see as a continuous concern with simplification from the earliest projects to Ogden, and we agree with Poole that these connections also point to the discourse of biblical

interpretation and religious debate. In summary, a mistrust towards the ambiguous nature of natural languages seems to have its roots in the distant past and in an anxiety over doubt and error. The pursuit of plainness became a banner and a mission that had evidently much to do with favouring clear unambiguous communication in English, a language that was only starting to gain confidence in the 16th century and would eventually rise over Latin as a means for international communication and trade through the following centuries. At the same time, the pursuit of plainness retained its deeper connections with the exact interpretation of the Scripture, of the Fathers, of the word of God. By the 19th century, the religious underpinnings of this idea appear to have been forgotten and to have completely shifted onto another plain; yet, a mistrust towards ambiguity appears to have remained solidly anchored in the mind of language reformers, and the proposals on how to correct or overcome it display striking continuities through the centuries.

4. *Conclusion*

The primary aim of this study was to show that Ogden's approach to language simplicity was deeply entrenched in the history of the English language since the Reformation. To support this claim, we have focused on two levels of analysis: linguistic aspects and contextual aspects.

The linguistic aspects revolve around Ogden's semiotic theory. The design of Basic as a systematic vocabulary with minimal grammar derived from standard English is deeply rooted in the ideas and principles outlined in Ogden's semiotic dissertation *The Meaning of Meaning* (1923). Panoptic conjugation – the method for selecting Basic's vocabulary – and lexical isomorphism were simply a practical implementation of the method of definition proposed in the book. The idea that morphosyntax can be seen as a mere natural consequence of vocabulary and that analytic structures are inherently superior are similarly overarching topics in the essay. The notion of overall language simplicity as a consequence of a "simple" vocabulary stretches back to the 16th century concept of "plain speech" as a source of truth and virtue; the opposite of this idea of "plainness" was the obscurity resulting from the ambiguity of a language prone to figures of speech and rhetorical artifice (cf. Sleidan 1560, Day 1586), which is what Ogden meant by coining the term "word-magic". Whereas in the 16th century the main concern was the danger of tampering

with the interpretation of the word of God and the scriptures (cf. More 1533), Ogden aimed at a fetishised (possibly idolised) notion of objectivity that could be attained only through a “panoptic language” to be seen “at a glance” without any ambiguity; this is why the English vocabulary needed to be stripped of unnecessary, redundant, nebulous words, exactly as Rollock (in Poole 2019: 125) argued for simplicity “in sense and signification” in the crucial lexicon of religious practice. Therefore, the concepts of “plain” and “simple” overlap with the notion of “literal” (explained in visual terms, i.e. “seen at a glance”) as opposed to “figurative”, meant as “deceptive”. Objectivity is thus what the eye can see. In the 17th century this visual aspect is even more emphasised by Dalgarno, Wilkins and Leibniz. Unlike Ogden, Wilkins’ idea of simplification encompassed phonology and syntax; nevertheless, we can clearly distinguish the same underlying approach in establishing “scientific” criteria to select or create a finite number words for an effective communication, because in this view the main source of semantic obscurity (and thus conflict) is the “multiplying of words” (Wilkins 1668: 24). A perfect language would result in a perfect thought to be shared by all human beings, thus leading to the end of divisions, whereas figurative and vague expressions are doomed to lead to misunderstanding, inconsistency, contradiction and error. Ogden seems to support the same strong version of linguistic determinism, since the overarching assumption of Basic English is that language and its structures limit and determine human knowledge or thought, as well as thought processes such as categorisation, memory, and perception (“the idea that because our thought is based on language, and because it is important for our thought to be clear, a great respect for form might be a help in the development of our minds”, Ogden 1968: 58). By using the same language, humans should be able to overcome conflict and live in peace.

In fact, the communities of language of reformers of the 19th and of the 17th centuries both worked in post-war contexts (whereby the contextual aspects). Ogden developed his project in the aftermath of the First World War, whilst Wilkins and contemporaries lived through the English civil wars. Both endorsed the universality of language as a means to favour understanding and consent, and thus to end human conflicts. Ogden’s admiration for the scientific method and the “achievements of science” had a significant role in his conception of a global language as a set of

criteria established by an authority (1968); similarly, though increasingly understood as problematic, the link between the emergence of universal language projects and the “new science” has been the object of extensive scholarly attention (Maat 2004: 7-9). If the 17th century notion of simplicity is to be understood in the scientific/philosophical sense of that which is not compound, as Wilkins vocabulary helpfully emphasises (1668: 28), simplification denotes process of reduction to “singleness”. In constructed language projects, the number of words must be reduced, but also their components, or their length, and their associated meanings. The idea of word elimination would continue all the way to Ogden, and insofar as it was heralded as a way to promote monoreferentiality and avoid ambiguity, it brought forth some of the principles that help us see constructed languages akin to scientific languages. In the 19th as in the 17th centuries, a language viable for nurturing scientific progress, international commerce, peace and prosperity, had to be a plain and simple language.

Alongside the evident similarities that can be highlighted across the centuries, however, a long view of constructed language projects enables us to appreciate a number of distinctive aspects. In particular, a comparative diachronic approach helpfully assists relating universal language schemes to the cultural and linguistic contexts in which they were conceived. Herein lies a crucial difference: if the 16th and 17th centuries reformers faced the breaking down of a “unity”, first religious and shortly after linguistic (Latin as the *lingua franca* of the Renaissance “Republic of letters”), for 19th century linguists there was not an immediate notion of cultural nor linguistic unity to refer to, and by that time English had acquired a completely different status. Ogden’s position with respect to English as the natural language basis for his controlled language project was simply inconceivable two centuries earlier.

In view of the above arguments, Ogden’s Basic English holds a special significance in the history of international auxiliary languages as he anticipated two major trends both within the movement and in the linguistics policies of the 20th century: he anticipated the “analytic shift” in constructed languages, which from Volapük and Esperanto tended to develop more isolating structures; more importantly, he envisaged the role of the English language as a global *lingua franca* in the 20th and 21st centuries.

Acknowledgements

An earlier version of this paper was presented at the conference SHESL-HTL “Simplicity and complexity in the history and theory of artificial languages” (Paris, 2020). The authors would like to express their gratitude to the participants of the conference. Their support and feedback had a significant impact on the development of this study. The authors would also like to thank the anonymous reviewers for their helpful comments and suggestions.

References

- Aarsleff, Hans, 1992, “John Wilkins”. In: Subbiondo, Joseph L. (ed.), *John Wilkins and 17th century British Linguistics*, Amsterdam / Philadelphia, John Benjamins: 3-41.
- Alfandari, Arturo, 1961, *Cours Pratique de Neo*, Bruxelles, Éditions Brepols.
- Anderson, Judith H., 2001, “Language and History in the Reformation: Cranmer, Gardiner, and the Words of Institution”. *Renaissance Quarterly* 54/1: 20-51.
- Anderson, Judith H., 2005, *Translating Investments: Metaphor and the Dynamic of Cultural Change in Tudor-Stuart England*, New York, Fordham University Press.
- Blank, Paula, 2006, “The Babel of Renaissance English”. In: Muggleston, Lynda (ed.), *The Oxford History of English*, Oxford, Oxford University Press: 212-239.
- Brown, Sylvia, 2001, “Women and the Godly Art of Rhetoric: Robert Cawdrey’s Puritan Dictionary”. *Studies in English Literature* 41/1: 133-148.
- Castiglione, Baldassarre, 1561, *The courtier of count baldessar castilio [...] done into english by thomas hoby*, London, Wyllyam Seres.
- Cawdrey, Robert, 1604, *A Table Alphabeticall*, London, I. R[oberts].
- Charke, William, 1581, *An answere to a seditious pamphlet lately cast abroade by a Iesuite with a discoverie of that blasphemous sect*, London, Christopher Barker.
- Cram, David, 1985, “Universal language schemes in 17th-century Britain”. *Histoire Épistémologie Langage* 7/2: 35-44. DOI: <https://doi.org/10.3406/hel.1985.1313>
- Day, Angel, 1586, *The english secretorie VVherin is contayned, a perfect method, for the inditing of all manner of epistles and familiar letters, [...]*, London, Robert Waldegrave. <https://search-proquest-com.pros.lib.unimi.it:2050/docview/2240921611?accountid=12459> (accessed April 19, 2020).
- De Witt, Starnes T. / Noyes, Gertrude E., 1991, *The English Dictionary from Cawdrey to Johnson 1604-1755*, Amsterdam / Philadelphia, John Benjamins

- Eco, Umberto, 1993, *La Ricerca della Lingua Perfetta nella Cultura Europea*, Bari / Roma, Laterza.
- Elyot, Thomas, 1537, *The boke named the gouvernour*, London, Thomas Berthelet
<https://search-proquest-com.pros.lib.unimi.it:2050/docview/2240916953?accountid=12459> (accessed April 19, 2020).
- Esterhill, Frank, 2000, *Interlingua Institute: A History*, New York, Interlingua Institute.
- Fulke, William, 1581, *A reioynder to bristows replie in defence of allens scroll of articles and booke of purgatorie also the cauils of nicholas sander D. in diuinitie about the supper of our lord [...]*, London, H. Middleton for George Bishop,
<https://search-proquest-com.pros.lib.unimi.it:2050/docview/2240906714?accountid=12459> (accessed April 19, 2020).
- Gode, Alexander / Blair, Hugh E., 1955 [1951], *Interlingua; a grammar of the international language*, New York, Frederick Ungar Publishing.
- Jespersen, Otto, 1928, *International Language*, London, Allen and Unwin.
- Jewel, John, 1567, *A defence of the apologie of the churche of englande*, London, Henry Wykes, <https://search-proquest-com.pros.lib.unimi.it:2050/docview/2248525278?accountid=12459> (accessed April 19, 2020).
- Joseph, John E., 1996, “The immediate sources of the ‘Sapir-Whorf Hypothesis’”. *Historiographia Linguistica* 23: 365-404.
- Kilgarriff, Adam / Rychlý, Pavel / Smrž, Pavel / Tugwell, David, 2004, “Itri-04-08 the Sketch Engine”. *Information Technology* 1: 105-116.
- Kilgarriff, Adam / Baisa, Vít / Bušta, Jan / Jakubíček, Miloš / Kovář, Vojtěch / Michelfeit, Jan / Rychlý, Pavel / Suchomel, Vít, 2014, “The Sketch Engine: ten years on”. *Lexicography* 1: 7-36.
- Knowlson, James, 1975, *Universal Language Schemes in England and France, 1600–1800*, Toronto, University of Toronto Press.
- Maat, Jaap, 2004, *Philosophical Languages in the Seventeenth Century: Dalgarno, Wilkins, Leibniz*, Dordrecht, Kluwer.
- McArthur, Tom / Lam-McArthur, Jacqueline / Fontaine, Lise, 2018 [1997], *The Oxford Companion to the English Language*, Oxford, Oxford University Press.
- McElvenny, James, 2015, “The application of C.K. Ogden’s semiotics in Basic English”. *Language Problems and Language Planning* 39/2: 187-204.
- McElvenny, James, 2018, *Language and meaning in the age of modernism. C.K. Ogden and his contemporaries*, Edinburgh, Edinburgh University Press.

- More, Thomas, 1533, *The second parte of the co[n]futacion of tyndals answere in whyche is also confuted the chyrche that tyndale deuyseth*. London, Wyllyam Rastell, <https://search-proquest-com.pros.lib.unimi.it:2050/docview/2240932958?accountid=12459> (accessed April 19, 2020).
- Myers, Adolph, 1938, *Basic and the Teaching of English in India*, Bombay, The Times of India Press.
- Nevalainen, Terttu, 2000, "Early Modern English Lexis and Semantics". In: Lass, Roger (ed.) *The Cambridge History of the English Language*, Cambridge, Cambridge University Press: 332-458.
- [OED] *Oxford English Dictionary Online* <https://www.oed.com>
- Ogden, Charles K. / Richards, Ivor A., 1989 [1923], *The Meaning of Meaning*, San Diego, Harcourt Brace Jovanovich (original published London, Kegan Paul).
- Ogden, Charles K., 1930, *Basic English: A General Introduction with Rules and Grammar*, London, Kegan Paul.
- Ogden, Charles K., 1931, *Debabelization*, London, Kegan Paul.
- Ogden, Charles K., 1932, *The ABC of Basic English*, London, Kegan Paul.
- Ogden, Charles K., 1934, *The System of Basic English*, New York, Harcourt, Brace and Co.
- Ogden, Charles K., 1935, *Basic English versus the Artificial Languages*, London, Kegan Paul.
- Ogden, Charles K., 1968, *Basic English: international second language*, Ed. E.C. Graham. New York: Harcourt, Brace and World.
- [Persons, Robert], 1581, *A brief censure vppon two bookes written in answere to M. Edmonde Campions offer of disputation*, Doway, John Lyon [Stonor Park [Pyrton], Greenstreet House Press].
- Peano, Giuseppe, 1903, "De Latino sine Flexione". *Revue des Mathématiques* 8/3: 74-83.
- Phạm, Xuân Thái, 1957, *Frater (Lingua sistemfrater). The simplest International Language Ever Constructed*, Saigon, TU-HAI Publishing-House.
- Poole, Kristen, 2019, "Theology, Plain and Simple: Biblical Hermeneutics, Language Philosophy, and Trinitarianism in the Seventeenth Century". *Reformation* 24/2: 122-137. DOI: 10.1080/13574175.2019.1665276.
- Poole, William, 2003, "The Divine and the Grammarian. Theological disputes in the 17th-century Universal Language Movement". *Historiographia Linguistica* 30/3: 273-300.
- Richards, Ivor A., 1935, *Basic in Teaching: east and west*, London, Kegan Paul.

- Russo, Paul, 1989, *I. A. Richards: his life and work*, Baltimore, Johns Hopkins University Press.
- Schleyer, Johann M., 1982 [1880], *Volapük: die Weltsprache*, Hildesheim, Georg Olms.
- Searight, Kenneth, 1935, *Sona: an auxiliary neutral language*, London, Kegan Paul.
- Sherry, Richard, 1550, *A treatise of schemes [and] tropes very profytable for the better vnderstanding of good authors, [...]*, London, Iohn Day. <https://search-proquest-com.pros.lib.unimi.it:2050/docview/2240902860?accountid=12459> (accessed April 19, 2020).
- Sleidan, John, 1560, *A famouse cronicle of oure time, called Sleidanes commentaries*, translated by John Daus, London: I. Daye.
- Southwell, Robert, 1587, *An epistle of comfort*, London, John Charlewood <https://search-proquestcom.pros.lib.unimi.it:2050/docview/2240903153?accountid=12459> (accessed April 19, 2020).
- Stillman, Robert, 1995, *The New Philosophy and Universal Languages in Seventeenth-Century England: Bacon, Hobbes, and Wilkins*, Lewisburg, Bucknell University Press.
- Stow, John, 1565, *A summarie of englyshe chronicles*, London, Thomæ Marsh. <https://search-proquest-com.pros.lib.unimi.it:2050/docview/2240931632?accountid=12459> (accessed April 19, 2020).
- Wilkins, John, 1668, *Essay Towards a Real Character and a Philosophical Language*, London, Gellibrand and Martyn.
- Wilson, Thomas, 1553, *The arte of retorique*, London, Richard Grafton.
- Zamenhof, Ludwig L., 1889 [1887], *Dr Esperanto's International Language: introduction and complete grammar*, trans. Richard H. Geoghegan, Warsaw, self-published (trans. of *Международный языкъ*, Warsaw, self-published).
- Zamenhof, Ludwig L., 1905, *Fundamento de Esperanto*, Paris, Hachette.

ALESSANDRO ZIRONI
(Alma Mater Studiorum Università di Bologna)

*Tracce e reminiscenze dei Goti in area ravennate
tra il sesto e il decimo secolo: diversità e inclusione
alla luce dei dati antroponimici¹*

This essay takes into account the Germanic personal names and their name-giving in the area of Ravenna from the sixth up to the end of tenth century. The middle of the sixth century records the peak of Gothic names in charters, in many cases connected to donations or sales of real estate due to material necessities after the loss of the Ostrogothic kingdom. Even when the Gothic community struggles to survive, the linguistic analysis on personal names shows that their language is still alive and in good health. Surprisingly enough, from the end of the same century, Germanic names completely disappear from charters: what did it happen to the community? The study of the charters of the following centuries demonstrates that Germanic names emerge again from the very end of the eighth century. New people with Germanic names are recorded, and some of them evoke the great aristocratic families of the Gothic past. This essay mainly focuses on the concepts of diversity and inclusion: the disappearing of Gothic names cannot be considered as a natural inclusion of Ostrogoths among the Roman population but reveals instead a sort of forced inclusion. The presence of name-giving referring to the Goths still at the end of the tenth century offers evidence that the Gothic past had survived in some way.

1. *Introduzione*

È il mese di maggio dell'anno 540. Dopo una lunga e controversa trattativa, l'esercito dell'imperatore Giustiniano, guidato dal generale Belisario, prende possesso della città di Ravenna. I Goti, che dal 493 dominavano quella che era stata la capitale dell'Impero romano d'Occidente, perdono il proprio potere. Il re, Vitige, è portato a Bisanzio, insieme alla moglie Matasunta, nipote di Teoderico il Grande, colui che aveva condotto gli Ostrogoti sul suolo italico. È lo storico Procopio di Cesarea che racconta questi fatti nella sua *Guerra gotica*. Giudicato autore attendibile in quanto diretto testimone di molti avvenimenti, lo è ancor

¹ Contributo sviluppato all'interno del progetto di eccellenza *DIVE-IN Diversity & Inclusion* del Dipartimento di Lingue, Letterature e Culture Moderne – Alma Mater Studiorum Università di Bologna (iniziativa Dipartimenti di Eccellenza MIUR [L. 232 del 01/12/2016]).

più per quelle notizie che non si prestano a manipolazioni ideologiche. Ai fini di questa ricerca, una di esse merita attenzione. Procopio riferisce che Belisario lasciò liberi di tornarsene a coltivare le proprie terre molti di quei Goti che si erano raccolti a difesa della città, tanto che i Romani, dopo quella partenza, si sentono rassicurati giacché, e qui le parole di Procopio nella traduzione di Domenico Comparetti, “(...) i Romani trovaronsi al sicuro, non essendo essi ormai superati in numero dai Goti in Ravenna” (Procopio di Cesarea, II, 29, 1896: 199).

Da tale asserzione prende le mosse il seguente studio. Cosa è restato, nelle tracce documentarie, della presenza dei Goti nell’area ravennate? Il tema non sembrerebbe aprire nuove prospettive di ricerca, dato che Nicoletta Francovich Onesti (2007), in un suo preziosissimo volume dedicato all’antroponimia ostrogota, ha già raccolto e discusso linguisticamente i nomi di persona testimoniati nelle fonti (come, ad es., il già citato Procopio, le *Variae* di Cassiodoro ma pure testi più inconsueti, come le *Epistole* di Gregorio Magno) spingendosi anche oltre i limiti della fine del regno degli Ostrogoti in Italia (a. 553); tuttavia, per come è organizzato il volume, non compare una sistemazione geografica e cronologica dei nomi, aspetti, invece, centrali a questa indagine. Ancor più recentemente, Wolfgang Haubrichs si è anch’egli occupato degli antroponimi nelle carte ravennate, proponendo la sua ricerca in alcuni articoli molto simili fra loro (Haubrichs 2015, 2016, 2019). La sua indagine non si limita ai nomi di persona di origine gotica, ma si amplia, con dati statistici, ai nomi di diversa provenienza linguistica, confrontando poi i risultati con la diffusione percentuale delle forme così raggruppate (germanica, latina, greca ecc.) con quelle attestate in altre aree dell’Italia settentrionale ricavandone, come esito, peraltro prevedibile, la maggiore ed estesa diffusione di forme germaniche al di fuori dell’area ravennate e della Pentapoli. Il presente studio, però, pur facendo tesoro di quei lavori, che riportano dati necessari ed acquisiti, si fa altre domande e cerca altre risposte. Si esaminerà l’antroponimia che rimanda all’ambito linguistico e culturale dei Goti non tanto per indagarne nuovamente i dati linguistici ma, partendo da essi (e pertanto riproposti laddove risultassero necessari quali base di discussione e riflessione), si analizzeranno i dati linguistici e documentari per ricostruire le sorti dei Goti in un’area geografica, quella ravennate e della Pentapoli, in cui la loro presenza era stata particolarmente significativa (nel contesto più generale della numerosità degli

Ostrogoti in Italia). Per tal ragione risulta necessario partire dagli antroponimi gotici testimoniati nei documenti vergati durante il VI secolo, per misurare la ragione della presenza di quei nomi in negozi giuridici, che aiutano a spiegare, al di là del mero dato quantitativo (peraltro limitato) i destini di una comunità. In altre parole, la presenza o l'assenza di antroponimi gotici si presta a motivare cosa accadde agli Ostrogoti nell'area presa in esame, permettendo di valutare se ci si trova dinnanzi a un processo di cancellazione o piuttosto di inclusione in relazione alla comunità numericamente e politicamente dominante. Occorre cioè chiedersi se l'inclusione degli Ostrogoti nel tessuto maggioritario della popolazione romana sia riconducibile a un andamento naturale o, piuttosto, sia assimilabile a casi di inclusioni in un qualche modo forzose, obbligate dalla necessità o da altre ragioni. Per questo motivo l'indagine qui proposta si spinge sino alle soglie dell'anno Mille perché solo una così ampia estensione temporale permette di cogliere le sopravvivenze dei Goti: della loro memoria o, al limite della loro reminiscenza.

Ravenna, presa dai Bizantini nel 540, passerà di mano nel 751, conquistata dal re longobardo Astolfo, ma siamo oramai alla fine di quel regno che sarà assoggettato ai Franchi nel 774. Occuparsi di nomi di persona in area ravennate pone perciò interrogativi interessanti sulla sopravvivenza di un'antroponimia di matrice linguistica gotica, ma anche sull'ingresso di nuovi nomi germanici e sulla loro diffusione in un tessuto sociolinguistico in cui la presenza di genti germaniche successive ai Goti lascia aperte riflessioni sulla fortuna di determinati nomi in un contesto linguistico profondamente romanizzato. L'indagine antroponimica permette di capire il successo di alcuni nomi a scapito di altri e, dunque, di misurarne il prestigio di cui sono testimoni. Insomma, spingersi nella ricerca antroponimica sino alle soglie dell'anno Mille aiuta ad indagare la sopravvivenza di un retaggio onomastico di matrice gotica in un territorio in cui quella gente germanica aveva vissuto come dominatrice nel corso di circa mezzo secolo, nuovi arrivati che avevano portato con sé una lingua vitale e documentata nella scrittura, scomparsa tuttavia nel breve corso di un secolo. L'antroponimo, se sopravvissuto agli strali del tempo, se continuato nell'uso, permette di seguire da vicino la fine della lingua gotica, così come la presenza o assenza di quei nomi nell'Esarcato prima, durante il breve dominio longobardo poi, per sfociare infine nei destini dei Franchi in Italia. Allo stesso tem-

po il nome di persona di origine germanica, permetterà di intravedere se e quanto quei nuovi dominatori hanno lasciato tracce di sé nel territorio ravennate e in quale forma linguistica. Il percorso di indagine porta perciò a riflettere su quanto può rivelare un nome di persona, in merito ai temi dell'identità culturale, dunque il mantenimento di una diversità rispetto ad altri gruppi, oppure all'inclusione di una determinata comunità (in questo caso gli Ostrogoti) all'interno di un contesto linguistico e culturale diverso. Come già ribadito, in quest'ultimo caso occorrerà chiedersi se un'eventuale inclusione rifletta una progressiva e pacifica fusione fra gli individui o sia invece dettata da dinamiche più coercitive provocando perciò inclusioni più arrendevoli e meno spontanee.

2. I documenti

La documentazione in cui si raccolgono gli antroponimi è dispersa in mille rivoli: le fonti più antiche sono rappresentate da atti vergati su papiro, materiale scrittorio che a Ravenna resta in uso piuttosto a lungo, tanto da essere impiegato anche in forma di codice: esemplare è il caso del cosiddetto *Codex Bavarus*, ora alla Staatsbibliothek di Monaco di Baviera (München, Staatsbibliothek, clm 44), databile fra il 966 e il 983 (Rabotti 1985: XXXI). Conosciuto come *Breviarium Ecclesiae Ravennatis*, raccoglie gli estremi di negozi giuridici che sanciscono i diritti di possesso da parte dell'arcivescovado nei territori della Pentapoli dell'Esarcato (per lo più enfiteusi, livelli, donazioni) di pertinenza fra la prima metà del VII secolo sino agli anni della compilazione. Occorre poi rifarsi ai documenti, papiracei o pergamenei, che registrano negozi giuridici.

Il corpus è perciò piuttosto vario ed articolato, e contempla per la pressoché totalità carte pertinenti a beni immobili, per lo più terreni, concessi dal proprietario (quasi sempre un ente ecclesiastico) a diverso titolo, soprattutto in enfiteusi o livello; seguono, come frequenza, atti di donazione verso la Chiesa e qualche altro raro negozio giuridico (come ad es. la compravendita). Il dato non deve stupire, perché sono gli archivi vescovili e monastici i luoghi in cui la documentazione si è preservata con maggiore facilità e, dunque, trasmessa ai nostri giorni. Abbiamo perso, cioè, una larga parte di rapporti intercorsi tra singole persone così come i lasciti testamentari, che non hanno superato il trascorrere del tempo e le vicende umane che spesso hanno danneggiato la conserva-

zione degli archivi (come, ad es., il sacco e la devastazione della città di Ravenna l'11 aprile 1512 da parte dei francesi). Le complesse vicende che hanno accompagnato la storia della ex capitale dell'Impero nel travagliato passaggio fra età tardo-antica ed altomedievale, con la successione del potere nelle mani dapprima di Odoacre, poi dei Goti, poi dei Bizantini, a seguire per poco tempo dei Longobardi e, infine, del controllo esercitato dai Franchi, la stessa decadenza della città – già evidente alla fine della guerra greco-gotica (a. 553) con la perdita di prestigio e centralità (connessa all'impaludamento dello strategico porto di Classe, peraltro saccheggiato più volte in età longobarda) – non giovarono di certo alla preservazione della documentazione, tant'è che lo stesso patrimonio librario prodotto in città ha lasciato sparute tracce di sé, spesso palinsesti, ora conservato in biblioteche non ravennati.

La parte più antica di questa documentazione è rappresentata da papiri. Si tratta di una collezione di 59 documenti che spaziano dalla metà del quinto secolo e arrivano all'ottavo secolo. Non tutti sono stati redatti a Ravenna, ma in un qualche modo fanno riferimento alla chiesa della città e pertanto erano conservati nell'archivio arcivescovile prima di essere dispersi in vari luoghi (Tjäder 1955: 17-23). Nomi di origine germanica compaiono però a partire dall'inizio del sesto secolo e trovano una maggiore concentrazione soprattutto nel decennio immediatamente successivo la fine della guerra greco-gotica:

Papiri con nomi germanici	
anni	n.
510 – 520	1
521 – 530	0
531 – 540	0
541 – 550	1
551 – 560	6
561 – 570	4
571 – 580	4
581 – 590	0
591 – 600	2
601 – 641	1

È vero che la certa perdita di carte non può portare a dati statistici attendibili, ciononostante l'ascesa della curva nel decennio 551-560 resta significativa, e trova una spiegazione nella contingenza intervenuta nell'anno 554, quando entra in vigore la *Pragmatica Sanctio* emanata dall'imperatore Giustiniano, atto col quale si tenta di ricostituire l'ordine giuridico riportando le lancette dell'orologio addirittura ai tempi precedenti la guerra scoppiata nel 535. La perdita di Ravenna nel 540, la definitiva sconfitta gotica nel 553, dovettero incrementare non poco le incertezze di chi, fra i Goti, deteneva beni immobili, ancor più se essi erano in proprietà dell'abborrita chiesa ariana che aveva ovviamente goduto della protezione regia ma a cui, con la fine del regno degli Ostrogoti, veniva a mancare ogni tipo di appoggio. Nei documenti ravennati non vi sono tracce di confische, neppure di restituzioni a legittimi proprietari (molti di quei beni erano peraltro stati acquisiti a seguito di opere di bonifica, come ad esempio nel caso delle terre appartenenti alla chiesa ariana di Santa Anastasia, quindi non sottratti a precedenti proprietari, oppure giunte in possesso di gente con nome germanico attraverso processi di compravendita). Dovette però essere importante, alla fine della guerra, la vendita di proprietà terriere per ragioni di sopravvivenza, materiale e sociale: tra le righe delle opere storiografiche del tempo, in particolare quelle del già citato Procopio di Cesarea ma anche di Agazia (1967), traspare come l'élite gotica, costituita da non più di 5.000 – 10.000 individui, scompare man mano, vuoi perché deceduta durante i conflitti armati, vuoi perché deportata nei territori dell'Impero romano d'Oriente (Heather 1998: 273-6). Occorre inoltre considerare che molti atti di donazione a favore della Chiesa ravennate dipendono dalla necessità di salvaguardare in un qualche modo i propri beni: alcuni papiri si riferiscono a Goti che abbandonano il credo ariano aderendo alla fede ortodossa presumibilmente non per convinzione ma per adeguarsi ai nuovi tempi e non subire ritorsioni (Cosentino 2016: 146). La fine della guerra greco-gotica parrebbe allora cancellare in maniera radicale la presenza di Goti – e di conseguenza i nomi di origine gotica – dalle testimonianze documentarie, ma merita compiere un'indagine non limitata alle grandi tendenze perché anche negli interstizi, nelle esigue tracce, si possono ricavare dati preziosi per aiutare una ricostruzione di ordine storico-culturale.

3. *L'età dei Goti*

Come si è prima accennato, nomi germanici compaiono a partire dal VI secolo, quindi a seguito della conquista dell'Italia da parte degli Ostrogoti. Gli antroponimi riconducibili a forme germaniche si interrompono però piuttosto bruscamente alla fine del medesimo secolo:

(I nomi, se declinati, sono presentati al caso nominativo)

anno	nome	ricostruzione gmc. *	informazioni desumibili dal papiro	edizione
510	Nonni (onomatopeico)	<i>nunn</i>	<i>vir clarissimus.</i> ufficiale del prefetto del pretorio a Ravenna.	Tjäder 1982: 47-48
520 (nato)	Otratarit (prob. Ostrarit)	<i>ostra-rēðaz</i> ostrogoto-consiglio		Tjäder 1955: 6
539	Tulgilo	<i>tulg-ilō-n</i> saldo+suffisso	<i>honesta femina.</i> Moglie di Parianis e madre di Domnica e Deutherius.	Tjäder 1982: 30
539	Witterit	<i>wīti-rēða-z</i> lotta-consiglio	<i>vir devotus,</i> <i>scutarius.</i> È un proprietario terriero.	Tjäder 1982: 30
542	Ardica	<i>hardu-ika</i> duro+suffisso	<i>vir honestus.</i>	Tjäder 1982: 43
542	Cessi	<i>keusan?</i> lancia? micio?	suocero di Ghiveric.	Tjäder 1982: 43
542	Ghiveric, Giberit	<i>gebō- -rika-z,</i> <i>-rēða-z</i> dono-potente/consiglio	<i>vir devotus.</i> Genero di Cessi, testimone di Waduulfus.	Tjäder 1982: 43
542	Riccifrida	<i>rīkja-friþō</i> potente-amico	<i>honesta femina.</i> Moglie di Waduulfus.	Tjäder 1982: 43
542	Waduulfus	<i>wada-wulfa-z</i> viandante-lupo	<i>vir devotus.</i> Marito di Riccifrida.	Tjäder 1982: 43
550 ca.	Sifilo	<i>siþjō-ilō-n</i> concordia+suffisso	moglie di Bilesarius, serva.	Tjäder 1955: 9
550 ca.	Nasa	<i>nasō</i> naso		Tjäder 1982: 55
551	Amalatheus	<i>amala-þewa-z</i> amalo- dedito/schiavo	<i>spodeus</i> della chiesa ariana di S. Anastasia a Ravenna.	Tjäder 1982: 34

551	Costila	<i>kust-ila</i> virtù+suffisso	<i>ustiarus</i> della chiesa ariana di S. Anastasia a Ravenna.	Tjäder 1982: 34
551	Gudelivus	<i>guḡa-leuḡa-z</i> dio-caro	<i>ustiarus</i> della chiesa ariana di S. Anastasia a Ravenna.	Tjäder 1982: 34
551	Guderit	<i>gōḡa-rēḡa-z</i> buono-consiglio	<i>ustiarus</i> della chiesa ariana di S. Anastasia a Ravenna.	Tjäder 1982: 34
551	Hosbut	<i>huzda-buḡa</i> tesoro-messaggero	<i>ustiarus</i> della chiesa ariana di S. Anastasia a Ravenna.	Tjäder 1982: 34
551	Igila	<i>eg-ila</i> riccio+suffisso	<i>spodeus</i> della chiesa ariana di S. Anastasia a Ravenna.	Tjäder 1982: 34
551	Merila	<i>mērja-ila</i> famoso+suffisso	<i>bokareis</i> (copista) della chiesa ariana di S. Anastasia a Ravenna.	Tjäder 1982: 34
551	Minnulus	<i>minn-</i> piccolo+suffisso lat.	<i>spodeus</i> e <i>clericus</i> della chiesa ariana di S. Anastasia a Ravenna. Ha anche nome Willienant.	Tjäder 1982: 34
551	Optarit	<i>uḡta- rēḡa-z, -harja-z</i> frequente- consiglio/esercito	<i>praesbyter</i> della chiesa ariana di S. Anastasia a Ravenna.	Tjäder 1982: 34
551	Sindila	<i>sinḡa-ila</i> cammino+suffisso	<i>spodeus</i> della chiesa ariana di S. Anastasia a Ravenna.	Tjäder 1982: 34
551	Sunjaifritha	<i>sunja-friḡu-z</i> vero-pace	<i>diakon</i> della chiesa ariana di S. Anastasia a Ravenna.	Tjäder 1982: 34
551	Theudila	<i>ḡeuḡō-ila</i> popolo+suffisso	<i>spodeus</i> e <i>clericus</i> della chiesa ariana di S. Anastasia a Ravenna.	Tjäder 1982: 34
551	Ufithari	<i>uḡta-harja-z</i> frequente-esercito	<i>praesbyter</i> della chiesa ariana di S. Anastasia a Ravenna.	Tjäder 1982: 34
551	Wiliarit	<i>wilja-rēḡa-z</i> volontà-consiglio	<i>clericus</i> della chiesa ariana di S. Anastasia a Ravenna.	Tjäder 1982: 34

551	Wiljarith	<i>wilja-rēða-z</i> volontà-consiglio	<i>spodeus / bokareis</i> (copista) della chiesa ariana di S. Anastasia a Ravenna.	Tjäder 1982: 34
551	Willienant	<i>wilja-nanþa-z</i> volontà-audace	<i>clericus</i> della chiesa ariana di S. Anastasia a Ravenna. Ha anche il nome Minnulus.	Tjäder 1982: 34
552	(morto ante)Bonila	<i>bonus-ila</i> lat. bonus+suffisso	<i>praeogativarius</i> a Ravenna.	Tjäder 1955: 4-5
553	Ademunt	<i>apa-mundu-z</i> nobile-difensore	figlio naturale di Aderit e fratellastro di Ranilo. Porta anche il nome lat. Andreas.	Tjäder 1955: 13
553	Felithanc	<i>felu-banka</i> molto-concorde	<i>vir sublimis</i> . Marito di Ranilo.	Tjäder 1955: 13
553	Ranilo	<i>rahnjan-ilō-n</i> depredare+suffisso	<i>sublimis femina</i> . Moglie di Felithanc e sorellastra di Ademunt, figlia di Aderit. Cattolica.	Tjäder 1955: 13
553	(morto ante)Aderit	<i>apa-rēða-z</i> nobile-consiglio	<i>vir gloriosus</i> . Padre di Ademunt e di Ranilo.	Tjäder 1955: 13
557	Faffo	<i>fanhan-</i> prendere	Porta anche il nome Rosemud.	Tjäder 1955: 7
557	(morto ante)Gudahals	<i>gōða-halsa</i> buono-collo	<i>vir illustris</i> . Marito di Gundihildi, padre di Landarit.	Tjäder 1955: 7
557	Gundihildi	<i>gunþjō-hildjō</i> battaglia-battaglia	<i>inlustris femina</i> . Moglie di Gudahals e madre di Landarit e Lendarit.	Tjäder 1955: 7
557	Landarit	<i>landa-rēða-z</i> terra-consiglio	<i>clarissimus puer</i> . Figlio di Gudahals e Gundihildi.	Tjäder 1955: 7
557	Lendarit	<i>linþi-rēða-z</i> scudo/tiglio-consiglio	<i>clarissimus puer</i> . Fratello di Landarit.	Tjäder 1955: 7

557	Gudila	<i>gōða-ila</i> buono+suffisso	figlio di Tzalico.	Tjäder 1982: 49
557	Gundila	<i>gunþjō-ila</i> battaglia+suffisso	proprietario terriero. Ariano divenuto ortodosso (<i>reconciliatus</i>) per riottenere i propri beni.	Tjäder 1982: 49
557	Rosemud	? <i>hrausa-mōða</i> gloria-animo	porta anche il nome Faffo.	Tjäder 1982: 49
557	Sitza	got. <i>sitan?</i> sedere	<i>vir honestus, comes.</i> Testimone.	Tjäder 1982: 49
557	Tzalico	? <i>þal-ika-z</i> rametto+suffisso	<i>comes.</i> Padre di Gudila, membro dell'esercito dei Goti.	Tjäder 1982: 49
557 (morto ante)	Adiut	<i>afa- -wīti, -wiðu, -weit, -wīt</i> nobile-lotta/bosco /ripagare	<i>vir inlustris.</i>	Tjäder 1955: 7
564	Ranihilda	<i>rani-hildjō</i> depredare-battaglia	<i>ancilla</i> , serva di Gratianus e Stephanus.	Tjäder 1955: 8
564	Gunderit	<i>gunþjō-rēða-z</i> battaglia-consiglio	proprietario terriero.	Tjäder 1955: 8
564	Guderit	<i>gōða-rēða-z</i> buono-consiglio	liberto di padrone romano.	Tjäder 1955: 8
572	Gunderit	<i>gunþjō-rēða-z</i> battaglia-consiglio	<i>exceptor curiae</i> della Chiesa di Ravenna.	Tjäder 1955: 14-15
572 (morto ante)	Lucerna	<i>lukarn?</i> lume	padre di Andreas <i>augustalis.</i>	Tjäder 1982: 35
575	Riccitanc	<i>rīkja-þanka</i> regno-concorde	<i>vir clarissimus.</i> Figlio di Montanus che fu <i>notarius sacri vestiarii</i> nel 540 ai tempi di re Vitige.	Tjäder 1955: 6
575-591	Oderic, Oderit	<i>auða-rīka-z</i> fortuna-potente	<i>vir clarissimus.</i> Figlio di Boherde, testimone.	Tjäder 1982: 36

575 – 591 (morto)	Boherde	<i>bowis-herdja-z</i> abitazione-pastore	padre di Oderic.	Tjäder 1982: 36
575 (morto nel)	Manna	<i>manna-n</i> uomo	<i>vir devotus</i> . Figlio del defunto Nanderit. Padrone del servo Albanio.	Tjäder 1955: 6
575 (morto ante)	Nanderit	<i>nanþa-rēða-z</i> audace-consiglio	padre di Manna.	Tjäder 1955: 6
575 -591	Hildigernus	<i>hildjō-gera-z</i> battaglia-desideroso	<i>vir clarissimus</i> . Proprietario di un terreno presso Rimini. Presente in due papiri.	Tjäder 1982: 36; Tjäder 1982: 37
600 ca.	Sisivera	<i>sisi-wērō</i> vittoria-leale	<i>honesta femina</i> . Liberta di Theudifara.	Tjäder 1955: 20
600 ca. (morta ante)	Theudifara	<i>þeudō-farō</i> popolo-viaggio	già padrona di Sisivera.	Tjäder 1955: 20
613 – 641	Wililiva, Wilileva	<i>wilja-leubō</i> volontà-cara	<i>clarissime femine donatricis gutae</i> .	Tjäder 1955: 28

Se si volessero considerare soltanto i nomi di area ravennate, bisognerebbe espungere quelli dell'anno 557, appartenenti a due papiri (Tjäder P 7 e Tjäder P 49), redatti il primo a Rieti, il secondo forse a Nepi, località posta sulla via Amerina, che collegava Ravenna a Roma attraverso l'Umbria. Tutti i nomi attestati permettono attribuzioni a forme linguistiche gotiche. Ne è una chiara traccia la presenza del suffisso vezzeggiativo masch. *-ila* (= gmc. **-ila*) o femm. *-ilo* (< gmc. **-ilō*): Bonila, Costila, Gudila, Gundila, Igila, Merila, Ranilo, Sifilo, Sindila, Theudila, Tulgilo. Si tratta di una formazione tipica dell'antroponimia germanica orientale che trova un'alta frequenza, specie per le forme maschili, fra gli antroponimi gotici (Francovich Onesti 2007: 126). Altro tratto che, al di là delle ovvie contingenze storiche, fa propendere per nomi gotici è la presenza degli elementi dei composti più produttivi fra gli Ostrogoti, quali gmc. **rēða-z* (Aderit, Giberit, Guderit, Gunderit, Landarit, Lendarit, Nanderit, Oderit, Optarit, Ostrarit, Wiliarit, Witterit). Un ulteriore elemento tipico dei nomi gotici è l'uso, come secondo membro di un nome composto, del gmc. **leuba-z* 'caro' per i maschili, gmc. **leubō* 'cara' per i femminili (Gudelivus, Wi-

liliva / Wilileva) oltre all'impiego del nome della stirpe degli Amali, gmc. **amala-* (Amalatheus). Compare invece soltanto una volta l'altrimenti frequente elemento gmc. **rīka-z* (Oderic), particolarmente produttivo in gotico.

Di particolare interesse è l'oscillazione tra i formanti *-ric* e *-rit*: si veda, ad esempio la coppia Ghiveric / Giberit, genero di Cessi e detto *vir devotus*, che compare con entrambe le forme nel medesimo documento in cui è testimone; stessa variante si trova per Oderit / Oderic, *vir clarissimus*, anch'egli testimone in un documento redatto fra il 575 e il 591. I due formanti rimandano a due radici diverse, gmc. **rēða-z* 'consiglio' e gmc. **rīka-z* 'potente' ma, specie nelle fasi più tarde della lingua gotica, i due elementi si ritrovano talvolta usati indifferentemente per lo stesso antroponimo (Francovich Onesti 2007: 72). Probabile che la confusione sia intervenuta a causa della redazione degli atti da parte di un non parlante gotico. Si tratta comunque di una oscillazione forse influenzata dalla frequenza nell'uso di forme composte con entrambi i formanti ma con lo stesso determinante. Si veda, ad es., la ricorrenza dei formanti got. *-ric*; *-rit* con il medesimo determinante gmc. **wilja-*, che non è limitata soltanto ai casi attestati nei papiri, ma si incontra, ad esempio, anche nel diffuso nome gotico Wiljarith, quello che con più frequenza si trova tra le fonti gotiche (Francovich Onesti 2013a, 21) alternandosi con Viliaric con diverse forme grafiche: un certo Viliaric, *pater pauperorum*, è citato in un'iscrizione di Vienne (Gallia Narbonense) e forse riconducibile a un burgundo (C.I.L. XII, 2150; Reichert 1987, 779) ma sempre il nome Wiliaric compare in un'iscrizione tombale romana della chiesa di S. Prassede e datata 589: si tratta del nipote del *magister militum* Trasaric (Reichert 1987: 779; Francovich Onesti 2007: 110). A questi due si affianca Uiliaric, *magister* della bottega libraria che suggella col suo nome il manoscritto col testo di Orosio ora conservato alla Biblioteca Laurenziana di Firenze (Plut. 65,1, f. 144v) (Cavallo 1983: 29-30; Francovich Onesti 2013b: 90). A queste forme con il formante *-ric* se ne aggiungono altre tre composte con *-rit(h)*: il documento ravennate papiro Tjäder P 34 dell'anno 551 presenta due nomi, Wiliarit *clericus* della chiesa gota ariana di Santa Anastasia (ai rigli 83 e 106) e Viliarit (in caratteri latini) / Wiljaritþ (in caratteri gotici) rispettivamente ai rigli 85 e 136, sempre nello stesso papiro e *bokareis*, dunque *lector*, ma anche copista (Scadigli 1973: 286-7), della stessa comu-

nità di Santa Anastasia. Sicuramente i due nomi si riferiscono a due persone differenti: il primo sigla il documento apponendovi una croce poiché incapace a firmare per *invecillitate oculorum* (Tjäder 1982: 95); il secondo, invece, firma il papiro usando l'alfabeto gotico con la formula *ik Wiljariþ bokareis handau meinai ufmelida* (io Wiljariþ copista ho firmato di mio pugno). Molti, a partire da Jan-Olof Tjäder (1982), hanno avanzato la suggestione che Uiliaric *magister* e Wiljariþ *bokareis* siano, alla fine, la stessa persona: coincide l'attività scrittoria, tanto che si potrebbe pensare a una chiusura della bottega libraria e all'ingresso del *magister* nella comunità di Santa Anastasia. Sono congetture ad oggi non dimostrabili. Certo, la frequenza del nome (addirittura due omonimi nella stessa congregazione) ma soprattutto l'uso diverso del formante (-*ric* nel manoscritto laurenziano, -*riþ* nella firma di propria mano sul papiro) debbono invitare alla prudenza: risulterebbe in effetti piuttosto curioso che il *magister* usi due forme onomastiche diverse per identificarsi. Sempre nella forma con formante gmc. *-*rēða-z*, Wiliarit, si ritrova in un'iscrizione della chiesa di S. Martino ai Monti (Roma), databile all'anno 532, e si riferisce a un eunuco che doveva essere nato intorno al 477, perciò precedentemente all'arrivo degli Ostrogoti in Italia (489) (Reichert 1987, 779; Francovich Onesti 2007, 111). Infine, è giunto il nome Wiliarit all'interno delle *Variae* di Cassiodoro (I,38); in questo caso si tratta di un adolescente e la datazione è attribuibile al lasso temporale 507-511, quindi in piena età teodericiana (Reichert 1987: 779; Francovich Onesti 2007: 111).

Da un punto di vista linguistico, l'oscillazione tra i due formanti -*ric*, -*rit*, potrebbe essere causata dall'esito di got. -*ē*- in posizione interna di parola: sebbene in gotico la vocale tenda a conservarsi, ciò non avviene nello specifico caso di gmc. *-*rēða-z* in cui il timbro vocalico compare come -*i*- che paleserebbe, nel caso di got. -*rit(h)*, una mutazione di pronuncia registrata dalla stessa firma olografa nel papiro ravennate di S. Anastasia Tjäder P 34, che va a intaccare anche la grafia stessa del nome Wiljariþ, dove ci si sarebbe attesi una forma **Wiljarēþ (Francovich Onesti 2007: 144-5). Una medesima oscillazione fra gmc. *-*ē*- e got. -*i*- si incontra poi anche negli antroponimi con secondo membro gmc. *-*mērja-z* 'famoso' che alternano got. -*mer* e got. -*mir* (Francovich Onesti 2007: 123), mentre più stabile risulta la -*ē*- nel formante gmc. *-*wērō* 'leale' ancora alle soglie del VII secolo (cfr. Sisivera).

Questa poliedricità di forme è segnale interessante di una lingua in movimento, non fossilizzatasi in forme arcaiche, e allo stesso tempo non ancora fagocitata dal predominante latino: ne sono una traccia evidente i nomi non declinati, o il mantenimento dell'uscita degli ipocoristici in *-a* per i maschili e in *-ō* per i femminili con il conseguente inserimento dei primi nella prima declinazione latina (talvolta nella terza), mentre i femminili entrano nella terza declinazione (Francovich Onesti 2007: 157-160). Ci si trova allora in un contesto linguistico ancora dinamico e vitale, in cui probabilmente già la prima generazione doveva aver acquisito un certo grado di bilinguismo passivo (Lazard 1991: 115): si ricordi, tuttavia, che gli Ostrogoti, risalendo la penisola balcanica nel V secolo si erano stanziati in aree dell'impero di lingua latina, ma l'incontro dei Goti col latino risale addirittura al III secolo (Gheller 2016: 47). La lingua gotica, da un punto di vista del numero di parlanti (occorre mettere nel novero anche i Visigoti), del prestigio culturale e per il suo uso anche liturgico è dunque impiegata ed elemento non secondario nella consapevolezza di una diversità rispetto alla maggioranza romana. Il processo di assimilazione (e dunque di inclusione nel sistema linguistico del latino) stava procedendo ma, sino alla caduta del regno ostrogoto, senza mettere in crisi la sopravvivenza della lingua gotica in Italia. La scomparsa degli antroponimi gotici in quella che potremmo determinare come la quarta generazione dall'ingresso nella penisola testimonia invece una cesura profonda che porta alla sparizione non solo della lingua ma anche alla mancata attribuzione di nomi gotici ai nuovi nati.

4. *Reminiscenze gotiche*

La già più volte rammentata fine del regno degli Ostrogoti, nel territorio ravennate pare una vera e propria cesura, anche a livello antroponimico. Tuttavia la comunità gotica, seppure uscita fortemente compromessa nella sua numerosità dal lungo conflitto, non era del tutto scomparsa, nonostante la sottrazione di tutti i luoghi di culto ariani (a parte, forse, la chiesa di S. Andrea) e il patrimonio ad essi collegato (Giovannini / Ricci 1985: 52, 56; Restelli 1984: 217). Ve ne è traccia in un celeberrimo passo del *Liber Pontificalis* di Agnello Ravennate che fa risalire tale azione ai tempi dell'omonimo vescovo Agnello (556-569) (Agnellus 2006, cap. 86, 253-4). Un concomitante papiro (ChLat 1982,

ch. 711, 94-97) databile agli anni 565-570 rammenta l'ottenimento da parte della Chiesa di tutti i beni appartenuti alla chiesa gotica (Tjäder 1971: 19; Zironi 2009: 28-29). I documenti ravennati redatti nel corso del VII secolo non lasciano però tracce di antroponimi di origine germanica; può essere che molti Goti avessero assunto nomi di tradizione latina, biblica o, al limite, greca, e che quindi sfuggano ad ogni possibile identificazione. Dobbiamo allora rivolgerci a un'altra testimonianza, purtroppo frammentaria e che non riguarda l'area ravennate ma soltanto alcuni territori della Pentapoli, ovverosia il già citato *Codex Bavarus*. Il regesto copre inoltre un periodo molto esteso, che va – come già detto – dal VII al X secolo, dunque percorre secoli che hanno visto la breve dominazione longobarda e poi della Chiesa sotto la tutela dei Franchi. Di conseguenza l'antroponimia germanica è piuttosto variegata, proponendo forme spiccatamente longobarde o tedesco superiori, come gli elementi con occlusiva labiale sorda in posizione iniziale assoluta (ad es. -*perga*; -*pert*; -*prand*) ad altri che, invece, non presentano tale assordimento e sono perciò più agilmente (ma non esclusivamente) riconducibili a una forma linguistica fràncone o, più generalmente, transalpina (ad es. -*berg*; -*bert*; non compaiono invece nomi in -*brand*). Rimanda invece all'area fràncone l'uso del formante gmc. *-*hardu* 'duro, solido' (ad es. Berardus, Bernardo, Gebeardo, Gerardo) che non compare in questa posizione nei nomi longobardi. Se ne ricava l'impressione generale che la maggioranza dei nomi germanici riflette delle forme consuete dell'antroponimia longobarda, anche se molti nomi non permettono di identificare un'esclusività di uso proprio di quella lingua, ossia si tratta di antroponimi ai quali mancano le caratteristiche fonetiche peculiari di un determinato dialetto. Un'indagine, allora, su nomi di persona forse riconducibili alla tradizione gotica deve appoggiarsi sull'uso di determinati elementi che linguisticamente o culturalmente rimandano con maggiore agio a forme già riscontrate nell'onomastica dei Goti. Tra i documenti su cui si può costruire questo corpus vi è appunto il *Breviarium Ecclesiae Ravennatis* (*Codex Bavarus*) (Breviarium 1985), tuttavia non fondamentale, giacché i nomi oggetto di questo studio registrati nelle annotazioni sono altresì pervenuti grazie alla documentazione sopravvissuta, a parte quello relativo al nome Imelperga per cui occorre ricorrere al *Breviarium* (Rabotti 1985: 28 n° 54). Elenchiamo comunque i nomi che si tratteranno a breve presenti nel *Codex Bavarus*:

nomeri	costruzione gmc. *
Adelgauso	<i>aþala-gauta</i> ‘nobile-goto’
Amelgarda	<i>amala-garði</i> ‘amalo-difesa’
Amelricus	<i>amala-rīka-z</i> ‘amalo-potente’
Imelperga	<i>amala-bergō</i> ‘amalo-protezione’
Theoderico, Theodericus	<i>þeuðō-rīka-z</i> ‘popolo-potente’

Per i secoli dall’ottavo al decimo è in effetti preferibile ricorrere alle carte ravennati, raccolte in diverse edizioni, che rimandano per lo più all’archivio arcivescovile di Ravenna, a quello di Stato sempre di Ravenna, o a quelli di alcuni monasteri cittadini (Sant’Andrea Maggiore, San Vitale e Sant’Apollinare in Classe, Santa Maria in Cereseo e San Martino dietro la Chiesa Maggiore), i cui documenti si trovano ora sparsi in diversi luoghi, tra cui, principalmente, l’Archivio Arcivescovile di Ravenna, l’Archivio di Stato di Ravenna e l’Archivio di Stato di Modena (quest’ultimo in quanto ha incamerato l’Archivio degli Estensi) (Federici / Buzzi 1911; Federici / Buzzi 1931; Muzzioli 1987; Cosma 1999; Benericetti 1999; Rabotti / Santoni 2000; Benericetti 2002a; Benericetti 2002b; Benericetti 2006; Montorsi / Rabotti 2006; Benericetti 2010). Da quei repertori si possono ricavare i seguenti nomi:

anno	nomeri	costruzione gmc. *	informazioni desumibili dalla carta	edizione
759	Anscauso	<i>ansi-gauta</i> ‘dio-goto’	vescovo di Forlimpopoli e abate	Benericetti 2006: 6, ch. 3
858	Amengausus, Elmengausus (2 volte)	<i>helma-gauta</i> ‘elmo-goto’	tabellio	Benericetti 2006 41-2: 44, ch. 17 e 18
873	Amelrico	<i>amala-rīka-z</i> ‘amalo-potente’	marito di Franca (Casumaro, FE)	Federici / Buzzi 1931: appendice, 329, ch. 1.

891	Gutta	<i>guta-z</i> 'goto'	figlia di Leone	Benericetti 2006: 112, ch. 42
894	Tedericus	<i>þeuðō-rīka-z</i> 'popolo-potente'	presbitero della chiesa di S. Pietro in Scotto (Forlì)	Benericetti 2006: 135, ch. 50
896	Vuandilo	<i>wandila-</i> 'vandalò'	deceduto, padre di Iohannes console (territorio di Faenza)	Benericetti 2006: 136, ch. 51
915	Theodericus	<i>þeuðō-rīka-z</i> 'popolo-potente'	<i>comes</i> . Atto redatto nel castello di Lausiniano, presso Lugo (RA). Il castaldo di Theodericus agisce per liberare dei suoi arimanni.	Benericetti 1999: 62-3, ch. 26
949	Teuderico	<i>þeuðō-rīka-z</i> 'popolo-potente'	<i>cubicularius</i> della chiesa di Ravenna	Benericetti 1999: 150, ch. 65
958	Adelgausus	<i>aþala-gauta</i> 'nobile-goto'	<i>honesto puero</i> (area dell'osimano)	Benericetti 2002a: 14-5, ch. 94
966	Amelgarda	<i>amala-garði</i> 'amalo-difesa'	moglie di Munaldo	Benericetti 2002a: 102, ch. 125
968	Adelgausus	<i>aþala-gauta</i> 'nobile-goto'	figlio di Gislerius <i>honestus</i> <i>vir</i> . Atto redatto nell'osimano; probabilmente coincidente con Adelgausus dell'anno 958.	Benericetti 2002a: 151-2, ch. 144
968	Amelricus	<i>amala-rīka-z</i> 'amalo-potente'	figlio di Iohannes <i>dux</i> . Testimone. Probabilmente coincidente con Amalricus dell'anno 978.	Benericetti 2002a: 153, ch. 145
978	Amalricus, Amelricus	<i>amala-rīka-z</i> 'amalo-potente'	figlio del defunto Iohannis, <i>dux, vir magnificus</i> . Testimone	Benericetti 2010: 95, 98, ch. 309
979	Amelgarda	<i>amala-garði</i> 'amalo-difesa'	moglie di Tetbaldo, figlia di Esmidionis <i>ex genere Francorum</i>	Benericetti 2002b: 53-4, ch. 211
980	Amelfredus	<i>amala-frīþu-z</i> 'amalo-pace'	figlio di Amelfredus	Benericetti 2002b: 63-4, ch. 215
992	Amalricus	<i>amala-rīka-z</i> 'amalo-potente'	<i>laudabilis vir, dux</i> , (area del forlivese)	Benericetti 2002b: 162, 164, ch. 255

La prima constatazione riguarda il sistema flessivo dei nomi. Mentre nei papiri ravennati del VI secolo alcuni nomi restavano non declinati (come quelli con formante got. *-rit*), nei documenti successivi tutti gli antroponimi sono inseriti in una declinazione latina. Nessun nome femminile presenta più l'uscita in *-ō*, così come non si trovano maschili in *-a*, entrambi tratti tipici del gotico. In poche parole, l'antroponimo sembra aver ampiamente dimenticato la morfologia della lingua gotica, tanto che non possiamo che confermare quanto giunge da altri dati secondo i quali il gotico scomparve già alla fine del VI secolo o agli inizi di quello successivo (Zironi 2009: 31) anche se resisterebbero in Italia centrale sacche antroponimiche con nomi di origine gotica con le uscite al maschile in *-a* addirittura nel corso dell'VIII secolo (Kögel 1892: 45-6; Bruckner 1895: 3) sebbene esse vadano giudicate con molta prudenza (Arcamone 1976: 151-2). Ciò, invece, non avviene nell'ambito ravennate, ma le ragioni possono essere molteplici e rimando alle conclusioni per alcune riflessioni complessive.

Restando invece sul dato linguistico, compare l'antroponimo Vuan-dilo, testimoniato, ma in forma non declinata, in gotico come Wandil (Reichert 1987: 764; Wolfram 1985: 518; Francovich Onesti 2007: 108), specificatamente nelle *Variae* di Cassiodoro (III, 38), ed è il nome di un ufficiale goto ad Avignone nel 508 (territorio a quei tempi soggetto agli Ostrogoti). L'antroponimo si trova anche tra i Longobardi, in una fonte pisana del 748, Uuandalo (Jarnut 1972: 252, n° 1463), ma quella ravennate spicca per la sua conservatività, mantenendo sia la *a* radicale che la *i* post-tonica, tipica della formazione ipocoristica del gotico (*-ila*, *-ilō*). Il suono vocalico, poi, non provoca la metafonia palatale (es. formazioni del tipo ***Wend-* ***Went-*, come, ad es. si riscontra nel termine *wentilseo* 'mare dei Vandali' nel *Hildebrandslied* (v. 43) di età carolingia (Zironi 2019: 112-3): in questo contesto si comporta similamente al longobardo (che conosce forme Uuandal- Wandel-, Guandel- Guandil-) (Francovich Onesti 2000: 220). Allo stesso tempo la *i* post-tonica non si indebolisce come spesso accade in longobardo (ove si trovano forme soprattutto con *e*; si registra soltanto un Guandilpert in un documento del 768) (Jarnut 1972: 330, n° 399), aspetto tanto più significativo alla luce della forma attestata in un documento redatto a Faenza nell'896 sebbene il nome sia di una generazione precedente in quanto designa il padre del *dux* Iohannes (Benericetti 2006: 138). Ovviamente non è pos-

sibile asserire che Uuandilo sia riconducibile alla *natio* gotica, ma di sicuro interesse risulta il contesto sociale legato al mondo militare cui appartiene. Il conservatorismo della forma non è però un dato da tralasciare per eventuali identificazioni linguistiche.

Merita poi un appunto l'elemento *-gaus* nelle forme composte maschili Adelgausus, Elmengausus, Anscausus. La ricostruzione germanica rinvia a **gauta-* 'goto', ma occorre prestare attenzione prima di giungere a possibili attribuzioni agli Ostrogoti. L'elemento infatti è altamente diffuso in tutto il bacino germanico (Förstemann 1900: 606-621) ma non in gotico, ove per l'etnonimo si usa la forma gmc. **-guta-z*, sia come nome monotematico (es. Gutta, Francovich Onesti 2007: 58) oppure come secondo elemento di un composto, in questo caso usato per nomi sia maschili (Ostrogotha) che femminili (Ostrogotho, Vulthogotho) o in femminili come Suavegotta e Sabigoto (Francovich Onesti 2007: 74-5, 90, 115). Le forme in **-gauta* non possono perciò essere derivazione diretta dal gotico, ma senza dubbio riflettono – almeno per i casi italiani – più che un diretto rimando all'etnonimo 'goto', un riferimento alla dinastia longobarda dei Gausi, che aveva dato a quelle genti due re: Alboino (530-572), che entra in Italia nel 568, e suo padre Audoino (morto nel 560, re dei Longobardi dal 547). Audoino, nel *Prologus* all'*Editto di Rotari* è detto *ex genere Gausus* (Edictum Rothari 2005: 14), dunque 'goto', perché, per linea femminile, il bisnonno di sua moglie Rodelinda era Teoderico il Grande. Audoino, insomma, si imparenta con la discendenza degli Amali (tant'è che la madre e la nonna di Rodelinda si chiamano rispettivamente Amalaberga e Amalafrida), entrando dunque in quell'ampia rete di grandi famiglie dell'élite aristocratica germanica continentale di cui gli Amali rappresentavano il ramo gotico più importante. I nomi presenti nei documenti ravennati rinviano dunque alla dinastia dei Gausi (che si interrompe bruscamente, nel suo ramo diretto, con Alboino), come conferma del resto il riscontro soltanto di forme che dimostrano l'adesione alla mutazione consonantica alto-tedesca antica che aveva interessato anche il longobardo (gmc. **-t- > long. -s-*; gmc. **-g- > long. -c- /k/*, cfr. ad es. Ans-causo).

Va osservato anche l'elemento **amala-* che, come in gotico, si ritrova utilizzato come primo membro di una forma composta; al contrario del gotico, però, non caratterizzato dall'indebolimento della vocale *-a-* atona, sono qui testimoniate forme rintracciabili nell'alto-te-

desco antico in cui la seconda *-a-* è invece sottoposta agli effetti di indebolimento di suono vocalico atono interno di parola (Braune / Eggers 1987: 63): nel nostro *corpus* si osservino le forme col formante gmc. **-rīka-z*, gmc. **-garði* e gmc. **-friþu* (Amelricus, a. 968; Amelgarda, a. 979 e Amelfredus, a. 980). Spesso, nella stessa fonte, si alternano forme con *-a-* e altre con vocale *-e-*, senza che sia possibile determinarne un criterio: le due varianti risultano concorrenti ma la presenza di forme con *-e-* allontana dalla permanenza di forme gotiche. Si deve poi rammentare che il determinante Amal- / Amel- è anch'esso diffuso sul continente europeo (Förstemann 1900: 88-95) senza che esso possa essere strettamente collegato a legami di parentela con la famiglia degli Amali. La sua estesa diffusione rende però ragione al loro prestigio e non escluderei motivi connessi alla politica estera di Teoderico il quale, mandando in sposa proprie figlie e sorelle presso altri regni romano-barbarici, ampliava l'influenza della sua schiatta (e dunque anche il suo nome) per l'Europa: tra l'altro proprio alcuni membri della famiglia unitisi in matrimonio fuori dai confini del regno ostrogoto portano nomi formati con Amal- (Ensslin 1959: 84-90; Tönnies 1989: 85-92). Se, dunque, il nome Amal- nelle carte ravennati non può direttamente legarsi alla *natio* gotica dei nominati, conferma però la fortuna del nome che ancora nel decimo secolo richiama – assai probabilmente in maniera inconsapevole – la fama di una delle élite aristocratiche germaniche altomedievali, prestigio che ne aveva decretato la diffusione.

Pare invece proprio un relitto gotico il nome femminile Gutta, proveniente dall'area di Faenza. Troverebbe infatti una corrispondenza nel nome ostrogoto maschile Guttus, forma monotematica derivata da gmc. **-guta-z* con geminazione consonantica. Qui, il nome verrebbe inserito nella prima declinazione latina. Non si può però nutrire una certezza assoluta in termini di attribuzione germanica del nome, tanto che anche per Guttus è stata avanzata l'ipotesi di una derivazione dal lat. Cottus, Cotius (Francovich Onesti 2007: 58). Del resto, nessuno dei parenti di Gutta nominati nell'atto porta un nome germanico (Benericetti 2006: 112), sebbene questo dato non sia alla fine particolarmente essenziale nel determinare l'appartenenza linguistica del nome. A favore di una possibile origine gotica gioca il mantenimento della vocale *-u-*, tipica proprio dell'etnonimo in quella lingua.

5. Conclusioni

Dopo la fine del regno ostrogoto in Italia, resta traccia documentaria di nomi di Ostrogoti che compaiono in atti che, per lo più, segnalano una palese difficoltà della comunità gotica (vendite e donazioni, in questo caso soprattutto alla chiesa ravennate). Si tratta di un passaggio storico-culturale importante che si gioca poco oltre la metà del VI secolo. La guerra greco-gotica aveva sicuramente decimato la componente maschile adulta fra i Goti e lo stesso clero ariano, come si evince dalle celebri sottoscrizioni all'atto di vendita del 551 da parte della comunità della chiesa di Sant'Anastasia, è invecchiato e senza protezioni (Lazard 1991: 127). Ne discende un forte ridimensionamento numerico della comunità, che porta alla progressiva assimilazione ai Romani e la conseguente perdita della lingua. La stessa antroponomia gotica sembra subire un collasso: con l'arrivo dei Bizantini e la definitiva cancellazione della comunità ariana (e, di conseguenza, dell'etnia gotica che in quella fede si identificava) (Berndt / Steinacher 2014: 233) viene a mancare anche l'attribuzione dei nomi gotici alle nuove generazioni.

Nel periodo di circa duecento anni, che va dalla metà del sesto secolo alla conquista del territorio dell'Esarcato da parte longobarda, scompaiono gli antroponimi germanici nelle carte prodotte a Ravenna: nella ventina di documenti pergamenei e papiracei prodotti tra la seconda metà del VI secolo fino all'ottavo secolo, raccolti in diversi volumi delle *Chartae Latinae Antiquiores* (voll. 9, 17, 20, 21, 25, 28, 29, 45, 54, 55, 90, 91, 92, 99), appare un unico nome germanico, quello di Anscuso, vescovo di Forlimpopoli, sicuramente morto nel 759 (Zaghini 1996: 118). Possiamo perciò dedurre che la residuale popolazione gotica non marca più una propria appartenenza attraverso l'impiego di nomi tradizionali ma, laddove giunge alla stesura di documenti, non è più distinguibile dalla maggioranza romana giacché ne ha adottato il sistema antroponomico, saldamente latino nonostante il dominio bizantino (Lazard 1985). A partire dall'ottavo secolo, con la conquista longobarda prima, e il controllo dei Franchi poi, le cose iniziano a mutare: nelle carte emergono, fra gli altri, nomi linguisticamente spiccatamente longobardi come Adrepaldus, Ansiperga, Ansprandus. Allo stesso tempo sono evidenti formazioni propriamente d'Oltralpe, come ad es. il nome della *comitissa* Engelrada / Ingelarada, dai chiari tratti fonetici germanico-occidentali (ad esempio il passaggio da

gmc. *-ē- > gmc. occ. -ā-) (Arcamone 2009: 78). Mancando poi l'antroponimo fra i Longobardi, se ne deduce che l'origine del nome della contessa, e anche la *natio*, fossero transalpini: è stato sostenuto che la sua casata, stanziatasi nel ravennate, avesse le proprie origini tra i Franchi ripuari (Rinaldi 2009: 20). A prima vista, allora, parrebbe che nulla distinguerrebbe il territorio ravennate dal resto dell'Italia settentrionale negli ultimi secoli del primo millennio cristiano, che conosce, in merito all'antroponimia germanica, la compresenza di nomi foneticamente longobardi e franchi. Come però si accennava agli inizi di questo lavoro, è negli interstizi della documentazione che si annidano dati interessanti. Dal secolo ottavo, per poi proseguire nei due secoli successivi, termine di questa indagine, fanno capolino nomi che richiamano una tradizione socioculturale gotica: formazioni in -**gauta-z*; -**amala-* o altri antroponimi rintracciabili tra gli Ostrogoti. Non è possibile sostenere una linea sottotraccia marcante una continuità tra il tempo dei Goti e i secoli successivi: si è visto, infatti, come *gaus-* rinvii alla stirpe dei Gausi piuttosto che all'etnonimo. Emergono tuttavia alcuni aspetti singolari, come il fatto che quei nomi si incontrino spesso in zone periferiche (come, ad esempio Comacchio, il territorio di Osimo, le campagne del ferrarese). È possibile che in aree più marginali si siano conservati più a lungo dei tratti distintivi, come testimonia del resto anche il documento del 1045 redatto a Brescello (Reggio Emilia, borgo sulle rive del fiume Po) riguardante un certo Obizo e sua moglie Dominica, di Goito (Mantova) che dichiarano *legem vivere gothorum* (Tamassia 1902: 402). Non sappiamo quanti, fra i discendenti di coloro che erano entrati con Teoderico in Italia nel 489, conoscessero la propria storia familiare nei secoli ottavo, nono e decimo, e quanti di loro ancora professassero la legge dei Goti. Qualche *enclave* potrà essere rimasta per un certo tempo: Goito, ad esempio, il cui toponimo rimanda ai Goti (Mastrelli 1994: 278) poteva essere una di quelle.

La storia del patrimonio antroponimico degli Ostrogoti in area ravennate e la sua fortuna nei secoli successivi alla perdita del regno documentano, contemporaneamente, diversità e progressiva inclusione. L'attribuzione del nome, come ha sottolineato anche Haubrichs in più occasioni (2015: 105; 2016: 253; 2019: 214), rivela l'orientamento culturale di una famiglia, la sua intenzione di integrarsi o di escludersi da un gruppo, scegliendo rispettivamente nomi diffusi presso una determinata comunità o, al contrario, nomi che marcano la diversità dalla comunità pri-

vilegiando un diverso senso di appartenenza identitaria. Il caso emblematico riguarda coloro che, fra i Goti, si convertirono all'ortodossia: abbiamo tracce documentarie dell'uso di un doppio nome, quello germanico – probabilmente attribuito dalla famiglia – e quello latino o biblico, che sarà verosimilmente assunto nel momento dell'abbandono della fede ariana (Lazard 1991: 128). Di questa usanza non vi è traccia negli anni precedenti, quando gli Ostrogoti erano parte integrante del regno d'Italia. Il progressivo abbandono della lingua gotica, che pare svilupparsi nel corso di tre, quattro generazioni a partire da quella entrata in Italia a fine V secolo (Lazard 1991: 128), la perdita del controllo politico-militare della penisola, il collasso demografico della popolazione maschile germanica, l'insicurezza economica, spingono ad abbandonare la propria diversità palesata dagli antroponimi per aderire a un'integrazione che, per molti versi, potremmo definire forzosa. Si tratta perciò di un'inclusione in larga misura dettata da ragioni di sopravvivenza, scelta che annienta la diversità culturale e linguistica. Sulla base di valutazioni statistiche (che debbono però essere assunte soltanto come linee di comportamento piuttosto che dati certi, vista la lacunosità che contraddistingue il corpus documentario su cui si misura l'indagine numerica) si può argomentare che l'area geografica ravennate e della Pentapoli resta isolata per quanto riguarda la diffusione di onomastica germanica (per lo più longobarda, poi franca) che interessa invece l'Emilia (Haubrichs 2019: 228-9); in Romagna e nella Pentapoli il numero degli antroponimi di formazione germanica non supera il quarto dei totali registrati anche nel X secolo. In tutto ciò l'antroponimia gotica non ha scampo. Permangono però elementi costitutivi dei nomi che rimandano a quella cultura e società, specialmente in due ambiti precisi: il legame dinastico (Amali e Gausi) oppure l'etnonimo (goto, vandalo); il caso di Gutta è poi un esempio raro peraltro non certissimo nella sua ricostruzione germanica. Gli Ostrogoti, in un'area come quella ravennate ove la loro presenza era stata sicuramente cospicua prima della guerra greco-gotica (Lazard 1991: 122; Maioli 1994: 232) hanno dunque lasciato flebile traccia del proprio patrimonio antroponimico, soffocato da quelle vicende umane che spesso accompagnano gli sconfitti, che talvolta in maniera eufemistica o, peggio ancora, superficiale, si definiscono inclusione. È probabile che la differenza numerica tra Goti e Romani avrebbe comunque portato alla scomparsa della lingua, ma non così automatica sarebbe invece stata la sparizione dei nomi.

Gli studi antroponomastici insegnano che laddove un nome è prestigioso per un determinato gruppo, esso persiste nell'uso (basti pensare ai nomi dei santi patroni all'interno delle comunità di riferimento). È stata giustamente rimarcata la perniciosa volontà dei nuovi dominatori bizantini di piegare lo spirito identitario degli Ostrogoti (Cosentino 2016: 147); pertanto, in maniera apparentemente paradossale, la fama dei Goti sopravvive al di fuori delle mura di Ravenna. Nella ex-capitale non resta traccia degli Amali, sottoposti a una *damnatio memoriae* di cui i mosaici della chiesa palatina teodericiana di Sant'Apollinare Nuovo sono palese evidenza: le figure dei dignitari e comandanti goti e quella dello stesso re sono cancellate (forse resta su una parete il suo volto perché identificato con Giustiniano) (Fiaccadori 1977: 167; Giovannini / Ricci 1985: 40; Maioli 1994: 236). Il nome degli Amali e dei Goti sopravvive invece tra le genti germaniche, che li portano con sé quando divengono i nuovi dominatori. Si palesano allora sommesse tracce che rinviano al tempo del regno ostrogoto, spesso lontane dalla città, piccoli resti, interstiziali, così come lo sono i frammenti di quei Goti che resistono fra le tessere musive della chiesa palatina: ora una mano, ora un braccio, spesso un'ombra, ma quel poco basta a rammentarne la loro, anonima, silente memoria.

Alessandro Zironi
Alma Mater Studiorum Università di Bologna
Dipartimento di Lingue, Letterature e Culture Moderne
Via Cartoleria 5, Bologna
a.zironi@unibo.it

Bibliografia

- Agazia Scolastico, 1967, *Historiarum libri quinque*, recensuit Rudolfus Keydell, Berolini, Walter de Gruyter.
- Agnellus Ravennatis, 2006, *Liber Pontificalis Ecclesiae Ravennatis*, cura et studio Deborah Mauskopf Deliyannis, Turnhout, Brepols [CCCM, 199].
- Arcamone, Maria Giovanna, 1976, "Antroponimia germanica a Pisa durante l'età longobarda". In: Chiarini, Paolo / Mastrelli, Carlo Alberto / Scardigli, Piergiuseppe / Zagari, Luciano (a cura di) *Filologia e critica. Studi in onore di Vittorio Santoli*, I, Roma, Bulzoni: 133-158.

- Arcamone, Maria Giovanna, 2009, “Onomastica guidinga”. In: Canaccini, Federico (a cura di), *La lunga storia di una stirpe comitale. I conti Guidi tra Romagna e Toscana. Atti del Convegno di Studi. Modigliana – Poppi, 28-31 agosto 2003*. Firenze, Leo S. Olschki: 71-89.
- Benericetti, Ruggero, 1999, *Le carte del decimo secolo nell'Archivio Arcivescovile di Ravenna, 1, 900-957*, Ravenna, Società di studi ravennati.
- Benericetti, Ruggero, 2002a, *Le carte ravennati del decimo secolo, Archivio arcivescovile, 2, aa 957-976*, Imola, University Press Bologna.
- Benericetti Ruggero, 2002b, *Le carte ravennati del decimo secolo, Archivio arcivescovile, 3, aa 976-999*, Imola, University Press Bologna.
- Benericetti, Ruggero, 2006, *Le carte ravennati dei secoli ottavo e nono*, Faenza, University Press Bologna.
- Benericetti, Ruggero, 2010, *Le carte ravennati del secolo decimo, IV, Archivi Minori (Monasteri di Sant'Andrea Maggiore, San Vitale e Sant'Apollinare in Classe)*, Faenza, s.n.
- Berndt, Guido M. / Steinacher, Rolland, 2014, “The *ecclesia legis Gothorum* and the Role of ‘Arianism’ in Ostrogothic Italy”. In: Berndt, Guido M. / Steinacher, Rolland (eds.), *Arianism: Roman Heresy and Barbarian Creed*, Farnham / Burlington, Ashgate: 219-229.
- Braune, Wilhelm / Eggers, Hans, 1987, *Althochdeutsche Grammatik*, 14. Auflage, Tübingen, Max Niemeyer.
- Breviarium, 1985, *Breviarium Ecclesiae Ravennatis (Codice Bavaro) secoli VII-X*, cur. Giuseppe Rabotti, Roma, Istituto Storico Italiano per il Medio Evo [Fonti per la Storia d'Italia, 110].
- Bruckner, Wilhelm, 1895, *Die Sprache der Langobarden*, Straßburg, Karl J. Trübner.
- Cavallo, Guglielmo, 1983, “La cultura a Ravenna fra Corte e Chiesa”. In: Capitani, Ovidio / Cavallo, Guglielmo / Fumagalli, Vito / Montanari, Massimo / Nobili, Mario / Peroni, Adriano / Ropa, Giampaolo / Vecchi, Giuseppe, *Le sedi della cultura nell'Emilia Romagna. L'alto Medioevo*, Milano, Silvana: 28-51.
- ChLat, 1982, *Chartae Latinae Antiquiores. Facsimile-Edition of the Latin Charters Prior to the Ninth Century, Part XX, Italy I*, published by Armando Petrucci / Jan-Olof Tjäder, Zürich, Urs Graf Verlag – Dietikon.
- Cosentino, Salvatore, 2016, “Social instability and economic decline of the Ostrogothic community in the aftermath of the imperial victory: the papyri evidence”. In: Herrin, Judith / Nelson, Jinty (eds.), *Ravenna its role in earlier medieval change and exchange*, London, Institute of Historical Research: 133-149.

- Cosma, Rita, 1999, cur., *Chartae Latinae Antiquiores, 2° series, Ninth Century, Part LV, Italy XXVII, Ravenna II, Roma, Città del Vaticano, Zürich, Urs-Graf Dietikon Verlag.*
- Edictum Rothari, 2005, *Edictum Rothari*. In: Azzara, Claudio / Gasparri, Stefano (a cura di), *Le leggi dei Longobardi. Storia, memoria e diritto di un popolo germanico*, Roma, Viella: 13-127.
- Ensslin, Wilhelm, 1959, *Theoderich der Grosse*, München, Münchener Verlag.
- Federici, Vincenzo / Buzzi, Giulio (a cura di), 1911, *Regesta Chartarum Italiae. Regesto della Chiesa di Ravenna. Le carte dell'Archivio Estense, I*, Roma, Ermanno Loescher.
- Federici, Vincenzo / Buzzi, Giulio (a cura di), 1931, *Regesta Chartarum Italiae. Regesto della Chiesa di Ravenna. Le carte dell'Archivio Estense, II*, Roma, Casa Libraria Editrice Italiana.
- Fiaccadori, Gianfranco, 1977, “Sulla memoria teodericiana di S. Martino in Ciel d’Oro”, *Felix Ravenna* IV s., 1/2 (113-114): 161-179.
- Förstemann, Ernst, 1900, *Altdeutsches Namenbuch, I, Personennamen*, 2. Aufl., Bonn, P. Hanstein.
- Francovich Onesti, Nicoletta, 2000, *Vestigia longobarde in Italia (568-774). Lessico e antroponomia*, 2° ediz., Roma, Artemide.
- Francovich Onesti, Nicoletta, 2002, *I Vandali. Lingua e storia*, Roma, Carocci.
- Francovich Onesti, Nicoletta, 2007, *I nomi degli Ostrogoti*, Firenze, Firenze University Press.
- Francovich Onesti, Nicoletta, 2013a, “I nomi dei papiri di Napoli e di Arezzo”. In: Francovich Onesti, Nicoletta, *Goti e Vandali. Dieci saggi di lingua e cultura altomedievale*, Roma, Artemide: 9-30.
- Francovich Onesti, Nicoletta, 2013b, “Latino e gotico nell’Italia del VI secolo”. In: Francovich Onesti, Nicoletta, *Goti e Vandali. Dieci saggi di lingua e cultura altomedievale*, Roma, Artemide: 89-100.
- Gheller, Viola, 2016, “Prima di Ulfila, accanto a Ulfila: missionari (in)volontari tra i Goti”. In: Piazza, Emanuele (a cura di), *Qui es qui ligno pugnat? Missionari ed evangelizzazione nell’Europa tardoantica e medievale (secc. IV-XIII) / Qui es qui ligno pugnat? Missionaries and Evangelization in Late Antique and Medieval Europe*, Verona, Alteritas 2016: 45-64.
- Giovannini, Carla / Ricci, Giovanni, 1985, *Ravenna*, Bari, Laterza.
- Haubrichs, Wolfgang, 2015, “Varianz und Hybridisierung eines Namensystems: Das Beispiel der Umformung der romanischen Namenwelt des Exarchats von Ravenna im frühen Mittelalter“. In: Czachur, Waldemar / Czyżewska, Marta / Zielińska,

- Kinga (eds.), *Sprache in der Zeit – Zeit in der Sprache*, Warszawa, Zakład Graficzny UW: 105-123.
- Haubrichs, Wolfgang, 2016, “The early medieval naming-world of Ravenna, eastern Romagna and the Pentapolis”. In: Herrin, Judith / Nelson, Jinty (eds.), *Ravenna its role in earlier medieval change and exchange*, London, Institute of Historical Research: 253-295.
- Haubrichs, Wolfgang, 2019, “Die frühmittelalterliche Namenwelt von Ravenna, der östlichen Romagna und der Pentapolis (Marche)”. In: Haubrichs, Wolfgang / Jochum-Godglück, Christa (hrsg. von), *Kulturelle Integration und Personennamen im Mittelalter*, Berlin – Boston, Walter de Gruyter, [Ergänzungsbande zum Reallexikon der Germanischen Altertumskunde, 108]: 214-253.
- Heather, Peter, 1998, *The Goths*, Oxford, Blackwell.
- Jarnut, Jörg, *Prosopographische und sozialgeschichtliche Studien zum Langobardenreich in Italien (568-774)*, Bonn, Ludwig Röhrscheid Verlag.
- Kögel, Rudolf, 1892, “rec. a Ferdinand Wrede, Über die Sprache der Ostgoten in Italien, Straßburg 1891”. *Anzeiger für deutsches Alterthum und deutsche Litteratur* 18: 43-60.
- Lazard, Sylviane, 1985, “Studio onomastico del «Breviarium»”. In: Vasina, Augusto (a cura di), *Ricerche e studi sul «Breviarium Ecclesiae Ravennatis» (Codice Bavaro)*, Roma, Istituto Storico Italiano per il Medio Evo: 32-61.
- Lazard, Sylviane, 1991, “Goti e Latini a Ravenna”. In: Carile, Antonio (a cura di), *Storia di Ravenna, II, 1, Dall’età bizantina all’età ottoniana. Territorio, economia e società*, Venezia, Marsilio: 109-133.
- Maioli, Maria Grazia, 1994, “Ravenna e la Romagna in epoca gota”. In: Bierbrauer, Volker / von Hessen, Otto / Arslan, Ermanno A. (a cura di), *I Goti. Catalogo della mostra Milano, Palazzo Reale, 28 gennaio – 8 maggio 1994*, Milano, Electa: 232-243.
- Mastrelli, Carlo Alberto, 1994, “I Goti e il gotico”. In: Bierbrauer, Volker / von Hessen, Otto / Arslan, Ermanno A. (a cura di), *I Goti. Catalogo della mostra Milano, Palazzo Reale, 28 gennaio – 8 maggio 1994*, Milano, Electa: 276-279.
- Montorsi, William / Rabotti, Giuseppe, 2006, “Le carte di Sant’Andrea Maggiore di Ravenna del secolo decimo. Supplemento (912-997)”. *Ravenna. Studi e Ricerche* XIII: 13-55.
- Muzzioli, Giovanni (a cura di), 1987, *Le carte del monastero di S. Andrea Maggiore di Ravenna, I, 896-1000*, Roma, Edizioni di Storia e Letteratura (1° ediz. 1961).
- Procopio di Cesarea, 1896, *La guerra gotica*, a cura di Comparetti, Domenico, II, Roma, Istituto storico italiano [Fonti per la Storia d’Italia].

- Rabotti, Giuseppe, 1985, "Introduzione". In Rabotti, Giuseppe (a cura di), *Breviarium Ecclesiae Ravennatis (Codice Bavaro) secoli VII-X*, Roma, Istituto Storico Italiano per il Medio Evo [Fonti per la Storia d'Italia, 110]: XXXI-XCII.
- Rabotti, Giuseppe / Santoni, Francesca (a cura di), 2000, *Chartae Latinae Antiquiores, 2° series, Ninth Century, Part LIV, Italy XXVI, Ravenna I*, Zürich, Urs-Graf Dietikon Verlag.
- Reichert, Hermann, 1987, *Lexikon der altgermanischen Namen, 1. Teil: Text*, Wien, Verlag der Österreichischen Akademie der Wissenschaften.
- Restelli, Giuseppe, 1984, "Sopravvivenze della cultura gotica in Italia". *Rendiconti dell'Istituto lombardo, Classe di Lettere e Scienze Morali e Storiche* 115: 207-264.
- Rinaldi, Rossella, 2009, "Note sulla nascita e l'affermazione della stirpe comitale", in Canaccini, Federico (a cura di), *La lunga storia di una stirpe comitale. I conti Guidi tra Romagna e Toscana. Atti del Convegno di Studi. Modigliana – Poppi, 28-31 agosto 2003*". Firenze, Leo S. Olschki: 19-46.
- Scardigli, Piergiuseppe, 1973, *Die Goten. Sprache und Kultur*, München, C. H. Beck.
- Tamassia, Nino, 1902, "Una professione di legge gotica in un documento mantovano del 1045". *Archivio giuridico Filippo Serafini* n.s., 68: 401-428.
- Tjäder, Jan-Olof, 1955, *Die nichtliterarischen lateinischen Papyri Italiens aus der Zeit 445-700, I, Papyri 1-28*, Lund, C.W.K. Gleerup.
- Tjäder, Jan-Olof, 1971, "Ravenna ai tempi dell'arcivescovo Agnello". In: Società di studi romagnoli (a cura di), *Agnello arcivescovo di Ravenna. Studi per il XIV centenario della morte (570-1970)*, Faenza, Fratelli Lega: 1-23.
- Tjäder, Jan-Olof, 1982, *Die nichtliterarischen lateinischen Papyri Italiens aus der Zeit 445-700, II, Papyri 29-55*, Stockholm, Paul Åström Förlag.
- Tönnies, Benrhard, 1989, *Die Amalertradition in den Quellen zur Geschichte der Ostgoten. Untersuchungen zu Cassiodor, Jordanes, Ennodius und den Excerpta Valesiana*, Hildesheim / Zürich / New York, Olms - Weidmann.
- Wolfram, Herwig, 1985, *Storia dei Goti*, ediz. italiana rivista e ampliata dall'autore, Roma, Salerno editrice.
- Zaghini, Franco, 1996, "Cronotassi dei vescovi di Forlimpopoli". *Forlimpopoli* 7: 113-128.
- Zironi, Alessandro, 2009, *L'eredità dei Goti. Testi barbarici in età carolingia*, Spoleto, Fondazione centro italiano di studi sull'alto medioevo.
- Zironi, Alessandro, 2019, *Il Carme di Ildebrando. Un padre, un figlio, un duello*, Milano, Meltemi.

ANDREA DROCCO, NEHA TIWARI
(Università Ca' Foscari Venezia, Independent scholar)

A Pragmatic Approach to Compound Verbs in Hindi/Urdu: The Case of (Inter)subjectivity¹

A number of different semantic and pragmatic factors govern the usage of compound verbs in Hindi/Urdu. In this paper, our focus is to set out and analyze one of the pragmatic factors that has been called '(inter)subjectivity' or 'prior knowledge' in detail and see how that correlates to the existing literature on compound verbs and to the emerging study of the codification of surprise ('mirativity') in Hindi/Urdu. Our main finding is that compound verb constructions with vector verbs jānā, denā and lenā are used in scenarios where the speaker and the listener are already aware of the event that is being spoken of. That is, the speaker's and the listener's minds are 'prepared' with respect to the context of the speaker's utterances.

Diversi fattori semantici e pragmatici regolano l'uso dei verbi composti in hindi/urdu. L'obiettivo del presente contributo è quello di definire e analizzare nel dettaglio uno di tali fattori di natura pragmatica, denominato in letteratura '(inter)subjectivity' o 'prior knowledge', e illustrare come esso possa contestualizzato all'interno della ricerca finora compiuta sui verbi composti e alle recenti analisi relative alla codificazione della "sorpresa" (i.e. mirativity) in hindi/urdu. La conclusione principale di questo articolo è che gli enunciati costruiti con gli ausiliari (in inglese 'vector verbs') jānā, denā e lenā sono utilizzati in scenari in cui il parlante e il destinatario sono già "mentalmente preparati" in merito a quanto riportato nell'enunciato.

1. Introduction²

Compound verbs in Hindi/Urdu (= H/U) have been a topic of extensive research for several decades. A lot of work has gone into defining and delimiting which constructions qualify as 'compound verb'

¹ This paper is a result of the discussions between the authors over a period of three years about compound verbs and their use in Hindi. Neha Tiwari is a native speaker of Hindi and teaches Hindi and Urdu as a second language.

² Our thanks first and foremost to Pinuccia Caracchi, Peter Hook, V.R. Jagannathan, Rupert Snell and other teachers at the Zabaan School for Languages for their insights. All errors and inadequacies are our responsibility.

The following abbreviations are used in this article: ACC = accusative; CP = conjunctive participle; DAT = dative; EMPH = emphatic; ERG = ergative; F = feminine; FUT = future; GEN =

constructions (see, for example, Hacker 1961; Hook 1974, 1978, 1991; Kachru / Pandharipande 1980; Nespital 1997; Slade 2016). At the same time, many scholars talk about scenarios where compound verbs are more likely to be used over simple verbs (cf., among others, Pořízka 1967-69; Hook 1974, *in press*; Kachru 1979, 1981; Poornima 2012). However, so far, a conclusive understanding of the factors determining the use of compound verbs still evades us. Needless to say, this makes the usage of these constructions confounding for learners of H/U, especially beginners. If it is true that the use of compound verbs may have semantic implications (presence of compound verbs connotes ‘completion’ or perfectivity), it is also true that there are other factors that govern their usage, and these are more pragmatic in nature and depend upon the informational context of the action. The aim of the present paper is to examine and discuss in depth one of the pragmatic factors that governs the use of compound verbs in H/U, pertaining to what Hook (1974) and Jagannathan (1981) have called ‘prior knowledge’³ and some other linguists call the ‘prepared mind’⁴ (see Slobin / Aksu 1982, Aksu-Koç / Slobin 1986, Bashir 1988, DeLancey 1997) and also ‘(inter)subjectivity’ (Nuyts 2001, 2012). In particular, our main goal is to develop Jagannathan’s analysis further. In a nutshell, Jagannathan (1981) has proposed that some compound verb constructions in H/U are more likely to be used in scenarios where the speaker and the listener already have information about the event being talked about or are familiar with it. This is something we discuss in detail in section 3. Moreover, anticipating a critique of the analysis presented in this paper, we also discuss the

genitive; HON = honorific; IMPF = imperfective; IMPV = imperative; INF = infinitive; INT = interrogative; LOC = locative; M = masculine; MIR = mirative; NEG = negation; OBL = oblique; PART = participle; PAST = past; PAST.PART = past participle; PERF = perfect; PL = plural; PRES = present; PROG = progressive; REL = relative; SG = singular; VV = vector verb.

³ In an earlier version of this paper, we used the term ‘prior information’ to talk about this pragmatic factor which was a translation of Jagannathan’s (1981) *pūrv jñān*, who most likely adopted it from Hook’s (1974) ‘prior knowledge’. In the present version, we have decided to use ‘(inter)subjectivity’ (Nuyts 2001, 2012) to talk about this factor as we think it is the most suitable term for what Jagannathan (1981) had in mind when he described *pūrv jñān*.

⁴ It is important to note here that Slobin / Aksu’s characterization of the idea of ‘prepared mind’ (1982; see also Aksu-Koç / Slobin 1986) is focussed on the ‘mental preparedness’ of the speaker (cf. Nuyts 2001). On the other hand, Jagannathan (1981) talks about the phenomenon of ‘prior knowledge’ from the point of view of the existence of shared information between the speaker and the listener. See below for more discussion on this topic.

findings of Montaut (2006) with respect to mirativity and the aorist⁵ in H/U and see how that relates to the use of compound verbs. We discuss some of Montaut's examples to show how mirativity and '(inter)subjectivity' or 'prior knowledge' can co-exist in certain scenarios.

The paper is organized as follows: In section 2 we provide a brief survey of the study of compound verbs that has been done so far in terms of the semantic and pragmatic factors that affect their usage. Subsequently, we explore one of the pragmatic factors, '(inter)subjectivity', that governs the use of compound verbs and develop our analysis in section (§ 3). We also try also to understand the case of the peculiar verb *girnā* 'to fall' with the aim of identifying some of the factors that possibly drive the choice between using *girnā* on its own and with other vector verbs. In the next section (§ 4), we look at how mirativity and '(inter)subjectivity' or 'prior knowledge' can co-exist in certain contexts. Finally (§ 5), we conclude this paper with our main findings.

2. *Compound Verbs in Hindi/Urdu: A bird's-eye view*

One of the salient characteristics of the New Indo-Aryan languages and one of their true innovations is the use of the so-called 'compound verbs' (Masica 1976, 1991: 325-6; Slade 2013). Compound verbs are a widely studied and discussed phenomenon, especially in H/U.⁶

As Hook (1974, 1991, 2001) and others (see, for example, Slade 2016) have noted, one of the defining characteristics of a compound verb construction is that the vector verb loses its primary meaning and becomes delexicalized. Following Hook (2001: 101) it is correct

⁵ The simple perfect/aorist is the perfective in H/U without the auxiliary verb, for example:

vah āyā 'He came' = Simple Perfect/Aorist.

Contrasted with:

vah āyā hai 'He has come' = Present Perfect;

vah āyā thā 'He was come' = Past Perfect.

⁶ For an overview of the study on compound verbs in Indo-Aryan, from both a synchronic and diachronic point of view, see Slade (2016), who also provides the reader a useful listing of studies on this topic, with respect to the Indo-Aryan group, selected by language (Slade 2016: 567). A very recent description of Dravidian compound verbs is Annamalai (2016), while for the most important studies on the presence and use of compound verbs in Tibeto-Burman and Munda languages see Hock (2016: 549-50).

to say that “[...] a compound verb (is) a sequence of two verbs AB (‘polar A’ plus ‘vector B’) that alternates with A (the ‘polar’) with little or no difference in meaning, that is, a difference not easily translatable into languages which do not have compound verbs.”⁷ The principal vector verbs of modern Indo-Aryan languages are those that, as full lexical verbs, mean ‘give’, ‘take’, ‘go’, ‘come’, ‘fall’, ‘rise’, ‘sit’, etc. Now consider the following H/U examples where the non-compound verb *lauṭā* (cf. 1) alternates with the compound verb *lauṭ gayā* (cf. 2):

- | | | | |
|-----|--|------------|-------------------------|
| (1) | <i>vah dukān</i> | <i>par</i> | <i>lauṭ-ā.</i> |
| | 3SG shop | LOC | go back-PERF.M.SG |
| | He went back to the shop. (adapted from McGregor 1977: 31) | | |
| | | | |
| (2) | <i>vah dukān</i> | <i>par</i> | <i>lauṭ ga-y-ā.</i> |
| | 3SG shop | LOC | go back go.VV-PERF-M.SG |
| | He went back to the shop. (adapted from McGregor 1977: 50) | | |

In (2) *jānā* does not convey its lexical meaning ‘go’, it has become dellexicalized and is serving a grammatical function. As we mentioned in the introduction, our aim in the current paper is to focus on one such grammatical function of the compound verb construction in H/U. However, before embarking on that, we present a brief summary here of the work that has already been done so far on the use of compound verbs and the different meanings they convey.

Here are the main features of compound verb constructions according to the literature so far:⁸

⁷ But see note 8.

⁸ In addition to what we have listed here, there are some other minor factors that are often correlated with the use of compound verb construction in H/U (such as using *ho jānā* to convey a change of state, and also other idiomatic usages). We have also talked about atypical compound verb constructions – where a transitive polar verb is compounded with an intransitive vector verb – in another paper (cf. Drocco 2018, Drocco / Tiwari 2020). Also, please note that in this paper we have focused on compound verb constructions where the polar verb appears as the verb-stem. But there are certain other compound verb constructions that involve the use of the perfective participle instead of the verb stem of the polar verb. These are beyond the scope of this paper. For a general overview of compound verb constructions and their usages please see, among others, McGregor (1977), Nespihal (1997) and Caracchi (2002).

Expression of Perfectivity: Compound verbs are used when the speaker wants to express that an action is completed or will be completed fully,⁹ without any exceptions (cf. Pořízka 1967-69; Hook 1974, 1993). Compound verbs shift the focus of the sentence towards the completion of the action, and consequently, as it has been noted by several scholars (Hook 1974, 1988; Nespital 1997; Caracchi 2002; McGregor 1977), the use of compound verbs with particles of negation and with the progressive aspect is very rare and specialized. We will briefly illustrate this with some examples taken from Hook (1974):

- (3) *maĩ ne parāñhā khā-y-ā lekin sār-ā nahĩ*
 1SG ERG parantha.M eat-PERF-M.SG but. all-M. NEG
 ‘I ate the *parantha* but not all of it.’ (adapted from Hook 1974: 167)
- (4) *maĩ ne parāñhā khā li-y-ā*
 1SG ERG parantha.M eat take.VV-PERF-M.SG
 ‘I ate up the *parantha*.’ (adapted from Hook 1974: 167)
- (5) *maĩ ne parāñhā khā li-y-ā lekin*
 1SG ERG parantha.M eat take.VV-PERF-M.SG but
sār-ā nahĩ???
 all-M NEG (adapted from Hook 1974: 167)

In (3) the subject did do the action but did not complete it all the way. Thus (5), which has *khā liyā* instead of *khāyā* as in sentence (3) above, is not compatible in the context of an incomplete action. Example (4), on the other hand, implies that the whole *parantha* (or as much as was expected, see below § 3) was eaten. Hook says:

⁹ In this paper we have not looked at apprehension constructions that are formed using *kahĩ...na/to nahĩ*. These constructions require a careful analysis of the informational context in which they are uttered. However, as Hook (*in press*) has noted, it is quite common for compound verbs to manifest in such constructions. We think that perfectivity is one of the factors that drives this usage in most cases, as the speaker is apprehensive or worried about the complete realization of a particular event. That said, there is also some expectation that is built into the context before the expression of apprehension is verbally realized, i.e. the speaker perceives or comes to know about something before the apprehensive state is triggered in him/her. We will discuss below how the presence of such an expectation can trigger the use of compound verbs. Additionally, Nespital (1997) notes that when simple verbs are used in the apprehension construction, the speaker is only expressing the possibility of something bad happening; he “does not assume a strong emotional position, such as fear or apprehension.” We hope to fully discuss this construction in another paper.

For them (some speakers) expression of transitive activity by a compound implies the exhaustion of the object or the exhaustive performance of the activity.

(Hook 1974: 166)

Predicate Orientation: As we noted above, *denā* ‘to give’ and *lenā* ‘to take’ are also used as vector verbs. The verb *denā* is normally used as a vector verb when the benefit of doing the action goes to someone other than the doer (cf. example 7), while *lenā* is used when the action is done for the benefit of the agent (cf. example 6) or simply put the benefit of the action comes back to the agent (cf. Butt / Ramchand 2001; Poornima 2012).

(6) *kavitā paṛh l-o*
poem read take.VV-IMPV
‘Read the poem to yourself.’
(adapted from Snell / Weightman 2010: 153)

(7) *kavitā paṛh d-o*
poem read give.VV-IMPV
‘Read the poem out.’
(adapted from Snell / Weightman 2010: 153)

Look at other two examples:

(8) *mā̃ ne ham-ē sār-ī kahānī sunā d-ī*
mother.F. ERG 1PL-DAT all-F story.F relate give.VV-PERF.F
‘Mother related the whole story to us.’
(adapted from Snell / Weightman 2010: 154)

(9) *m-er-ī jeb se cābī le lo*
1SG-GEN-F pocket from key take take.VV-IMPV
‘Take the key from my pocket.’
(adapted from Snell / Weightman 2010: 154)

In (8), the subject (the mother) is telling the story to the listeners (her children) and the benefit is going to them, hence the use of *denā*. In (9), the second person subject is being asked to take the key. The vector verb

lenā in the imperative (i.e. *lo*) adds the sense that he/she should take it for their benefit. In this particular case, it might benefit them because they asked for it or they need it for some work.

In addition to predicate orientation, as Kachru (1979) notes, *lenā* can also be used in cases where the speaker wants to focus on ‘modest capability or modest achievement’. For example:

- (10) *vah thorī-bahut hindī bol le-t-ā h-ai*
 3SG little Hindi speak take.VV-PRES-M.SG be.PRES-3SG
 ‘He can (or he manages to) speak a little Hindi.’
 (adapted from Kachru 1979: 164)

This seems especially true with verbs such as *jānnā* ‘to know’, *samajhnā* ‘to understand’, *pahcānnā* ‘to recognize’ (cf. Kachru 1981, Nespital 1997; Caracchi 2002; McGregor 1977), as in the following example taken from Kachru (1981):

- (11) *maĩ ne samjhā-y-ā to*
 1SG.DIR ERG explain-PERF-M.SG then
us ne savāl acchī tarah se
 3SG.OBL ERG question(M) well manner with
samajh li-y-ā
 understand take.VV-PERF-M.SG
 ‘I explained and he understood the question well’.
 (adapted from Kachru 1981: 187)

Certain vector verbs like *ḍālnā* add additional nuances to the meaning of the simple verb. As Nespital (1997) points out, the semantic meanings that the vector verbs help to bring out are already inherent in the simple verb. In these cases, vector verbs help to ‘monosemize’ a simple verb, i.e., they help to restrict the number of possible semantic meanings the simple verb can take in a particular sentence and context. For example, the aforementioned *ḍālnā* when used as a vector verb in a compound verb construction adds to the semantic verb of the polar verb the meaning that something is done thoroughly or intensely and sometimes even violently (cf. McGregor 1977; Nespital 1997; Butt / Ramchand 2001; Caracchi 2002; Poornima 2012). An unambiguous example of the specialization of meaning of an original simple verb

through the addition of the vector verb *ḍālnā* is *mārnā*. As simple verb *mārnā* means ‘to beat’, ‘to strike’, ‘to shoot’, ‘to kill’, ‘to bring down (a bird by killing)’ among others (cf. 12). However, when compounded with *ḍālnā*, the resulting restricted meanings are ‘to kill outright, to slay or murder (which is more intense than just hitting or striking)’ (cf. 13) (cf. Nespital 1997).

- (12) *vah sāmne jo laṛkā jā rah-ā h-ai,*
3SG in front REL boy go PROG-M.SG be.PRES-3SG
aḡar vahā jā-kar tum use do jhāpar mā-r-o,
if there go-CP 2PL 3SG-ACC 2 slap strike-IMPV
to māi abhī tum-hē do sau rupy-e
then 1SG right now you-DAT 2 hundred rupee-PL
d-ū-g-ā.
give-1SG-FUT-M.SG
‘The boy who is going/walking in front of us, if you go over there and slap him twice, I will give you 200 rupees right now’.
(adapted from Nespital 1997: 1002; originally from a native speaker Jitendra Bahri, Reader, Iswar Saran Degree College, University of Allahabad, English translation by Neha Tiwari)

- (13) *‘apn-ī gāy th-ī, māṛ ḍāl-ī, phir*
own-F cow.F be-F kill cast.VV-PERF.F then
kisī dūsre k-ā jānvar to nahī mā-r-ā?’
else anyone GEN-M.SG animal EMPH not kill-PERF.M.SG
‘It was my/our own cow, we killed it, we did not kill anyone else’s animal, did we?’
(adapted from Nespital 1997: 1003; originally from Premchand’s *Godān*, English translation by Neha Tiwari)

In the first sentence, *mārnā* just means ‘to strike’ (in (12) ‘to strike with a slap’: *jhāpar mārnā*). Contrast this with the second sentence where *māṛ ḍālnā* means ‘to kill’.

In addition to the semantic and aspectual meanings we have discussed above, some scholars (especially, if not only, Hook 1974 and Jagannathan 1981, as we have already said above; cf. also Nespital 1997 and Caracchi 2002) have also mentioned the informational context as a feature correlated with the presence/absence of compound verbs.

However, the majority of them have not examined this topic in depth.¹⁰ In accordance with the aims of the present paper, we will discuss this in the next section.

3. *Compound Verbs and (Inter)subjectivity*

Hook (1974) and Jagannathan (1981) in their analysis of compound verbs present the notion of ‘prior knowledge’ or lack of it. Thus, the first part of the current section is devoted to presenting and developing their idea and connecting it with the concept of ‘(inter)subjectivity’ as illustrated by Nuyts (2001, 2012) by giving more examples with our analysis.

Hook (1974: 248) says that in cases where the performance of an action is completely unforeseen by the speaker, he may not use compound verbs. He gives several examples of scenarios where a simple verb will be preferable to compound verbs. Here is one of them:

- (14) *maĩ ne śrĩ kathāvate kā nām kabhĩ nahĩ*
 1SG ERG Mr. Kathavate GEN-M.SG name.M never
sun-ā lekin kal un-k-ā
 hear-PERF.M.SG but yesterday 3PL.OBL-GEN-M.SG
khat mil-ā
 letter.M get-PERF.M.SG
 ‘I had never even heard of Mr. Kathavate before but yesterday I got a letter from him.’(adapted from Hook 1974: 248-249)

As we can see from (14), the speaker in this case was not even aware of the existence of some Mr. Kathavate. For that reason, he was not expecting any letter from him. When he gets the letter, it is entirely unexpected, and he expresses that using the simple verb *mil-ā*.

Jagannathan (1981) says that the defining characteristic of a vector verb based on ‘prior knowledge’ (*pūrv jñān*) context is that the speaker will use this construction in front of the listener only when he knows that the listener is aware/informed of the said action already.¹¹ He gives the following example to illustrate this idea:

¹⁰ Hook (*in press*) concludes that ‘prepared mind’ is a factor that can be correlated to the ‘manifestation’ of compound verbs in H/U, but he rightly concludes that it is not the only factor.

¹¹ Original Hindi text: अर्थात् रजक क्रिया का प्रयोग वक्ता श्रोता के सामने तभी करेगा, जब उसे मालूम हो कि सामने वाला भी कथ्य व्यापार से पहले से ही परिचित है ।

- (15) *mehmān ā ga-e*
guest.PL. come go.VV-PERF.M.PL
'The guests are here.' (adapted from Jagannathan 1981: 264)

Analyzing this sentence Jagannathan (1981) says:

When a speaker says this sentence then what is implied is that the members of the family already knew about the guests coming and were perhaps waiting for them. Or they had this inkling that the guests might drop by, and they had talked about it. On seeing guests that were not expected, the speaker would perhaps have said (cf. 16):¹²

(Jagannathan 1981: 264, 265, English translation: Neha Tiwari)

- (16) *ham-ār-e ghar mehmān ā rah-e haī*¹³
IPL-GEN-PL.OBL house guest.PL come PROG-M.PL 3PL
'Guests are coming to our house.' (adapted from Jagannathan 1981: 265)

We will call situations or contexts that are similar to (15) above as places where 'prior knowledge' is present and the context of (16) as cases where 'prior knowledge' is absent. Bashir (1988) also talked about this phenomenon based on Slobin and Aksu's (1982) characterization of 'prepared mind' (cf. also Aksu-Koç and Slobin 1986). However, as Nuyts (2001: 36-37) has pointed out, the concept of 'prepared mind' is about the speaker's subjective state of knowledge. Hence, he suggests the use of another term, i.e. '(inter)subjectivity', for scenarios where there is shared information between the speaker and the listener, and he points out that this phenomenon is observed in some Asian languages (2000: 36-37). Given the close parallel between Jagannathan's idea of 'prior knowledge' and the concept of '(inter)subjectivity', we have decided to use the latter term in our paper.

¹² Original Hindi text: कोई व्यक्ति घर वालों को सूचित करते हुए कहे मेहमान आ गये, तो उसका तात्पर्य यह है कि आने वाले मेहमानों की सूचना और शायद इतज़ार औरों (श्रोताओं) को भी है। या मेहमानों के आ टपकने की आशंका थी और इस बात की चर्चा हो चुकी थी। अप्रत्याशित रूप से आने वाले मेहमानों को देखकर वह शायद यही कहता, हमारे घर मेहमान आ रहे हैं।

¹³ We are aware of the fact that it would have been better had example (16) been in the same tense, the perfective, and not in the progressive. However, we have taken this example as it is from Jagannathan (1981).

Along the lines of Hook (1974, *in press*), Jagannathan (1981), Slobin / Aksu (1982), and Nuyts (2001, 2012) we define ‘(inter)subjectivity’ in the context of H/U as follows:

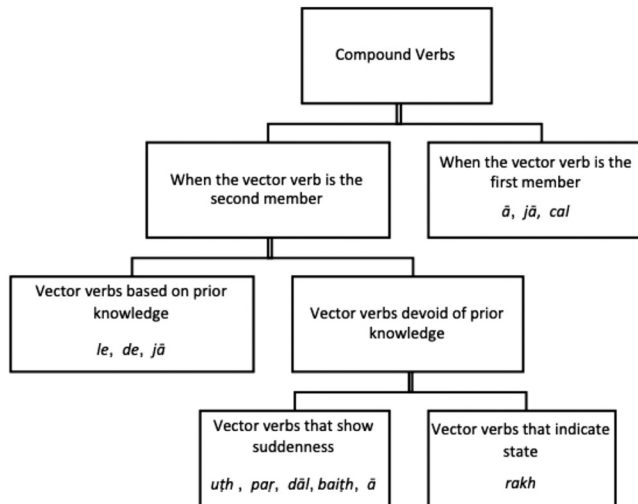
In the context of any sentence (or utterance), *(inter)subjectivity* is that state of the world in which the speaker (or the writer) and the listener (or the reader) both already have some information about the event that is being spoken of in the sentence, and they both know this to be true for the other person as well, i.e., the speaker knows that the listener is aware about the event and vice versa. In H/U, (inter)subjectivity can be encoded by the use of compound verb constructions with vector verbs *jānā*, *denā* and *lenā* restricted as follows:

jānā is used with intransitive verbs;¹⁴

denā is used with transitive verbs;

lenā is used with transitive verbs and some intransitive verbs.

We have taken this restriction based on the work of Jagannathan (1981: 264) who categorizes compound verbs as follows:



¹⁴ For an analysis on the use of *jānā* with transitive verbs see Kachru (1979: 162), but in particular Montaut (1991, 2004), Drocco (2018) and Drocco / Tiwari (2020).

Let us look at some more constructions from Jagannathan (1991) that exemplify the use of compound verbs in cases where an event is known and/or expected.¹⁵ These examples are taken from a self-study resource for Hindi learners. In this work, Jagannathan does not go into a detailed explanation of the concept of ‘(inter)subjectivity’ as some of his previous works (cf. Jagannathan 1981). Therefore, the analysis of (17) (and similarly the analysis of (18) and (19) below), is completely by the authors of this paper.

(17) a. *nāditā: āp ko vetan mil ga-y-ā?*
HON DAT salary.M get go.VV-PERF-M.SG

b. *śarmā: abhī nahī mil-ā*
now NEG get-PERF.M.SG

c. *nāditā: kyō? kyā bāt h-ai?*
INT INT matter be.PRES-3SG

d. *śarmā: āj baīk band h-ai kal*
today bank closed be.PRES-3SG tomorrow
mil jā-e-g-ā
get go.VV-3SG-FUT-M.SG

a. Nandita: ‘Did you get (your) salary?’

b. Sharma: ‘No, not yet.’

c. Nandita: ‘Why? What’s the matter?’

d. Sharma: ‘The bank is closed today. I will get it tomorrow (*don’t worry*).’¹⁶

(adapted from Jagannathan 1991: 278, English translation by Neha Tiwari)

In this situation, Nandita and Sharma are married and Nandita knows when Sharma usually gets his salary. When she asks the question (cf. 17a), she already expects that the event would have been completed; this is

¹⁵ Moreover, as already said (see note 2), Jagannathan is in conversations about the correlation of concept of ‘prior knowledge’ and the use of compound verbs in Hindi with one of the present authors of the paper.

¹⁶ In this last sentence we added ‘don’t worry’ in the translation, and we will talk about that below. See the analysis of (17d) and (23).

exactly a case of (inter)subjectivity, and for this reason she uses *mil gayā* and not *milā*.¹⁷

The expectation that the event would have been completed, conveyed by compound verbs, is also connected in H/U with an additional function of compound verbs—their use to add a sense of reassurance that the event is completed (or will be completed when talking about a future event). Accordingly, when the speaker announces the culmination or completion of an action that was expected by both the speaker and the listener, the use of compound verbs is reassuring. We can see the interplay between the sense of finality/completion on the one hand and that of reassurance on the other, by noticing the use of *mil jānā* instead of the simple verb *milnā* in the future tense *mil jāegā* of (17d). It adds the nuance that the event will be completed, and Nandita need not worry about it.

Some other examples from Jagannathan (1991) illustrate the fact that compound verbs are more likely to be used in the context of (inter)subjectivity (contrast 18 with 19):

(18) a: *tum ne kal kyā k-iy-ā?*
2PL ERG yesterday INT do-PERF-M.SG

b: *maĩ ne ek kitāb paṛhī.*
1SG ERG a book.F read-PERF.F

a: ‘What did you do yesterday?’

b: ‘I read a book.’

(adapted from Jagannathan 1991: 277, English translation by Neha Tiwari)

(19) a: *maĩ ne tumh-ē ek kitāb dī th-ī.*
1SG ERG 2PL-DAT a book.F give-PERF.F be.IMP.F

vah kitāb tum ne paṛh l-ī?
3SG book.F 2PL ERG read take.VV-PERF.F

¹⁷ In his response, Sharma uses simple verb *milā* in ‘*abhī nahīñ milā*’. Readers can refer to Hook (1974) for more information about the non-occurrence of compound verbs in negative sentences barring a few exceptions.

b: *hā̃ parh l-ī bahut acch-ī h-ai.*
yes read take.VV-PERF.F very good-F be.PRES-3SG

a: 'I had given you a book. Did you read it?'

b: 'Yes, I read it. It is very good.'

(adapted from Jagannathan 1991: 277, English translation by Neha Tiwari)

In (18) the first speaker, 'a', has no information about what the other person, 'b', did. As a consequence, when 'b' answers, he is giving new information to 'a'. In (19), however, we can see that 'a' had given the book to 'b' to read, and thus, she could anticipate that 'b' may have read the book. The use of compound form *parh lenā* in this case captures and expresses a couple of nuances:

- i) 'a' had prior knowledge about the action, she thought it may have happened, and she is expressing that in the question;
- ii) There is also a subtle implication that 'a' expects the answer to be 'yes'.

To sum up the analysis above, we have defined '(inter)subjectivity' as a situation where the speaker knows that the listener is aware of the context in which a certain action takes place. In other words, both the speaker and the listener share a certain informational context with respect to that action, and when they talk about it, this shared information is marked by the presence of a compound verb. Thus, in cases where both the simple and the compound verbs are grammatically possible, a speaker's choice for using a compound verb would be governed:

- 1) by the presence of shared information among other factors and/or
- 2) by the fact that the speaker announces the culmination or completion of an action that was expected by both the speaker and the listener, and hence the use of compound verbs is also reassuring.

We will build on both points of this analysis further with (20) and (21):

(20) *sāhab, āp se koī mil-n-e*
Sir HON with someone meet-INF-OBL

- ā-y-ā* *h-ai*.
come-PERF-M.SG be-PRES.3SG
'Sir, someone has come to meet you.'¹⁸
- (21) *āp* *kā* *kouriar ā* *ga-y-ā!*
HON GEN-M.SG.DIR courier come go.VV-PERF-M.SG
'Your courier came!' (i.e. 'Your courier is here!')¹⁹

With respect to point 1 above, it will be useful to analyze the information structure in both examples (20) and (21) by adopting the following parameters:

- i. Speaker's information set: Was the action new information for the speaker or he/she was already anticipating it?
- ii. Listener's information set: Was the action new information for the listener or he/she was already anticipating it?
- iii. Shared information: Was the speaker aware that the listener knows about the action or is expecting it?

When we analyze example (20) in terms of these parameters, we can make the following observations:

- i. Speaker's information set: Given the use of “*koī*” which means ‘someone’, it seems quite likely that the speaker had no prior knowledge about this person's arrival. It is, however, possible to imagine a scenario where the speaker can pretend no foreknowledge of the event by using this phrasing. In that case he/she would be feigning ignorance;
- ii. Listener's information set: There is no way to tell from the statement if the listener was aware of and/or anticipating the new person's arrival. If he/she was, he/she did not inform the speaker;
- iii. Shared information: The use of the present perfect with a simple verb and not compound verb in this case suggests that the speaker thinks that this arrival is new information for the listener. Hence, the speaker and the listener do not have shared/common information about the event.

¹⁸ This sentence has been created by us. However, it is very common for the contexts we have used them in. Readers can find a similar usage here: <https://hindi.pratilipi.com/read/दस-बजकर-दस-मिनट6-दस-बजकर-दस-मिनट6-Nmz8XGMYZDWb-5r558s07815m2g0>.

¹⁹ This example has also been created by us. A similar usage in the same context can be found here: <https://www.amarujala.com/haryana/jhajjar-bahadurgarh/crime/71536611917-jhajjar-bahadurgarh-news1>

In (21), we see the use of compound form *ā jānā* in the simple perfect or aorist.²⁰ An analysis of the information characteristic of the context of this sentence reveals the following information:

- i. Speaker's information set: The use of compound verb *ā jānā* in the sentence suggests that the speaker was already anticipating the event;
- ii. Listener's information set: The use of compound verb *ā jānā* indicates that the listener is most likely aware of the event because the speaker when using compound verb knows that the listener is aware of the context;
- iii. Shared Information: The speaker in this sentence knows that the listener is aware of the event and is already anticipating it. There is shared information between the two of them.

With regard to the point b) above, the reassuring sense implied in the announcement by the speaker of the culmination/completion of an action that was expected by both the speaker and the listener, the literature on compound verbs so far does not explain why in certain cases speakers feel the need to use compound verbs to add this sense of completion, and why in other cases they use simple verbs. In order to understand the reason for this contrast, consider again (21) above. If instead of *ā gāyā* (cf. 21), the speaker had said *āyā* (cf. 22), it would have sounded incomplete and, in some cases, could have made the listener apprehensive.

(21) *āp k-ā kouriar ā ga-y-ā!*
HON GEN-M.SG.DIR courier come go.VV-PERF-M.SG
'Your courier came!' (i.e. 'Your courier is here!')

(22) *āp k-ā kouriar ā-y-ā.*
HON GEN-M.SG.DIR courier come-PERF-M.SG

The listener on hearing (22), when he was expecting (21), would not be able to understand the context of the sentence fully without reinterpreting it. As a consequence, (22) on its own in this context is incomplete and, as a consequence, the listener, finding the speaker's statement somewhat unexpected, would reinterpret (22) as one of the three possibilities below:

²⁰ Regarding what we mean by the 'aorist' in H/U see note 5 above.

- (22.a) *āp k-ā kouriar ā-y-ā h-ai.*
 HON GEN-M.SG.DIR courier come-PERF-M.SG be.PRES-3SG

As the listener was expecting (21), there was already prior knowledge. However, if the speaker says (22) (*āyā*) instead of (21) (*ā gayā*), the listener might think that perhaps the speaker has forgotten about the courier and, hence is reporting it as new information. So, while the speaker says *āyā*, the listener, thinking that the speaker is presuming it is new information for him, really wanted to say *āyā hai*, but just said *āyā* perhaps because he was in a hurry (or maybe because he misheard him).

- (22.b) *āp k-ā kouriar ā ga-y-ā.*
 HON GEN-M.SG.DIR courier come go.VV-PERF-M.SG

The listener knows that the speaker wanted to say *ā gayā* (22.b) as in (21) (based on the tone of voice and/or because the listener knows that the speaker knows that the listener was expecting the courier), but perhaps he was in a hurry, or in case of children, still learning the language and nuances of compound verb constructions. Thus, the listener would reinterpret it as a case of an intended compound verb construction where the simple verb was used by mistake.

- (22.c) *āp k-ā kouriar ā-y-ā lekin ...*
 HON GEN-M.SG.DIR courier come-PERF-M.SG but ...

The listener while reinterpreting (22) as (22.a) or (22.b) would have to conclude that the speaker used the wrong construction by mistake or because he was in a hurry. For competent speakers, the chances of that happening are usually rare. So, in most scenarios, on hearing (22), the listener would conclude (22.c), that the courier is here but there is something missing or something off about it. The expected action is not fully completed, and he anticipates some bad news. Because if that were not the case, the speaker would have said (21).²¹

²¹ We think that the semantic aspect of perfectivity that sometimes qualifies the use of compound verbs is correlated with the presence of '(inter)subjectivity', that is to say, they are not two entirely independent meanings but normally occur together. This point, however, needs further research.

It is important to emphasize here that this reassuring element of compound verb constructions in cases of (inter)subjectivity plays a very important role in day to day communication.²²

A flip side to this is that compound verb constructions governed by the presence of the notion of ‘(inter)subjectivity’ are used mainly for getting or conveying confirmation about the occurrence of an expected event. For instance, when the speaker says sentence (21), he is also implying that ‘we do not need to worry about the courier anymore, it has arrived, and we can talk about something else or go back to what we were doing.’ If, however, the listener wants more information about the event, the conversation would most likely switch to simple verbs. Here is an example:

(23) a: *āp k-ā kouriar ā ga-y-ā!*
 HON GEN-M.SG.DIR courier come go.VV-PERF-M.SG

b: *acchā! kab ā-y-ā?*
 oh (lit.: ‘good’) when come-PERF-M.SG

a: *dophar mē ā-y-ā.*²³
 (this) afternoon LOC come-PERF-M.SG

a: ‘Your courier is here.’

b: ‘Oh! When did he come?’

a: ‘He came this afternoon.’

Jagannathan (1991) also illustrates this point in his continuation to example (19) above.

(24) a: *maī ne tumh-ē ek kitāb d-ī th-ī.*
 1SG ERG 2PL-DAT a book.F give-PERF.F be.IMP.F

vah kitāb tum ne paṛh l-ī?
 3SG book.F 2PL ERG read take.VV-PERF.F

²² Ubiquitous expressions like *ho jāegā* ‘it will be done’, convey not only information about completion of a certain task or work, but also convey a sense that the listener need not worry.

²³ This conversation has been created by us.

b: *hā̃ paṛh l-ī* *bahut acch-ī h-ai.*
yes read take.VV-PERFF very good-F be.PRES-3SG

a: *kitnī der mē paṛh-ī?*
how much time LOC read-PERFF

b: *cār ghaṅṅ-e mē.*
four hour-M.PL LOC

a: 'I had given you a book. Did you read it?'

b: 'Yes, I read it. It is very good.'

a: 'In how much time did you read it?'

b: 'In four hours.'

(adapted from Jagannathan 1991: 277, English translation by Neha Tiwari)

Notice that when 'a' asks how long it took 'b' to read the book, she uses a simple and not a compound verb. Most native speakers find this to be the most neutral way to ask questions soliciting for details. The use of compound construction is not appropriate in this case because it would carry certain implications about the time it usually takes 'b' to read a book—something about which 'a' may or may not have any idea or opinion.²⁴

Let us look at some examples of '(inter)subjectivity' at work in literature.

(25) *aur abhī ham cand kadam cale hōge ki acānak dhūdh mē se dākbābū*
kā cehrā ubharā

We would have only walked a few steps when Daakbabu's face appeared from the mist'

aur phir dākbābū ne kah-ā-
and then Daakbabu ERG say-PERF.M.SG

²⁴ Nespital (1997) points out that when the informational centre of the utterance is something other than the realization of the verbal action, then simple verbs are more likely to be used. Additionally, it also seems that the use of compound verbs under such cases is often idiomatic. For example, when a speaker says, '*kaun ā gayā?*' instead of '*kaun āyā?*', the question is not just to elicit an informational response, it also expresses some frustration or another similar emotion on the part of the speaker.

vah ciṭṭhī ā ga-ī, jis-kā mujh-e
 3SG.DIR letter.F come go.VV-PERF.F REL.OBL-GEN 1OBL-DAT
īmtazār th-ā.
 waiting be-IMPF.M.SG
 ‘and Daakbabu said—The letter I was waiting for has come.’

maī ne kah-ā – devyānī k-ī mā,
 1SG.DIR ERG say-PERF.M.SG Devyani GEN-F mother
*ḍākbābū ke nām ciṭṭhī ā ga-ī.*²⁵
 Dakbabu GEN name letter.F come go.VV-PERF.F
 ‘I said—Devyani’s mother, the letter in the name of Daakbabu has come.’

In (25) one of the speakers, the Daakbabu, tells the other that the letter he was expecting has arrived, and this person in turn informs his wife, Devyani’s mother, that the letter that was already expected has come. In both cases, the use of compound verb *ā jānā* corresponds with our observation that in situations where an event is expected by both the listener and the speaker, compound verbs are likely to be used.²⁶ But, consider example (26):

(26) *kuch der niṣprāṇ se pare rahe ki*
 for some time lifelessly lain stay-PERF.M.PL that
bāhar se saukhīlāl k-ī āvāj ā-ī,
 outside from Saukhilal GEN-F voice.F come-PERF.F
 ‘After he had lain there lifelessly for some time, he heard Saukhilal’s voice,’

“*kaumreḍ. ... ciṭṭhī ā-ī hai surājo k-ī.*”
 Comrade letter.F come-PERF.F be.PRES-3SG Surajo GEN-F
 ‘«Comrade! Surajo’s letter has come.»’

²⁵ <https://books.google.co.in/books?id=zIVQDwAAQBAJ&pg=PT35&lpg=PT35&dq=%22चिट्ठी+आ+गई%22&source=bl&ots=neOilvNRIs&sig=ACfU3U0GTCtdq1eBJ-beujfmOe5Q52TMzQ&hl=en&sa=X&ved=2ahUKEwiyy9nv7IrpAhUGWX0KHaX6D144ChDoATAAegQIAxAB#v=onepage&q=%22चिट्ठी%20आ%20गई%22&f=false>

²⁶ Also, it is interesting to see in this case that while the Daakbabu is the “speaker” giving information about the arrival of the letter, the use of the compound verb in his phrasing is affirming something that he himself knew and his listener knows as well. This beautifully illustrates our definition of ‘(inter)subjectivity’ in the context of compound verb constructions.

bolte hue ve dālān mē ā gae. guruji ko lagā, jaise sapnā dekh rahe hō. ardhcetnā kī nīd se ve jage aur ackacāe hue saukhīlāl ko dekhne lage. hāth mē rakhā liphāphā saukhīlāl ne guruji kī taraf barhā diyā. liphāphā dete hue bole, “saverē riphūjī gāv ke pās dākpīn ne yah ciṭṭhī dī. surājo kī ciṭṭhī hai. dekhie na, ciṭṭhī kī pīṭh par nām-patā likhā huā hai.”²⁷

‘Saying this he came to the verandah. Guruji felt as if he was dreaming. He woke up from his half-conscious sleep and, startled, started looking at Saukhilal. Saukhilal moved the envelope that was in his hand towards Guruji. While giving the envelope he said, «The postmaster gave this letter near the refugee village this morning. It is Surojo’s letter. Look, on the back of the letter (his) name and address are written.»’

Contrary to example (25), in (26) the listener, Guruji, is not expecting Surajo’s letter. When Saukhilal reports the arrival of the letter, he uses the simple verb in the present perfect (*āī hai*). Guruji is finding it hard to believe that a letter would have come from Surajo, so Saukhilal tells him how he got the letter and why he thinks it is Surajo’s (because his name and address are mentioned on the back of the letter).

Here is one example from the news:

- (27) *anya koronovāyras kī tarah yah vāyras (kovid-19) bhī*
 other coronavirus like 3SG.DIR virus.M (kovid-19) too
*jānvarō. se ā-y-ā h-ai.*²⁸
 animal.M-PL.OBL from come-PERF-M.SG be.PRES-3SG
 ‘Like other coronaviruses, this virus (COVID-19) too has also come from animals.’

In (27) the writer is trying to explain the origin of COVID-19. While people may generally be aware that the coronavirus family comes from animals, he is establishing that to be true for COVID-19 as well as he

²⁷ https://books.google.co.in/books?id=wReKDwAAQBAJ&pg=PT33&lpg=PT33&dq=%22चिट्ठी+आई+है%22&source=bl&ots=oynujNW9IK&sig=ACfU3U3AiI-vlBxP-Ggp6nQU6pw_aFuySw&hl=en&sa=X&ved=2ahUKEwjS9Pag7orpAhWaA3IKHTR_D1E4MhDoATAAegQICxAB#v=onepage&q=%22चिट्ठी%20आई%20है%22&f=false

²⁸ <https://www.downtoearth.org.in/hindistory/health/communicable-disease/what-is-this-coronavirus-how-dangerous-is-this-69599>

probably does not want to presume that the readers would already be aware of this information (even though they might be).²⁹ Hence the use of the simple verb *āyā hai*.

Sentence (28) below is taken from a local newspaper published on April 5, 2020, after the COVID pandemic had started spreading in India, and people were already anticipating that it would reach their towns/cities. In this example, the use of *ā gayā* confirms what the readers were already suspecting.

- (28) *alarṭ- koronāvāyras ā ga-y-ā h-ai*
alert coronavirus come go.VV-PERF-M.SG be.PRES-3SG
*rāybareli (mē) bhī.*³⁰
Raebareli (LOC) too
'Alert: The coronavirus has come in Raebareli too.'

To sum up, in this section we looked at how the phenomenon of '(inter)subjectivity' is expressed in H/U using compound verb constructions with *jānā*, *denā* and *lenā*. However, a question arises, do all occurrences of these compound constructions indicate the presence of (inter)subjectivity? We discuss this in the next section.

3.1. (Inter)subjectivity: The case of *girnā* 'to fall'

As it has been illustrated above through our analysis, our main hypothesis is that if both the speaker and the listener share a certain informational context with respect to the action in question, this shared information is marked by the presence of a compound verb. This raises an important question: are all usages of *jānā*, *denā* and *lenā* as vector verbs in H/U cases of (inter)subjectivity according to our definition given in section 3 above? In other words, if a speaker chooses to use a compound verb over a simple verb construction using these vector verbs, does that always imply presence of prior or shared information?³¹

²⁹ This is another interesting aspect about the use of compound verbs: there might be situations where the speaker may not acknowledge that some shared information exists in the context, and instead, he may prefer to use the simple verb. It is a complex speech act that deserves further careful analysis.

³⁰ <https://arjanadhikarmedia.page/article/alarṭ-korona-vaayaras-aa-gaya-hai-raayabarelee-bhee/Jt2Wza.html>

³¹ Hook (*in press*) has also discussed this in his paper.

According to Jagannathan (1981), compound verb constructions with *jānā*, *denā* and *lenā* are based on prior knowledge, but he concedes that it is not always possible to establish this for every such usage one encounters. He gives the example of *baiṭh jāo* instead of *baiṭho* and says that whether or not there is prior knowledge when this sentence is used is debatable. However, he goes on to argue that in most of the cases, one can find the context that is creating a prior knowledge/(inter)subjectivity like scenario (Jagannathan 1981: 265). He also mentions that there are certain simple verbs for which the compound form is usually preferred in the perfective and sometime in other tenses as well.³²

In this section, we will look at the case of one such verb, *girnā*, ‘to fall.’ This verb is frequently used along with the vector verb *jānā*, and as in the case of *baiṭh jāo* and *baiṭho* above, the argument for the applicability of (inter)subjectivity is not always obvious and/or perhaps even tenable. While it is not possible at this stage to lay down exact rules for when *girnā* can or should be used with vector verbs, it would be instructive to look at some of the usages as they can help bring out some of the main considerations native speakers have in mind while weighing the choice between using a simple verb versus a compound verb. Consider example in (29):

- (29) *ādhī se gir ga-y-ā 215 sāl*
 heavy wind because of fall go.VV-PERF-M.SG 215 year.PL
purān-ā per jar-ō ko dekhte hī dar
 old-M.SG tree.M root-M.PL.OBL ACC looking after scare
*gae log aur bulā lī pulis.*³³
 go.VV-PERF.M.PL people.M and call take.VV-PERF.F pulis.F
 ‘A 215 year old tree fell because of heavy wind, the people got
 scared after looking at its roots and called the police.’

³² Jagannathan (1981) also talks about *caīknā* (‘to be startled’), *jhuknā* (‘to bend’), *ulajhnā* (‘to be tangled’), *phisalnā* (‘to slip’), *phailnā* (‘to spread’), *bahnā* (‘to flow’, ‘to drift’), *bujhnā* (‘to be extinguished’), *ṭakrānā* (‘to collide’) which are used with *jānā* (‘to go’) as the vector verb but for which prior knowledge analysis does not always apply. These verbs are beyond the scope of the current paper. However, we think that our analysis for *girnā* (‘to fall’) can potentially be generalized to some of these other verbs. We hope to address this fully in another paper.

³³ <https://m.dailyhunt.in/news/india/hindi/my+khabar+hindi-epaper-mykrhin/aandhi+se+gir+gaya+215+sāl+purana+ped+jado+ko+dekhte+hi+dar+gae+log+aur+bula+li+pulis-newsid-96480643>

As this is a news headline, it is new information for the readers (which are akin to listeners in our analysis). Hence, there is no prior knowledge and yet *girnā* has been used in the compound form *gir jānā*; this seems to be true also for example in (30):

- (30) *us ne dhīre se samīr k-e kādhe par*
3SG.OBL ERG softly Sameer GEN-M.OBL shoulder LOC
hāth rakh-ā aur samīr k-e hāth se
hand place-PERF.M.SG and Sameer GEN-M.OBL hand from
*kāfī k-ā kap gir ga-y-ā.*³⁴
coffee GEN-M.SG.DIR cup.M fall go.VV-PERF-M.SG
'She softly placed (her) hand on Sameer's shoulder and the coffee cup fell from Samir's hand.'

In this case as well, neither the subject nor the readers knew beforehand that the first action of placing the hand on the shoulder will result in the second action, the cup falling.

So, while in (29) and (30) above, the events (the tree and the cup falling) were new and unexpected, the use of *gir jānā* in these scenarios sounds natural to native speakers. We can explain the use of compound verbs in these situations in three possible ways:

- i. Certain polar and vector combinations, especially with *jānā* 'to go' as the vector verb, have become fossilized, and it is not always possible to correlate their usage with the existence of prior knowledge in the context;
- ii. As we have noticed in section 1 above, the use of compound verbs emphasizes perfectivity or completion. When something falls, it falls all the way. Hence, the semantic aspect of perfectivity, and not the pragmatic aspect of (inter)subjectivity, could be driving the choice of the compound construction here;³⁵
- iii. Also, when we look at (30) from the point of view of the readers, as the context gets established in H/U, 'from Samir's hand the coffee cup', the readers start to expect that the cup is going to fall. Thus, while the readers did not have any information about the cup falling when the sentence

³⁴ From *Tomaaya Gaan Shonaabo* by Manav Kaul (<http://aranyamanav.blogspot.com/2010/06/blog-post.html>).

³⁵ Nespital (1997: *sub voce*) gives several examples of *gir jānā* that bring out the perfective aspect of the compound verb construction in contrast to the usage of *girnā* which is often seen in the non-perfective.

started, as the context was established, the expectation was created midway through the sentence. So, the use of the compound form in this case can also be to meet the expectation of the readers, and it gives them a sense of finality and completion. This point needs further explanation, and we will discuss it below by taking an example from Bashir (1988).

In a footnote in her study on Kalasha syntax, Bashir (1988) discusses what we advanced in point (iii) above in relation to the use of compound verbs in H/U with the below example:

- (31) *sunītā ke pāv anāyās³⁶ hī darvāz-e kī or*
Sunita GEN feet suddenly PTC door-OBL toward
barh ga-e.
move go.VV-PERF.M.PL
'Sunita's feet suddenly moved toward the door'.

The example was originally given by Hook³⁷ to show that the vector verb *jānā* can also express the semantics of 'unprepared mind'. However, regarding the example in (31) Bashir argues that:

[...] they are in my view not counterexamples at all. In (a) (i.e. our 31), for example, the previous context of the story has been leading up to the very act of going out the door. It is thus reporting an event which has been (carefully) foreshadowed for the character and for which the hearer/reader has been mentally prepared.

(Bashir 1988: 258, note 30)

As a consequence, according to Bashir, the readers who have read the build-up to (31) already, to an extent, anticipate that Sunita would go towards the door and her doing so was expected for them. Thus, while experiencing/reading a build-up to a particular event, one can attain a state of mind where one starts to expect it. This is something Slobin / Aksu

³⁶ *anāyās* can also be translated as 'spontaneously' in this context.

³⁷ Hook (*in press*) in a recent article talks about the pragmatic factor of 'prepared mind' (what we are calling '(inter)subjectivity' and 'prior knowledge' in this paper) and concludes that while it is operative in some cases when compound verbs are used in H/U, it is not enough to account for all the places where compound verbs are seen. This is something we agree with too, especially because in many instances it is not always possible to clearly delineate the semantic factor of perfectivity and the pragmatic factor of 'prepared mind'—they often work in tandem.

(1982) also drew attention to when talking about the Turkish particle *-miş* with reference to the concept of ‘prepared mind’:

Kemal gelmiş ‘Kemal came’ is appropriate in the context of encountering Kemal’s coat, but not in the context of hearing the approach of Kemal’s car. In both cases, the speaker has not SEEN Kemal or his arrival, but in the latter case the auditory sensory experience is part of the process of Kemal’s arrival, and thus the speaker’s consciousness was involved in the process before its actualization.

(Slobin / Aksu 1982: 192)

Thus, the examples from Bashir (1988) and Slobin / Aksu (1982) support what we said in point (iii) above, and also help to illustrate something more general about the phenomenon of (inter)subjectivity. The knowledge that is shared between people may not always be verbalized, it could be something that is inferred from circumstances or from one’s own experience, and it may not have any verbal component to it. This happens in many cases when people live together, and as a result, they come to acquire a fair amount of shared information regarding their context not all of which is necessarily verbalized (see sentence 17 above).

However, while we have said that *girnā* normally occurs with other vector verbs like *jānā* (and *parnā*), it is interesting to point out that it is also possible to find it by itself without any vector verb. This is normally the case when an accident is reported as in (32) below.³⁸

- (32) *tilak nagar-cēbūr steśan ke bīc gir-ā*
Tilak Nagar-Chembur station between fall-PERF.M.SG
*pul k-ā hissā rel sevā bādhit.*³⁹
pull GEN-M.SG part.M train service interrupted
‘A part of the pull between Tilak Nagar and Chembur fell, train service interrupted.’

³⁸ It must be noted that in most of the cases where *girnā* comes on its own without the vector verb, we find that the subject does not come in the beginning of the sentence but rather comes after the verb. While the same reversal of word order can be seen in (29) above as well in the case of *gir jānā*, it is more frequently encountered when *girnā* is used as a simple verb. We are, however, aware of the fact that in order to be sure of this last statement a statistical counting is necessary. This is beyond the scope of the present paper. Moreover, also the fact that this unusual syntax seems to be correlated to expressing surprise or reporting new events needs further research.

³⁹ <https://aajtak.intoday.in/story/weather-update-mumbai-tilak-nagar-chembur-railway-station-harbour-line-trains-bridge-rain-1-1107541.html>

This finding is reinforced by the fact that Hook (1974, *in press*) gives certain examples of situations where prior shared knowledge is missing. One general category is news headlines which report new and usually unexpected information.

- (33) *kal hindustān mē bhūkamp ā-y-ā.* (and not *ā gayā*)
 yesterday India LOC earthquake come-PERF-M.SG
 ‘There was an earthquake in India yesterday.’ (adapted from Hook 1974: 250)

Analyzing this sentence, Hook says that in situations of *prior ignorance* like earthquakes, H/U speakers prefer to talk about the event in question using a simple verb.

To conclude this section, we advance the hypothesis that certain polar and vector verb combinations such as *gir jānā* have become so grammaticalized – and thus more and more obligatory – in H/U that as a result of the semantic-pragmatic bleaching typical of all grammaticalization process (cf. Hopper 1991; Hopper / Traugott 2003; Traugott / Heine 1991) it is not always possible to see the pragmatic factor of (inter)subjectivity coming into play for such usages (see point ‘i.’ above). Additionally, the fact that it is also possible to see the usage of *girnā* on its own, which although not very common, is still sometimes seen in H/U (cf. Hook *in press*), is again illustrative of the gradualness of the grammaticalization process where different stages of the grammaticalization cline coexist in the same synchronic phase of a particular language (cf. Hopper 1991; Lehmann 1982). However, even with these particular combinations, it is possible to suggest that the pragmatic factor of (inter)subjectivity can lead to the manifestation of compound verbs. Indeed, as we saw in the discussion of (30) and (31) above, sometimes it may be possible for fore-knowledge about an event to develop even mid-way through the sentence or just prior to the sentence under consideration but not verbally acknowledged (cf. Bashir 1988). This point needs further research.

4. *Mirativity and (Inter)subjectivity*

As we have seen so far, the use of *jānā*, *denā* and *lenā* as vector verbs may not only have a semantic but rather a pragmatic meaning that is codified in the information structure of the event. More specifically,

compound verb constructions with *jānā*, *denā* and *lenā* as vector verbs (restricted as per our definition in section 3) are more likely to be used in situations where there is prior shared knowledge about the action or an expectation that it will take place. This brings us to the question of how the use of compound verbs can be correlated to the expression of surprise in H/U, and thus, if the use of simple verbs in H/U is in some way connected with mirativity. But before we go into the details of this topic, some introductory words are in order.

Mirativity⁴⁰ is now a widely recognized linguistic category and research is underway in many languages to identify how different grammatical constructions are used by speakers of different languages to express surprise. DeLancey (2012) defines mirativity as a category which marks a statement as representing information which is *new* or *unexpected*. He (1997, 2001, 2012) has given several examples of mirative constructions in different languages. Here is an example from the Athabaskan language Hare in which the particle ‘*lō̃*’ is used to express surprise and is therefore is classified as a marker of mirativity:

- (34) *juh̄ye sa k'ínayeda lō̃*
hereabout bear walk.around.IMPF MIR
‘There was a bear walking around here!’ (inferred from discovery
of bear tracks)
(adapted from DeLancey 2001: 375)

Talking about the use of *lō̃* in this sentence DeLancey says:

When I discussed the question with my consultants, they were quite clear about what the something else was. As one speaker put it, “*lō̃* is there because you didn’t know”.

(DeLancey 2001: 377)

Mirativity is reputed a valid linguistic category in some South-Asian languages also. For example, talking about the mirative in the Dardic language Khowar, Bashir (2010) gives the following example:

⁴⁰ On the correlation between ‘mirativity’ and ‘(inter)subjectivity’ see Nuyts (2001).

- (35) *tu khowár kor-ák bir-áu*
you Khowar do-AG.N become(PST.I)-2SG
'(Oh), you speak Khowar.'
(registering surprise)

Bashir explains (35) saying:

The sentence shown in (35) is a surprised reaction to an immediate, first-hand observation of an unexpected ongoing action, and clearly not reporting hearsay, inference from a resultant state, or non-witnessed action. Nor can its primary purpose be to inform the addressee (me, in this case) of a fact, since I can be presumed to already know that I speak Khowar. Its function is to communicate the fact that the speaker has learned something new and is surprised by it. The mirative meaning of this utterance is clear and distinct, regardless of the fact that the Khowar “inferential form” is also used for other indirective/mediative meanings.

(Bashir 2010: 3)

Starting with this short introduction to mirativity in different languages, in this section our focus is primarily on the possible grammatical expression of mirativity in H/U. We build on the general work on this field by Montaut's (2006) analysis of mirative usage of the simple perfect (what she calls the aorist) in H/U (see also Sigorskiy 2010). However, the purpose here is to extend her analysis further by exploring whether or not there are any mirative implications to cases where compound verbs are used in H/U.

In her analysis, Montaut (2006) looks at the use of the simple perfect or aorist to express unexpected events. Let us look at some of the examples from her paper:

- (36) *are! kitnā barā ho ga-y-ā!*
my how tall.M.SG be go.VV-PERF-M.SG
'My! How tall he has become!' (adapted from Montaut 2006: 75)
- (37) *ā^{kh}ir ā gae mehmān!*
finally come go.VV-PERF.M.PL guest.PL
'Here they are finally, our guests.' (adapted from Montaut 2006: 75)

- (38) *are ! kaise ā-e?*
my howc ome-PERF.M.PL
'My! How come you are (here)? / what's up?/ what a surprise!'
(adapted from Montaut 2006: 76)

Montaut's analysis of how the element of surprise is expressed in these sentences is restricted to the use of the aorist. She says regarding (36):

The speaker confronted with an unexpected fact or situation (here the size of the boy) uses the aorist and not the perfect or present, which would only mean a neutral statement.

(Montaut 2006: 75)

For (36) and (37), Montaut says:

The above examples [...] are classical 'mirative' extensions of evidential markers in the world language with evidential markers, and can be explained by a contrast between what is expected P' and what occurs P...

(Montaut 2006: 75)

Thus, in Montaut's analysis, the expression of speaker's surprise is conveyed by the use of the aorist, and thus the latter, according to her, is in H/U a marker of the linguistic category of 'mirativity'. She does not mention, in any way, how the expression of speaker's surprise is related to the presence or absence of compound verbs. However, as we have discussed above, compound verbs are often used in the context of (inter)subjectivity and especially so in the simple perfect/aorist. So, the question naturally arises, how do we reconcile the two scenarios? To be more explicit, according to Montaut (36) and (37) above are good examples of the expression of speaker's surprise because the verb is in the aorist, while following our analysis of (inter)subjectivity, one would conclude that since compound verbs are used in (36) and (37), there is presence of prior shared knowledge (so, on the surface, it appears to be the opposite of surprise). Although this seems a real contradiction, as we will show, it is a *prima facie* contradiction. To understand this, we propose here a more nuanced analysis of the notion of mirativity that could account for sentences (36) and (37), that is to say, we propose an analysis that is

not in contrast with Montaut's analysis and can still account for (inter)subjectivity.

According to our analysis, the same examples, (36) and (37) above, are used in a context of 'partial surprise', and because of the fact that the surprise is only 'partial', compound verbs are being used. In order to explain this new aspect, which we have called 'partial surprise', it would be useful to delineate two aspects of the event in (36) above:

- i. The subject of the verb 'to grow' has grown in size; this is encoded in the use of *barā ho gayā*;
- ii. The subject has grown a lot; this is expressed by the use of *kitnā* which means 'how much', normally an interrogative marker, but it is being used to express surprise here.

The use of the compound verb in (31) suggests that the speaker was already anticipating that the subject would have grown. However, what is surprising about the event, and thus the use of the simple perfect (or the aorist) according to Montaut, is that the speaker didn't anticipate the extent of growth (hence the use of *kitnā*). Thus, when we speak of surprise, it may not and should not be restricted to speaker's reaction to the whole event. As a matter of fact, the speaker can also choose to express his/her surprise about some unexpected characteristic of an expected event.

The same analysis explains the usage in (37). The speaker was already anticipating the arrival of the subjects (that they were supposed to come was known information). However, what was unexpected was the time when they came (when everyone had given up hope that they would come).⁴¹

Given the analysis above, sentence (38) presents a good contrast to (36) and (37). *āe* is a simple verb and in this case, the use of the simple perfect/aorist form of a simple verb (i.e. *āe*) indicates that the arrival of the subject was entirely, that is to say not partially, unexpected.

Also, notice the use of the present perfect in sentence (20) in section 3 repeated here for convenience (now 39):

⁴¹ In this particular sentence it is also important to note the change in word order. Putting the subject at the end of the sentence in our view also lends an element of surprise. However, more work needs to be done to understand the correlation between word order and the expression of mirativity in H/U.

- (39) *sāhab, āpse koī mil-n-e*
 Sir HON with someone meet-INF-OBL
ā-y-ā h-ai.
 come-PERF-M.SG be-PRES.3SG
 ‘Sir, someone has come to meet you.’

By our analysis of (inter)subjectivity, we know that in (39) the speaker did not use the compound form *ā gayā* as there is no shared information between the speaker and the listener. But, although, the information conveyed is new, the goal of the sentence is not to express surprise but only to report new information in a neutral tone. Thus, Montaut’s observation that the aorist is used in cases where the speaker wants to convey a sense of surprise holds true for this particular example. Thus, while Montaut’s analysis of the use of the aorist for encoding mirativity in H/U holds, we contend that the analysis can be extended further, trying thus to analyze more in depth the possible expression of the different nuances of mirativity in H/U—a difference that is conveyed by the use of simple versus compound verbs. So, when compound verbs get used in statements of surprise, it is possible that there is a simultaneous expression of overall knowledge about the event but surprise regarding a particular aspect of it. We will illustrate this point with two more examples taken from literature.

- (40) ... *patā nahī kyā hogā merā, budbudāte hue vo fon lagāte hue socne lagī, patā nahī ārav uṭhā ki nahī, pāgal hai bilkul, socte hue fon ārav ko lagā, usne ek sās mẽ bolnā śurū kiyā guḍ mārñig, uṭh jāo jaldī aur taiyār hokar jaldī ā jāo aur fon rakhte hue apne sapne samette hue vo taiyār hone cal parī. āj parīām part āne vālā thā, thoī ghabrāhaṭ to ho rahī thī par use pakkā viśvās thā ki vo apne mā bāp ke sapnō ko, unse kiye vāde ko zarūr pūrā karegī,*
tabhī us kī nazar ārav par parī,
 just then 3SG.OBL GEN-F eyes Arav LOC fall-PERF.F
are, tum ā ga-e,
 oh, 2PL(SG) come go.VV-PERF.M.PL
calo jaldī noṭis bord par rijalṭ dekhne...⁴²
 come-IMPV hurry notice board LOC result see-INF.OBL

‘Don’t know what will happen to me, muttering as she was calling Aarav, she started to think, don’t know if Aarav has woken up or not, he is a complete idiot, as she was thinking this, the phone call to Aarav

⁴² An excerpt from <https://storymirror.com/read/story/hindi/64zqf2jk/alphaaz/detail?undefined>

got connected, and she started saying in one breath, good morning, wake up fast, get ready and come here quickly, and as she was about to hang up, she gathered her things and set off to get ready. The results were coming out today, she was a bit nervous, but she was confident that she will definitely fulfil her parents' dreams and the promises she made to them. Just then, her eyes fell on Aarav. **Oh, you are here!** Come on, hurry, let's go to the notice board to check the result...'

As we can see from the context, the narrator was expecting her friend Aarav to come. However, when he got there, she was lost in her own stream of thoughts. So, she is surprised to see him not because she didn't know he was coming—she was already expecting him, but because she had for a moment stepped out of 'ordinary conscious awareness' and was not in the present moment. Thus, the aorist here along with the explicative *arey* serve to add the sense of surprise, while the use of *ā gae* confirms that it was after all expected. Here is one more example:

- (41) “*arey, yah kitab to pramilā bhūl hī gaī. śām ko uske ghar de ānā. aur jiske kaho – rekhā ke, nūtan ke, citravatī ke-kuch dene ke bahāne tumhen vahīm bhej dūngī.*”

kahte-kahte acānak jāne kyā huā, rūp maṇi kī āmkhem ḍabḍabā āī. bhāvuk hokār bolī “maī ādmī kī taklīf ko bakhūbī samajh saktī hū. paṛhātī bhī hū. yahī merā subjekt hai. sāl bhar se ūpar ho gayā hai. tum tapasyā kar rahe ho. cāhū bhī to...ḍākṭar ne haḍḍiyō kī taklīf ke kāraṇ thorā hilne-ḍulne se bhī manā kar rakhā hai.” rūp maṇi viclit ho calī thī, “maī tumhāre kisī kām kī nahī rahī nā.” uskī laṃbī aṃguliyā apnī ākhō tak calī gaī.

“*are-are yah kyā ho ga-y-ā*

tumhē!

Hey what be go.VV-PERF-M.SG 2PL(SG)-DAT

pāgal h-u-ī h-o? kyā ādmī sirf isīlie

mad be-PP-F be-PRES.2PL INT man only for this reason

śādī kar-t-ā h-ai?...”⁴³

marriage do-PRES-M.SG be-PRES.3.SG

“*Hey, what has come over you! Have you gone mad? Does a man get married only for this reason?...”*

⁴³ From *Teesri Kahani* by Hardarshan Sehgal. <https://books.google.it/books?id=Z050BQAAQBAJ&pg=PT3&dq=Hardarshan+Sehgal&hl=it&sa=X&ved=2ahUKEwiH0YnHodLrAhXjkIsKHSVaC2EQ6wEwAnoECAQQAQ#v=onepage&q=Hardarshan%20Sehgal&f=false>

“Oh, Pramila forgot this book. Go give it to her in the evening. And whoever you say—Rekha, Nutan, Chitravati—I will send you to their house on the pretext of delivering something.

Saying this out of nowhere Roop Mani’s eyes suddenly became full of tears. Getting emotional she said, “I understand the pain/discomfort of a man really well. I also teach it. This is my subject. It’s been over a year. You are doing a penance. Even if I wanted to...the doctor has asked me not to move even a bit because of the problem in my bones. Roop Mani had become agitated, “I am of no use to you, right?” Her long fingers went up to her eyes.

“**Hey, what has come over you!** Have you gone mad? Does a man get married only for this reason?...”

The context of these lines is that Roop Mani and the person she is talking to are married. But Roop Mani is confined to the bed and cannot be intimate with him. She is joking about setting him up with some other women and at the same time getting agitated. Her husband is surprised to see the sudden shifts in her mood, first joking and then crying. In this example, the case for (inter)subjectivity is not as straightforward. When the husband said, “*ye kyā ho gayā tum-hē*” he is expressing his surprise at the things Roop Mani is saying and by her mood. However, when we consider the context it is not hard to conclude that Roop Mani’s behavior and how she is feeling cannot be new for her husband in its entirety as she has been confined to the bed for over a year, and they have been together during this period. Again, what is surprising to him is not her mood per se, but the extent to which she is feeling worthless and blaming her own self in this particular context. Thus, we can see that in H/U, both surprise and (inter)subjectivity can co-exist in the same sentence and that the analysis of the codification of mirativity, especially in the case of Indo-Aryan languages, should take into account the entire informational context of the sentence and then identify aspects that are known, familiar or expected and aspects that are surprising and/or unexpected. One way this can be done in H/U is by looking at the form the verb takes in the sentence.

5. Conclusion

There is a wealth of psychological, contextual, and cultural detail hidden behind the usage of compound verb constructions in H/U, and hence, it is not surprising that they continue to present one of the greatest hurdles to attaining

oral proficiency for students learning either Hindi or Urdu as a second language. Among all the different verbs that can be classified as true vector verbs, Jagannathan (1981) thinks that there is a case to be made for treating compound verb constructions with vector verbs *jānā*, *denā* and *lenā* differently when it comes to '(inter)subjectivity', and as we have (exhaustively, we hope) showed, we agree with his observation. Indeed, in the present paper we have focused on the pragmatic meaning of compound verb constructions with *jānā*, *denā* and *lenā* as vector verbs and have shown that these constructions are used in cases of prior shared knowledge—when the speaker knows that the listener has foreknowledge of the event and/or is expecting it. As we have seen, this shared information can exist because of explicit past actions (as in sentence 19) or because of just being familiar with the context (like in sentence 17). We then looked at Montaut's study (2006) on the codification of surprise in H/U and tried to correlate her analysis to our analysis related to the use of compound verbs in the presence of prior shared knowledge. In her paper, Montaut provides several examples where statements of surprise have compound and not simple verbs in them, and so at first blush, it appears that the (inter)subjectivity framework might not be applicable to sentences that express surprise. However, we saw that it is important to understand what constitutes 'surprise' in any particular context. In H/U, it is possible to simultaneously express familiarity with the overall context and surprise about a particular aspect of it, and we looked at some of Montaut's examples to show how mirativity and inter-subjectivity can co-exist in certain scenarios and that the notion of 'surprise' itself deserves a more nuanced analysis. This means that mirativity does not exclude existence of prior knowledge, rather the presence of prior knowledge which is codified with compound verb constructions makes the expression of surprise in H/U more nuanced. While we have made an attempt here to analyse in depth only one of the pragmatic aspects of compound verb usage, we hope to continue to build this analysis further so that our understanding of this construction can gradually become more complete.

Andrea Drocco
Università Ca' Foscari di Venezia
Dipartimento di Studi sull'Asia
e sull'Africa Mediterranea
Ca' Cappello, San Polo 2035 - 30125 - Venezia
andrea.drocco@unive.it

Neha Tiwari
Independent scholar
tiwari.neha@icloud.com

References

- Aksu-Koç, Ayhan / Slobin, Dan, 1986, “A Psychological Account of the Development and Use of Evidentials in Turkish”. In Chafe, Wallace and Nichols, Johanna (eds.), *Evidentiality: the Linguistic Coding of Epistemology*, Norwood, New Jersey: Ablex, 159–67.
- Annamalai, Elay, 2016, “Expanded verbs in Dravidian”. In Hock, Hans H. / Bashir, Elena (eds.), *The Languages and Linguistics of South Asia: A Comprehensive Guide*, Amsterdam: John Benjamins, 550-9.
- Bashir, Elena, 1988, *Topics in Kalasha syntax: An areal and typological perspective*, University of Michigan, Ph.D. dissertation unpublished.
- Bashir, Elena, 2010, “Traces of mirativity in Shina”. *Himalayan Linguistics* 9/2: 1–55.
- Butt, Miriam / Gillian Ramchand, 2001, “Complex aspectual structure in Hindi/Urdu”. *Oxford Working Papers in Linguistics, Philology and Phonetics* 6: 1–30.
- Caracchi, Pinuccia 2002. *Grammatica Hindī*, Torino, Magnanelli. (4th edition).
- DeLancey, Scott, 1997, “Mirativity: The grammatical marking of unexpected information”. *Linguistic Typology* 1: 33–52.
- DeLancey, Scott, 2001, “The mirative and evidentiality”. *Journal of Pragmatics*. 33/3: 369–382.
- DeLancey, Scott, 2012, “Still mirative after all these years”. *Linguistic Typology* 16/3: 529–564.
- Drocco, Andrea, 2018, “An attempt to understand the encoding of reduced transitivity in Hindi: the case of compound verbs with *jānā*”. In Sharma, Ghanshyam / Bhatt, Rajesh (eds.), *Trends in Hindi Linguistics*, Berlin/Boston: Walter de Gruyter, 265–84.
- Drocco, Andrea / Tiwari, Neha, 2020, “Atypical Compound Verb Constructions in Hindi/Urdu. The Case of Transitive Polar Verbs with Intransitive Vector Verbs”. *Annali di Ca' Foscari. Serie orientale*, 56: 307-34.
- Hacker, Paul 1961, “On the problem of a method for treating the compound and conjunct verbs in Hindi”. *Bulletin of the School of Oriental and African Studies* 24, 484–516.
- Hock, Hans H., 2016, “Complex verbs. Introduction”. In Hock, Hans Henrich / Bashir, Elena (eds.), *The Languages and Linguistics of South Asia: A Comprehensive Guide*, Amsterdam: John Benjamins, 549–50.

- Hook, Peter E., 1974, *The Compound Verb in Hindi*, The University of Michigan: Center for South and Southeast Asian Studies.
- Hook, Peter E., 1978, "The Hindi compound verb: What it is and what it does". In Singh, Kripa Shanker (ed.), *Readings in Hindi-Urdu linguistics*, Delhi: National Publishing House, 130–157.
- Hook, Peter E., 1988, "Paradigmaticization: A case study from South Asia". In *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, Berkeley: Berkeley Linguistics Society, 293–303.
- Hook, Peter E., 1991, "The Emergence of Perfective Aspect in Indo-Aryan Languages". In Traugott, Elizabeth C. / Heine, Bernd (eds.), *Approaches to Grammaticalization*, Vols. I-II, Amsterdam, John Benjamins, 59–89.
- Hook, Peter E., 1993, "Aspectogenesis and the compound verb in Indo-Aryan". In Verma, Manindra (ed.), *Complex predicates in South Asian languages*, New Delhi, Manohar, 97–114.
- Hook, Peter E., 2001, "Where do Compound Verbs Come From? (And Where are They Going?)". In Singh, Rajendra (ed.), *The Yearbook of South Asian Languages and Linguistics, Tokyo Symposium on South Asian Languages. Contact, Convergence and Typology*, New Delhi / Thousand Oaks / London, Sage Publications, 101–31.
- Hook, Peter E., *in press*, "Births, earthquakes, meteors, and other autogenous expressions: The Hindi-Urdu compound verb and its covert semantics". In Kageyama, Taro / Hook, Peter E. / Pardeshi, Prashant (eds.) *Verb-Verb Complexes in Asian Languages*, Oxford: Oxford University Press, n.d.
- Hopper, Paul J., 1991, "On some principles of grammaticalization". In Traugott, Elizabeth C. / Heine, Bernd (eds.), *Approaches to Grammaticalization, Vol. I*, Amsterdam: John Benjamins, 17–36.
- Hopper, Paul J. / Traugott, Elizabeth, 2003, *Grammaticalization*, Cambridge, Cambridge University Press.
- Jagannathan, Veeravalli R., 1981, *Prayog aur prayog: Handbook of modern Hindi usage*, New Delhi, Oxford University Press.
- Jagannathan, Veeravalli R., 1991, *Svayam Hindi Seekhein*, New Delhi, Janepa Publishing House Pvt. Ltd.
- Kachru, Yamuna, 1979, "Pragmatics and Verb Serialization in Hindi-Urdu". *Studies in the Linguistic Sciences* 9/2: 157–69.

- Kachru, Yamuna, 1981, "Transitivity and Volitionality in Hindi-Urdu". *Studies in the Linguistic Sciences* 11/2: 181–93.
- Kachru, Yamuna / Pandharipande, Rajeshwari, 1980, "Toward A Typology of Compound Verbs in South Asian Languages". *Studies in the Linguistic Sciences* 10/1: 113–24.
- Lehmann, Christian, 1982, *Thoughts on Grammaticalization. A programmatic Sketch*. Vol. I, Arbeiten des Kölner Universalien-Projekts, Nr. 48. Köln: Institut für Sprachwissenschaft.
- Masica, Colin P., 1976, *Defining a Linguistic Area. South Asia*, Chicago, University of Chicago Press.
- Masica, Colin P., 1991, *The Indo-Aryan Languages*, Cambridge, Cambridge University Press.
- McGregor, Ronald S., 1977, *Outline of Hindi Grammar; with exercises*, Delhi, Oxford University Press (2nd edition).
- Montaut, Annie, 1991, *Aspects, voix et diathèses en hindi moderne*, Louvain-Paris, Éditions Peeters.
- Montaut, Annie, 2004, *A Grammar of Hindi*, München: Lincom Europa.
- Montaut, Annie, 2006, "Mirative Extensions of Aorist in Hindi/Urdu". In Singh, Rajendra (ed.) *The Yearbook of South Asian Languages and Linguistics 2006*, Berlin/New York: Mouton de Gruyter, 71–86.
- Nespital, Helmuth, 1997, *Dictionary of Hindi Verbs*, Allahabad, Lokbharti Prakashan.
- Nuyts, Jan, 2001, *Epistemic modality, language, and conceptualization: a cognitive-pragmatic perspective*, Amsterdam, John Benjamins.
- Nuyts, Jan, 2012, "Notions of (inter)subjectivity". *English Text Construction* 5/1: 53–76.
- Poornima, Shakthi, 2012, *Hindi aspectual complex predicates at the syntax-semantics interface*, State University of New York, Buffalo, PhD dissertation.
- Pořízka, Vincent, 1967–69, "On the Perfective Verbal Aspect in Hindi". *Archiv orientální*. 35: 64–88, 208–31; 36: 233–51; 37: 19–47, 345–64.
- Sigorskiy, Alexander A., 2010, "Evidentiality, inferentiality and mirativity in the Modern Hindi". *Lingua Posnaniensis* 52/1: 71–80.
- Slade, Benjamin, 2013, "The diachrony of light and auxiliary verbs in Indo-Aryan". *Diachronica* 30/4: 531–78.

- Slade, Benjamin, 2016, “Compound verbs in Indo-Aryan”. In Hock, Hans H. / Bashir, Elena (eds.), *The Languages and Linguistics of South Asia: A Comprehensive Guide*, Amsterdam, John Benjamins, 557–65.
- Slobin, Dan / Aksu, Ayhan, 1982, “Tense, aspect and modality in the use of the Turkish evidential”. In Hopper, Paul (ed.), *Tense-aspect: Between semantics & pragmatics*. Amsterdam: John Benjamins, 185–200.
- Snell, Rupert / Weightman, Simon, 2010 (1989), *Teach Yourself Complete Hindi*, London, Hodder Education.
- Traugott, Elizabeth C. / Heine, Bernd (eds.), 1991, *Approaches to grammaticalization*, Amsterdam, John Benjamins.

MAURO MAGGI
(Università degli studi di Roma La Sapienza)

*Annotations on the Book of Zambasta, VIII*¹

*For Raffaele Torella, in friendship,
on the occasion of his 70th birthday
on 28th June 2020*

The article contains annotations on various passages of the Old Khotanese Book of Zambasta and, compared to previous editions and translations of the work, provides refinements and improvements in their interpretation and, in some cases, their reading. The verses taken into account are 2.120-121, 5.81-83, 13.104, 24.174-178, 24.513, 24.516, and 24.650-652.

L'articolo contiene annotazioni su vari passi del Libro di Zambasta anticocotanese e, rispetto alle precedenti edizioni e traduzioni dell'opera, fornisce raffinamenti e miglioramenti nella loro interpretazione e, in alcuni casi, nella loro lettura. I versi presi in considerazione sono 2.120-121, 5.81-83, 13.104, 24.174-178, 24.513, 24.516, and 24.650-652.

The Book of Zambasta (Z for short) is an anonymous religious poem in Khotanese, an Eastern Middle Iranian language attested in manuscripts from the fifth to the tenth century². Composed in the older variety of the language and largely based on Indian sources, the work expounds the teachings of Mahāyāna Buddhism and is the longest and probably the earliest original Khotanese Buddhist work, as there are good reasons for dating it to the second half of the fifth century (Maggi 2004b; Martini 2013). It is, thus, of the utmost importance for our knowledge of the Khotanese language and the history of Buddhism in Khotan.

¹ I thank Bhikkhunī Dhammadinnā and Douglas A. Hitch for comments and suggestions on a first draft of this article. Doug Hitch was also so kind as to revise my English.

² Standard edition with English translation by Emmerick (1968a); editio princeps with German translation by Leumann (1933-1936). Emmerick's work forms the starting point for the passages discussed here. All quotations from the Book of Zambasta are cited with Emmerick's translation unless otherwise specified. Verse numbers may be followed by the letters a, b, c, d to refer to single pādas. For Khotanese texts, including Z, I use the conventions listed at the end of the article. Different conventions used by others are converted to the ones adopted here for the sake of consistency. The same applies to the translations quoted.

This article is the third and last batch of philological and linguistic annotations that provide a background to, and in a few cases refine, the Italian translation of chapters 1-2, 5, 13-14, and 24 of the Old Khotanese Book of Zambasta that I published in an extensive anthology of Buddhist texts (Maggi 2004a). That translation activity gave me the opportunity to read afresh parts of the work and prompted me to introduce a number of refinements and improvements in interpretation as compared to the edition and translation by Ronald E. Emmerick³.

1. Z 2.120-121

1.1. Emmerick's edition and translation

- 120 *mahāsamudru ūtca biśśa khaṣṭa hāmāte*
bātā ahāna kaṣṭa hāmāte vātā biśśā 9
- 121 *ne balysi hoto hve' harbiśśu butte karā*
balysa buvāre śśūjīye hota samu 20

¹²⁰ The water in the great ocean can all be drunk up; the wind, on attachment by a noose, can all be held—¹²¹ a man does not at all know all the power of a Buddha. Buddhas alone know one another's power.

1.2. Revised edition and translation

- 120 *mahāsamudru ūtca biśśa khaṣṭa hāmāte*
*bātā ahāna *kaṣṭa (MS kaṣṭa) hāmāte vātā biśśā 9*
- 121 *ne balysi hoto hve' harbiśśu butte karā*
balysa buvāre śśūjīye hota samu 20

¹²⁰ Can all the water in the great ocean be drunk up? Can all the wind be held by means of a noose (and) a whip? ¹²¹ A man does not at all know all the power of a buddha. Buddhas alone know one another's power⁴.

These verses of chapter 2—a paraphrase of the Bhadramāyākāravāyākaraṇa⁵—are part of a passage where the miraculous powers and true omniscience of the Buddha are compared with the limited knowledge and strength of

³ The first and second batches of such annotations are contained in Maggi (2009 and forthcoming). Other articles in the series “*Annotations on the Book of Zambasta*” devoted to other aspects of the work are Maggi (2005, 2017, 2019a, 2019b) and Maggi/Martini (2014). The present one closes the whole series.

⁴ Cf. Maggi (2004a: 1206, 1272 n. 40).

ordinary men like the magician Bhadra, whose alleged knowledge and illusory tricks are stigmatised. In my Italian translation, I adhered to the transmitted text for the sake of caution and basically retained Emmerick's interpretation apart from taking the two clauses in Z 2.120 as questions rather than statements⁶. In a note, however, I suggested a possible emendation because the manuscript reading *kašta* poses problems. Ernst Leumann did not offer a translation of the hemistich 120cd where it occurs (Leumann 1933-1936: 24-25). Emmerick's 'on attachment' follows the analysis of the word by Harold W. Bailey, who regarded *kašta* as an instr.-abl. sg. with the palatalising ending *-ia* from an otherwise unattested action noun **kaṣṭi-* 'attaching' to Ir. **kas-* 'to attach' and translated it 'on being attached, after capture' (KT 6.43)⁷. Even disregarding that the regular OKh. instr.-abl. endings of the *i*-declension are *-e* and *-ä*, whereas *-ia* is locative⁸, Bailey's analysis is impossible because the consonant cluster *-ṣṭ-* cannot be palatalised and allows palatalisation to go back to the previous vowel segment *-a-*, which would give **kīṣṭa* (cf. the palatalised infinitive *khīṣṭe* 'to drink' and the corresponding past participle *khaṣṭa-* 'drunk', Hitch 1990: 193, 188). In light of these difficulties and on account of the strong similarity of the Brāhmī letters *t* and *n*, I suggest emending the manuscript slightly to read **kaśna*, the instr.-abl. sg. and, admittedly, the only occurrence of Khotanese *kaśa-* 'whip' ← Skt. *kaśa-* 'whip, thong' (Mahābhārata)⁹.

2. Z 5.81-83

2.1. Emmerick's edition and translation

- 81 *nā ju ye hā jsāte mamūka ni vā vara handari hīṣṭi*
karā hāḍe pabanā ne haṣḍā ttāna cu viñāno vipākā 1
 82 *trāmu vijñāno ttaraṃgga vi—ṣayīgya māšta samudru*
kāmā na-ro sarbāte urma—ysde bvāmatīno māsā . 82

⁵ See Régamey (1938) for this text lost in Sanskrit but preserved in the Tibetan and Chinese Buddhist canons.

⁶ In this I follow Bailey, Dict. 57 s.v. *kas-* 'adhere'.

⁷ Cf. the similar past participle *kaṣṭa-* 'attached' from Ir. **kaz-* (SGS 22 with references).

⁸ Bailey himself took *kašta* as a locative in Dict. 57 ('can all the wind, being caught in a noose, be held?').

⁹ In Maggi (2004a: 1272 n. 40), I opted for the meaning 'bridle [briglia]' on account of one of the meanings of the Skt. feminine *kaśā-* 'whip; rein, bridle' (see MW 265).

- 83 *samu kho durutātu pachī—ysu pabasto khvīyā samu—
dru kāmu na-ru kalpā nāta’—stā ttāmu buro aṃttā nā nīstā 3*
⁸¹ One does not go from here to there nor, on the other hand, does another arrive there. But the connexion is not broken because there is *vipāka* in *viññāna*. ⁸² So in *viññāna* there are waves in the great ocean of objects until the great sun of *bodhi* rises for you. ⁸³ Just as one could think of distance as waves in continuity in the ocean, there is then no end for them until the *kalpa* ends.

2.2. Revised edition and translation

- 81 *nā ju ye hā jsāte mamūka ni vā vara handari hīsti
karā hāḍe pabanā ne haṣḍā ttāna cu viññāno vipākā 1*
82 *trāmu viññāno ttaraṃgga vi—ṣayīgya māsta samudru
kā mā na-ro sarbāte urma—ysde *bvāmatīnai (MS bvāmatīno)
māstā . 82*
83 *samu kho duru tātu pachī—ysu pabasto khvīyā samu—
dru kāmu na-ru kalpā nāta’—stā ttāmu buro aṃttā nā nīstā 3*
⁸¹ No one here goes away nor does another there arrive hither, yet the connexion is not broken at all because there is fruition (of former karma) in the (storehouse) consciousness. ⁸² So there are billows in the (storehouse) consciousness, in the great ocean of (sense) objects, until the great sun of awakening rises, ⁸³ just as—to consider the waves in the ocean continuously for a long time—there will be no end of them so long until the aeon ends¹⁰.

The term *viññāna-*, *viññāna-* (Skt. *viññāna-*) denotes ‘consciousness’. According to Buddhist psychology, there are six kinds of consciousness connected with the six sense organs, including the mind, and their objects. In 5.81-82, *vi(j)ññāna-* is used in the specialised sense of ‘(storehouse) consciousness’ (Skt. *ālayaviññāna-*), an additional kind of consciousness peculiar to the *Vijñānavāda* or *Cittamātra* (Mind Only) tradition. The storehouse consciousness is likened to an ever-changing stream that underlies a being’s cycle of existence and is said to contain the “seeds” for what is perceived, through the various kinds of consciousness, as the phenomenal reality in general and, in particular, the personal experiences resulting from karma. In connection with the passage under consideration,

¹⁰ Cf. Maggi (2004a: 1223).

Ernst Leumann had already referred to the waves of consciousness (“Wellen des Vijñāna”) and the storehouse consciousness (“Ālayavijñāna”) in the Laṅkāvatārasūtra, one of the founding texts of the Vijñānavāda tradition (Leumann 1933-1936: 84).

In my Italian translation, I kept to Emmerick’s reading and interpretation with two significant differences. Thus, I did not regard the two verses as two sentences, but as a single sentence with the correlatives *trāmu ... samu kho* ‘So ..., just as’ introducing the principal clause and a clause of comparison respectively; and I did not translate *pachīysu* in verse 83 as an unclear and unspecified modal form with Emmerick (‘one could think’), but as a variant spelling of the infinitive **pachīysā* ‘to consider’ with *-u < -ā* by assimilation to the following initial *p-* (comparable to the assimilation in Z 2.159 *pulsu < pulsā* ‘to ask’ due to the preceding *-u-*)¹¹. Like Emmerick, however, I took *bvāmatīno* in verse 82 as a contraction of a form of *bvāmatīnaa-* + *-ū* ‘of *bodhi* [...] for you’ and I regarded the verb *pachīysu* as governing *durutātu* ‘distance’¹², though better solutions had been provided by Ernst and Manu Leumann.

Ernst Leumann’s emendation (1933-1936: 85) **bvāmatīnai* for *bvāmatīno* is probably correct. Nowhere else is *-o* a nom. sg. ending of the *aa*-declension and it is unlikely that *-o* is contracted from long *-e* ([*ē*] < *-ai*) + *-ū* 2 pl. enclitic pronoun ‘for you’, since *-ū* is, as a rule, not used after a long vowel or a diphthong and has no counterpart in the Late Khotanese version of the passage (Mañjuśrīnairātmyāvātārasūtra 241 KBT 125 *khu būre vā ne sarbe masta *urmaysdi bvāmavīja* ‘until the great sun of enlightenment rises’)¹³. The mistaken *-no* may have been influenced by the last syllable of *viñāno* in verse 81 in the manuscript line immediately above *bvāmatīno*.

In reading *durutātu* as a single word meaning ‘distance’ (acc.) in verse 83a, Emmerick followed Bailey’s interpretation of the hemistich as ‘just as one might count as distance the waves continuous in the ocean’ (KT

¹¹ I also referred the adjective *viśayīgya* ‘pertaining to objects’ (loc. sg. or nom.-acc. pl.) to the nom.-acc. pl. *taramgga* (‘the waves of objects’) rather than to the loc. sg. *samudru* (‘in the ocean of objects’), but this is unnecessary and is at variance with the metrical segmentation of the hemistich.

¹² ⁸ ²Così, nel grande oceano della coscienza ci sono le onde degli oggetti (*viśaya*), finché non sorge per voi il grande sole del risveglio, ⁸ ³proprio come, a considerare la lontananza, le onde di continuo non hanno fine nell’oceano finché non termina l’evo cosmico’ (Maggi 2004a: 1223).

¹³ Bailey’s emendation for manuscript *urmaysdā* (*-ā* for *-i* is common in manuscript P 4099). Emmerick’s unpublished translation.

6.121). The reading and analysis of *durutātu* as an otherwise unattested *-tāti-* abstract from *dura-* ‘far, distant’ is kept by Almuth Degener, who understands the word as ‘(temporal) distance’¹⁴. However, rather than postulating a hapax, it seems safer to keep to the earlier word division *duru tātu*, a collocation that occurs elsewhere and fits well in the present context. In his edition, Ernst Leumann read *durutātu* but did not translate this and most of the hemistich, but Manu Leumann, in the glossary, suggested reading *duru tātu* as a variant spelling of Z 14.15 *duru gāvu* ‘for a long time’ (Emmerick’s translation) on account of the common intervocalic variation *-g/-t/-y/-v-* and the enclitic character of *gāvu/tātu*.¹⁵ Sten Konow greeted this as an “ingenious suggestion” that “removes a great difficulty”.¹⁶ Bailey himself later changed his mind and, in his *Dictionary of Khotan Saka*, read *duru tātu* ‘far away indeed’, “[w]ith *tātu*, *gāvu*, later *-vā*, *-ā* ‘ever’”¹⁷. This view was in the end adopted also by Emmerick in his unpublished edition and translation of the Mañjuśrīnairātmyāvatārasūtra, where he reads and translates the Late Khotanese version of the passage thus (241-242 KBT 125):

*dūrā pastava behī *pabastau tta khvī samvadrra*

**kūṣṭi būrai ja kalpa na ṇa ’ṣta ttā būra atta ne byaidai*

They have proceeded for long, extremely continuous those waves in the ocean. Until the aeon ends, so long is there no end of them¹⁸.

The occurrence of the collocation in Suvarṇabhāṣottamasūtra 6.1.54 *cerā duru gāvu* ‘for as long as’ as a rendering of Skt. *dirgha-rātram* ‘for a long time’ shows that *duru gāvu/tātu* has a temporal value, as already understood by Emmerick in Z 14.15 (‘for a long time’), rather than a local one¹⁹.

Notice that I consider the subsequent infinitive *pachīysu* as the (middle intr. = passive) verb of an absolute clause, which has *khvīyā* ‘waves’ as its

¹⁴ Degener (1989: 281) s.v. *durutāti-*: ‘(zeitliche) Entfernung’.

¹⁵ Leumann (1933-1936: 84-85, 417) s.v. *gāvu*. Cf. Emmerick (1989a: 214 § 3.2.3.3.3.3.6 (1-2)) on the variant spellings *-g/-t/-y/-v-* for [ʔ] between vowels.

¹⁶ Konow (1939: 46) s.v. *gāvu*.

¹⁷ Dict. 161 s.v. *dura-*.

¹⁸ **pabastau* and *ṇa ’ṣta* result from Emmerick’s emendation for manuscript *bastau* (cf. Z 5.83 *pabasto*) and his expunction for manuscript *ṇa ’ṣta ṇa ’ṣta* (by dittography) respectively; **kūṣṭi* is Bailey’s emendation for MS *kūṣṭā* (cf. fn. 13) I have very slightly modified the translation.

¹⁹ See Skjærvø (2004: vol. 1, 96-97) and cf. BHSD 265 s.v. *dirgharātra-*.

overt subject ('to consider the waves') and which stands in an adverbial relation to the remainder of the clause of comparison: *pachīysu pabasto khvīyā samudru* 'to consider the waves in the ocean continuously'.

3. Z 13.104

3.1. *Emmerick's edition and translation*

104 *būtajina karmyo bei`ṇa . agvahāna āchā hvṇ`ndā .
balysā kari pracai nāstā tca—mna āchinei hāmīya 4*

¹⁰⁴ Through a preparer of food, through *karmas*, through poison, through indigestion, illnesses arise for a man. For a Buddha there is no reason at all why he should become ill.

3.2. *Revised translation*²⁰

¹⁰⁴ Through (possession) caused by demons, through *karmas*, through poison, through indigestion, a man has illnesses. For a Buddha there is no reason at all why he should become ill²¹.

Emmerick's translation of *būtajina* as 'Through a preparer of food' agrees with Bailey's hypothesis that the word is the same as *būkaja-* in Z 24.513 ('devoted men', Emmerick 'devotees') and *Jātakastava* 30 (8r3) *bukajā* ('steward') and that, in the present passage, it means 'preparer of food' (KT 6.255-256)²².

The alternation *-k/-t-* is far from common, however, and it seems safer to regard *būtaja-* as a different word from *būkaja-* (see § 5 below). On the contrary, the alternation *-t/-v-* in the vicinity of *ū* is well known already in Old Khotanese²³, so that it is far more likely that *būtaja-* is a variant of the adjective *būvaja-* 'caused by spirits', a *-ja-* derivative from *būta-*, LKh. *būva-* ← Skt. *bhūta-* 'a spirit (good or evil)'²⁴.

²⁰ No revision in edition.

²¹ Cf. Maggi (2004a: 1235).

²² Cf. Bailey (1953: 99: "The adj. E 14.104 *būtaja-* means 'connected with food'"), KT 4.97-98, and Dict. 29 s.v. *būtaja-*. Likewise, Degener (1989: 211) has a single entry for "*būkaja-*, *būtaja-* [...] 'Küchenmeister', 'Anhänger'".

²³ Cf. Emmerick (1989a: 214 § 3.2.3.3.3.3.6 (2)).

²⁴ See Degener (1989: 211 s.v. *būvaja-*) and cf. Leumann (1933-1936: 478 s.v. ²*būta-*), Dresden (1955: 461 s.v. *būva-*), and MW 761 s.v. *bhūta*.

In actual fact, this interpretation was already offered by Ernst Leumann (1933-1936: 183), who translated *būtajina* as ‘durch von Bhūta(-Geistern) herrührendes (Besessensein)’ and it is scarcely to be doubted that he was right. The belief that illness was not only of natural or human origin, but could also be brought about by the action of supernatural and demonic forces—and, for that matter, of karma, also mentioned in the passage under consideration—was common in ancient India and demonic possession and attacks formed an important part of the classical treatises of āyurvedic medicine²⁵. Thus, among the branches of medicine listed, for example, in § 1.1 of Ravigupta’s Siddhasāra, included in the Tanjur division of the Tibetan Buddhist canon and largely preserved in Khotanese, there is *bhūta-cikitsita*- ‘cure of (illnesses caused by) spirits’, which is rendered as Tibetan *gdon bsal-ba* ‘removing demons’ and, rather freely, as LKh. Siddhasāra 1.1 (2v5 KT 1.4) *sparśā*’ *pverq̄ma* ‘the removal of (illnesses due to) contact (with spirits)’²⁶.

In a medical context, the Khotanese adjective *būvaja*- ‘caused by spirits’ and its base lexeme occur together in Jīvakapustaka [2-3].9-10 (45v3-5 KT 1.137) *harbīśām būvajām āchām vīra . śā*’ *agada haṃdrauja pīsālyāna : naṣa māra-m jsa harbīśa vittāla : būva u kūra hīra* ‘in (the case of) all diseases due to Bhūtas, this remedy must be smeared on the forehead. Thereby will be pacified all Vetālas, Bhūtas, and evil things’, where *harbīśām būvajām achām* ‘in (the case of) all diseases due to Bhūtas’ translates Skt. **sarva-bhūta-vikāreṣu* (MS *sarva-bhūta-vakāriṣū*)²⁷. Spirits and diseases are referred to again in Jīvakapustaka [47].10 (91r4 KT 1.173) *būvyām jsa u yikṣyām jsa paśāva āchā* ‘diseases let loose by bhūtas and yakṣas’ (Skt. **bhūta-yakṣa-vikāreṣu*, MS *bhūta-yak<ṣ>a-vakāriṣū*)²⁸.

²⁵ See Wujastyk (1999) on the effects of possession by evil spirits and Leslie (1999) on the effects of bad karma according to classical Indian medicine.

²⁶ See Emmerick (1980-1982: vol. 1, 17; vol. 2, 14-15). Emmerick’s unpublished translation of the Khotanese.

²⁷ Ed. and trans. by Emmerick (1992b: 67-68). His omission of the last syllable and the following colon in *pīsālyāna* : is but a slip.

²⁸ Trans. by Konow (1941: 55).

4. Z 24.174-178

4.1. *Emmerick's edition and translation*

- 174 *maskhalä nā nāma hvāṣṭā ttārthā paḍauysā*
trāmu māñandā . dātā vaska kho ttā 'śā' .
- 175 *śātā vāṭco pūraṇi haṃberāte māstā*
śamā rrundi kuṃbiye ttāna pūraṇa nāma 5
- 176 *dādā aysāte nāma asāda hāra vāte nyūte*
ysamaśśandau ttānai aysāte nāma yidāndi .
- 177 *prabhūttā nāma śā baśdiye prahūtā .*
sañai kāḍe asāda śā vā saññai nāma 7
- 178 *niyaṇḍā nāma nyātā pūrā kye kaṣṭe*
ysamaśśandau harbiśśu śśavamera gyaḍiṃgya 78

¹⁷⁴ Maskarin by name is their chief, leading heretic. He is similar to a thief with regard to the Law. ¹⁷⁵ The second, moreover, Pūraṇa, fills the large Pot(hell)s of King Yama. Therefore he is Pūraṇa by name. ¹⁷⁶ The third, Asita by name, has taught the world about evil things. Therefore they have made him Asita by name. ¹⁷⁷ He is Prabhūta by name who is abounding in evil. Very imperfect are his plans: he is Saṃjaya by name. ¹⁷⁸ He is Nirgrantha by name, the son of Jñātr, who attached the whole world to the night-time of ignorance.

4.2. *Revised translation*²⁹

¹⁷⁴ Maskarin by name is their chief, leading heretic: he is similar to a thief (Skt. *taskara-*) with regard to the Dharma. ¹⁷⁵ The second, moreover, Pūraṇa, fills (Skt. *pūrṇa-* 'full') the large cauldrons of King Yama (in hell): therefore he is Pūraṇa by name. ¹⁷⁶ The third, Asita by name, has taught the world about unproved things (Skt. *asiddha-*): therefore they have made him Asita by name. ¹⁷⁷ He is Prabhūta by name who is abounding (Skt. *prabhūta-*) in evil. Very imperfect are his expedients (Kh. *sañā-*): he is Saṃjaya by name. ¹⁷⁸ He is Nirgrantha by name, the son of Jñāta, who attached the whole world to the impenetrable darkness of ignorance (*ajñāna-*)³⁰.

²⁹ No revision in edition.

³⁰ Cf. Maggi (2004a: 1251, 1281-1282 n. 139). See Skjærvo in Studies 2.140-143 for the reading and translation *śśavamera* 'impenetrable darkness' instead of *śśava-mera* 'night-time' and Emmerick (1989b) for its etymology from **-mahā-tantra-* 'moon darkness' (hence *śśavamāra-* 'darkness of a moonless night').

This passage presents the six chief heretical masters of the time of the Buddha and explains their names through folk-etymological puns based on the phonic resemblance of the names with other words and aimed at disparaging the heretics. The Sanskrit, Gandhari, and Pali equivalents of the Khotanese names are³¹:

Khotanese	Sanskrit	Gandhari	Pali
<i>Maskhala</i>	<i>Maskarin Gośālīputra</i> ³²	<i>Makhali</i> ^o	<i>Makkhali Gosāla</i>
<i>Pūraṇa</i>	<i>Pūraṇa Kāśyapa</i>	<i>Purana Kaśava</i>	<i>Pūraṇa Kassapa</i>
<i>Aysāta</i>	<i>Ajita Keśakambala</i>	—	<i>Ajita Kesakambala</i>
<i>Prabhūta</i>	<i>Krakuda Kātyāyana</i>	—	<i>K/Pakudha Kaccāyana</i>
<i>Samñāa</i>	<i>Samjayin Vairāṭīputra</i> ³³	<i>Samjaia</i> 'follower of S.'	<i>Sañjaya Belaṭṭhiputta</i>
<i>Niyaṇḍa</i>	<i>Nirgrantha Jñātaputra</i> ³⁴	—	<i>Nigaṇṭha Nātaputta</i>

Comparison of the heretical masters' names with the Sanskrit and Khotanese terms involved in their folk-etymological explanations, summarised in the following list, shows that folk etymology is based on Indian rather than Khotanese genuine words and that, therefore, the passage under consideration is probably not a Khotanese original invention but depends on some Indian source, be it in Sanskrit, Gandhari, or other³⁵:

Khotanese name	Sanskrit folk etymology	Khotanese equivalent
174 <i>Maskhala</i>	<i>taskara-</i> 'thief'	<i>ttā'sa'</i>
175 <i>Pūraṇa</i>	<i>pūrṇa-</i> 'full'	<i>hamber-</i> 'to fill'
176 <i>Aysāta</i>	<i>asiddha-</i> 'unproved'	<i>asāda-</i> (LW)
177 <i>Prabhūta</i>	<i>prabhūta-</i> 'abundant'	<i>prahūta-</i> (LW)

³¹ The Khotanese names *aysāta-* and *maskhala-* also occur elsewhere in the Book of Zambasta: see Leumann (1933-1936: 390, 483). See also KT 6.5 s.v. *aysāta-*, 137 s.v. *niyaṇḍā*, 276 s.v. *maskhalā*, 355 s.v. *saṃñāi*. Other variants of the Sanskrit and Pali names exist: see BHSD 7, 163, 300, 351, 420, 551 for Sanskrit and Malalasekera (1938: vol. 1, 37-38; vol. 2, 64-65, 89-90, 242-243, 398-400, 999-1000) for Pali. For the Gandhari names, see Falk (2010: 16 and 2014: 35 l. 20).

³² See Vogel (1970: 23 n. 19) for a Skt. patronymic *Gośālāputra* evinced from Tibetan instead of the metronymic *Gośālīputra*.

³³ See Vogel (1970: 25 n. 29) for variants of the metronymic or patronymic containing the verb *vi-vad-* 'to contradict, oppose' instead of **vi-vaṭ-* 'to shout'.

³⁴ So in Avadānaśataka.

³⁵ I refer only to Sanskrit for the sake of simplicity.

177 <i>Samñāa</i>	<i>saṃjñā-</i> ‘idea’	<i>sañā-</i> ‘expedient’ instead of <i>saṃñā-</i> (LW) ‘idea’!
178 <i>Niyaṇḍa</i>	<i>a-jñāna-</i> ‘ignorance’	<i>gyaḍīnaa-</i> ‘pertaining to ignorance’.

The probable dependence on some Indian source is especially clear in the case of 174 *Maskhala*, which corresponds to Skt. *Maskarin* (basically ‘religious mendicant’³⁶) and is explained by means of Kh. *ttā`ṣa`* - ‘thief’. This makes sense only if Kh. *ttā`ṣa`* - is here a rendering, so far unnoticed, of Skt. *taskara-* ‘thief’ and indicates that the original wordplay used the rhyme *MaskarV* ~ *taskarV*, essentially lost in Khotanese.

Likewise, 178 *Niyaṇḍa*, notwithstanding its decidedly Prakrit appearance, presupposes Skt. *Nirgrantha Jñātaputra*, whose patronymic, omitted in Khotanese, is explained by means of Kh. *gyaḍīnaa-* ‘pertaining to ignorance’ from *gyaḍia-*, the usual equivalent of Skt. *ajñāna-*, *avidyā-* ‘ignorance’.

The substitution of Kh. 176 *Aysāta* (Skt. *Asita*) for Skt. *Ajita* is presumably due to the influence of the name of the Indian sage Asita—mentioned in Z 2.21 as *Aysāta* together with the mythical sages Bhārgava, Vyāsa, and King Bhagīratha in a speech by Maskarin himself—or of the chaplain of King Śuddhodana, the father of the Buddha, who features in Z 24.230-239. Notice that, in the explanation of this heretical master’s name, the loanword 176 *asāda* (← Skt. *asiddha-*) does not simply mean ‘evil’, but has to be taken in its philosophical sense of ‘unproved’, as was already understood by Ernst Leumann³⁷.

Of special interest is the pun on *Samñāa/Saṃjayin*, which most likely originally connected Skt. *Saṃjayin* and *saṃjñā-* ‘idea, conception, imagination’. Metrics requires that 177 *sañai kāḍe* ~ ~ ~ ~ scans as a five-mora segment and, thus, shows that *sañai* ~ - ‘his expedients’ (< *sañā* + -*ā* 3 sg. enclitic pronoun) is not a defective spelling for *sa(m)ñai* -- ‘his ideas’ (< *saṃñe* + -*ā*)³⁸. So, some misunderstanding, or deliberate adaptation, of the source occurred in the explanation of the name 177 *Samñāa* by means of Kh. *sañā-* ‘plan, expedient, means’ (a genuine

³⁶ MW 794.

³⁷ Leumann (1933-1936: 318) (‘unbewiesene’). See MW 120 s.v. *asiddha* and 1215 s.v. ²*siddha*.

³⁸ So mistakenly understood in Maggi (2004a: 1251).

Iranian word) instead of expected Kh. *saṃñā-* ‘idea’ (a borrowing ultimately from Skt. *saṃjñā-* through Prakrit)³⁹. As well as in the Book of Zambasta, the terms *sañā-* ‘expedient’⁴⁰ and *sañā-brīcā-* ‘familiarity with expedients’ render Skt. *upāya-* ‘(skilful) means’ and *upāya-kausalya-* ‘skilfulness in means’ in fifth-century translations of Mahāyāna sūtras such as the *Saṅghāṭasūtra* and the *Ratnakūṭa*⁴¹. It is significant that *sañā-* occurs in Z 2.18 in a statement on the uselessness of the ascetic practices and teachings of the heretics who opposed the Buddha: *sañā ni pharu harbi[ś]ś[ä] kūra tteri atāca kho ni rro parāhā* ‘Their many expedients are all false, as ineffective as their restraint’ (Emmerick). Given the specifically Mahāyānist conception of “skilful means”, that is, the Buddha’s and bodhisattvas’ ability to appear and teach in different ways in order to adapt to different beings so as to bring them to deliverance⁴², the replacement of expected *saṃñā-* ‘idea’ with *sañā-* ‘expedient’ suggests that the passage was misunderstood or, more probably, was changed on purpose in order to coat it in a Mahāyānist varnish. Though the two abstract concepts of *saṃñā-* ‘idea’ and *sañā-* ‘expedient’ are close in shape and semantics and Khotanese novices could confuse them, the composer of the Book of Zambasta likely understood the Indian pun and made an independent contribution that would work with his audience.

5. Z 24.513

5.1. Emmerick’s edition and translation

513 *tcei`mañinu bāgya [nā]tasta būkaja balysä*
auysāra buṇḍa kyau tta śśānye daimä 13

⁵¹³ The devotees of the Buddha have fallen before our eyes, enraged, wicked, you whom I see lying thus.

³⁹ This *Samñāa* (Skt. *Samjāyin*, Pali *Sañjaya*) was erroneously equated by Leumann (1933-1936: 10-11) (followed by Régamey 1938: 5 n. 17) with the different person Z 2.26 *sacai nigranthā* ‘Satyaka the Nirgrantha’ (Skt. *Satyaka*, Pali *Saccaka*), “a contemporary of Buddha, described as a great debater” (BHSD 554).

⁴⁰ See Leumann (1933-1936: 510), Emmerick (1966: 177, 1967: 44), and Dict. 417. Bailey’s idea (KT 6.354-355, essentially followed by Tremblay 2005: 434) that Kh. *sañā-* is from Gandhari *sañā-* (i.e. *saṃñā-* < OInd. *saṃjñā-*) is untenable on phonological and semantic grounds.

⁴¹ See Maggi (forthcoming) on Z 14.11-12, with further references.

⁴² On “skilful means”, see Pye (2003).

5.2. *Revised edition and translation*

513 *tcei`mañinu bāgya [vā]ta sta būkaja balysä*
auysära buṇḍa kyau tta śśānye daimä 13

⁵¹³ Instead of (tearing out my sinner's) eyes, you have been the slaughterers of the Buddha, enraged, wicked, you whom I see lying thus⁴³.

This verse is part of an episode which occurs towards the end of a paraphrase of the prophecy of decline of Buddhism narrated in the Candragarbhasūtra⁴⁴ and which contains the mourning of King Duṣprasaha for the death of the whole community of monks, who have fought and murdered one another, and for the consequent complete loss of any religious guidance. Emmerick's translation of the first hemistich can be compared with Bailey's working translation 'under our (their) eyes fallen the devoted men of the Buddha' (KT 6. 255). Apart from the postposition *bāgya*, which does not mean 'before' but 'instead of' and the like as was later shown by Emmerick himself⁴⁵, the interpretation of the verse is problematic because of the gap in [+] *ta sta* and the rare word *būkaja*, which occurs only here and once more as LKh. Jātakastava 30 (8r2) *bukajā*⁴⁶.

In a first very tentative attempt at translation, Leumann, who was unaware of the Late Khotanese occurrence of *būkaja-*, translated this provisionally as 'grief' or 'tears' and read and emended the defective word as *n{ät}a< >sta* 'has seated' 3 sg. f. or pl. m. prf. intr. of *nād*-⁴⁷. Bailey and Emmerick opted for the hapax legomenon [*nä*]*tasta* '[they] have fallen', 3 pl. prf. intr. from a participle **kasta-* (with preverb **ni-*) to an alleged Ir. root **kah-* 'to move'⁴⁸. Since the king, who is holding the dead bodies of the religious leader Śiṣyaka and his opponent Sūrata in his arms, first addresses his own preceptor Śiṣyaka in the 2 sg. in verse

⁴³ Cf. Maggi (2004a: 1266, 1284 n. 167), where, however, I translated *tcei`mañinu bāgya* as 'instead of (being) my eyes [invece di essere i miei occhi]'.

⁴⁴ Cf. Nattier (1991: 170-173).

⁴⁵ See Emmerick in Studies 1.86 and Skjærvø in Studies 3.112-114.

⁴⁶ Z 13.104 *būtajina* has been considered to belong here but is actually a different word (see § 3.2 above).

⁴⁷ Leumann (1933-1936: 346-347: 'Uns auf die [auf meine] Augen ist niedergesessen [hat sich gelegt] Trauer [?, oder: Tränen ?] um den Priester'). See SGS for the form.

⁴⁸ KT 6.44 and Dict. 57, 73; cf. EWAia 1.332-333.

512 (*thu paro dritai balysānu utāru* ‘You have kept the noble Buddha-command’), he is then likely, in the immediately following verse under consideration, to address both leaders or, more probably, all slaughtered monks in the 2 pl. Accordingly, I supplement the 2 pl. prf. intr. [*vā*]*ta sta* ‘you have been’ (cf. Z 23.102 *uhu nu hā ggīhā vāta sta* ‘you have been their helpers’).

We have seen above (§ 3.2) that Emmerick’s translation ‘devotees’ for *būkaja* in the verse in question followed a suggestion by Bailey, who tried to reconcile in this way the present occurrence of the word with the *bukajā* occurring in Jātakastava 30 (8r3) and interpreted by him as ‘preparer of food, steward’ on account of a supposed Pali counterpart *bhatta-kāraka*⁴⁹. Bailey’s interpretation is based on the assumption that *būkaja-* is a *-ja-* derivative from a *būka-* occurring only once in Old Khotanese in a fragmentary and not entirely clear context (Z 18.8 *būku* acc. sg.) and understood by him as ‘food’⁵⁰. Apparently, Bailey thought of a semantic development such as ‘preparer of food → steward → [attendant →] devoted person’. Apart from the little convincing outcome ‘devoted person’, however, the meaning of the supposed base lexeme *būka-* itself is far from certain in my opinion, since this *būka-* may, but need not, be a different word from LKh. *būka-* in the eighth-century documents, where it seems to mean ‘troop’ or the like⁵¹.

The rendering ‘slaughterers [macellai]’ for *būkaja* in my Italian translation was conjectural. It fits well the context in the Book of Zambasta, with figurative reference to the monks who have all ‘slaughtered’ one another in a fierce battle, and the context in the Jātakastava, where a king of the deer—the Bodhisattva, the Buddha-to-be in one of his births—offers himself as a ransom to the servant who is about to slaughter a pregnant doe for the king of Benares, who used to have meat at every meal:

⁴⁹ See fn. 22 and cf. Dresden (1955: 482 s.v. *bukaja-*).

⁵⁰ See KT 6.256 s.v. *būku* (where the reference “291⁸” is to the folio and verse numbers), Emmerick (1966: 162-163, 174), and Dict. 295. On a possible but uncertain Late Khotanese occurrence of *būka* ‘food’ in the Chinese-Khotanese bilingual P 2927 24 KT 3.103 see Bailey, KT 4.97-98 and Takata (1988: 207 [食べ物 *tabemono* ‘food’ with no Chinese parallel], 227).

⁵¹ So Skjærvø, Catalogue lxxvii: “*troop’ (? there are first, second, etc. *būkas* [...])”. Bailey suggested some kind of “workmen or employed at the guard post” (KT 4.95) and an “official title” (cf. Dict. 295).

- 30 *barbīrye gūysna ja va mu 'śda' jsa pātcā .*
bukajā puṣa tsuai . rruṃdāña hurāśña
cu didi pha ysairka yuḍḍā pāñe paṃna .
śau na ye vena tvī cuai jīye hera :

³⁰ Afterwards for the sake of the pregnant deer, in compassion, you went at once to the slaughterer in the royal kitchen [?], [the deer] which made so much lament in front of all. There was no one save you who would give his life for her⁵².

I suggest separating *būkaja-* from *būka-* and, instead, seeing in *būkaja-* a loanword from a variant of the OInd. word **bokkasa-*—continued as Prakrit *bokkasa-* ‘a non-Aryan people’, Nepali *bokso* ‘wizard’, *boksi* ‘witch’⁵³—and variously spelled, on account of its origin from some non-Indo-Aryan language, as Vedic *paulkasá-* ‘a type of half-caste’, Classical Skt. *pulkasa-* ‘name of a despised mixed tribe’ (with the Epic Skt. variants *pukkasa-*, *pukkaśa-*, *pukkaṣa-*, and the variant *bukkasa-* ‘a *caṇḍāla*’ in lexicographical works), *pulkasaka-* ‘a barbarian tribe’, and Pali *pukkusa-* ‘a despised caste’⁵⁴.

In particular, Kh. *būkaja-* mirrors an Indic variant **bokkaśa-* with initial *b-* like Skt. *bukkasa-*, Prakrit *bokkasa-*, Nepali *bokso*, *boksi* and with *-ś-* like Epic Skt. *pukkaśa-*. The change Skt. *o* → Kh. *ū* is common and is to be found, for instance, in *koṭi-* ‘crore, ten millions’ → *kūla-* (Z 22.160 +), *vimokṣa-* ‘release, deliverance’ → *vāmūha-* through Prakrit (Z 10.30 +), etc. The occasional rendering of Skt. *-ś-* through Khotanese *-j-* can also be observed in OKh. *ājāvāṣa-* ‘venomous snake’ ← Skt. *āśvīṣa-* (Z 3.76 etc.), *vijīṣḍā-tara-* ‘more special, very special’ ← Skt. *viśīṣṭa-* (Suvarṇabhāṣottamasūtra 6.3.5), LKh. *dajagraiva* ← Skt. *daśagrīva-*, the name of a demon (Book of Vimalakīrti 352 KBT 111, Rāmāyaṇa P 2781.119 KT 3.70), and LKh. *arja-* ‘piles’ ← Skt. *arśas-* (Siddhasāra 0.11 [2r1 KT 1.4] +)⁵⁵.

⁵² Ed. and trans. by Dresden (1955: 425), with ‘slaughterer’ substituted for his ‘steward’. Jātakaṣṭava 30-31 summarises the Nigrodhamigajātaka (see Dresden 1955: 447 for references).

⁵³ See Turner (1966: 528 s.v. **bōkkasa-*).

⁵⁴ See MW 631 s.vv. *pukkaśa* and *pukkasa*, 638 s.v. *pulkasa*, 652 s.v. *paulkasá*, 733 s.v. *bukkasa*, and PED 461 s.v. *pukkusa*. On this word family see Kuiper (1991: 54-57) and EWAia 2.152 s.v. *pulkasa-*. Vedic *pūklaka-*, *pūlkaka-*, *pūtlaka-* ‘little boy’ are not variants of *pulkasa-* (so e.g. MW 631) but dialectal or popular variants of *putraká-* ‘little son’ (Hoffmann 1982: 86-90).

⁵⁵ See KT 6.11 for *ājāvāṣa-*, Skjærvø (2004: vol. 2, 340) for *vijīṣḍā-tara-*, Maggi (2013: 143) for *dajagraiva*, and Konow (1941: 85) for *arja-*.

As for the meaning, it should be recalled that, in lexicographical works arranged systematically, Skt. *bukkasa-* and *caṇḍāla-* ‘outcaste’ are considered to be synonymous and are listed in the proximity of people involved in killing animals like hunters, fishers, fowlers, butchers, and the like (e.g. in Hemacandra’s *Abhidhānacintāmaṇi* 927-933⁵⁶). Similarly, Pali *pukkusa-* denotes ‘a low social class, the members of which are said (in the *Jātakas*) to earn their living by means of refuse-clearing’ and which features right at the end in enumerations of social classes and despised clans that usually also include the term *caṇḍāla-* ‘outcaste’⁵⁷. All this suggests that the task of killing was left to outcastes such as *caṇḍālas* or *bukkāsas/pukkusas*, as is confirmed by the occurrence of Buddhist Skt. *caṇḍālaka-* ‘outcaste, *caṇḍāla*’ in a list of enemies of birds⁵⁸. In the Book of *Zambasta* itself, when the heretics plot to get rid of the Buddha, one of them even suggests to have him killed by a *caṇḍāla*: *Z* 2.30 *o ye caṇḍālā ttata heḍā se thvī ttuṣṣe yana ṣṣīve* ‘or (suppose) one gives money to a *caṇḍāla*: “Destroy him in the night”’.

In light of the preceding, it seems likely that Kh. *būkaja-* is a borrowing from Indic **bokkaśa-* ‘outcaste’ and that, in the *Jātakastava*, LKh. *bukajā* is the butcher who has to slaughter the deer and, in the Book of *Zambasta*, OKh. *būkaja* are, metonymically, the slaughterers of the Buddha, that is, the destroyers of his teachings entrusted to them.

6. *Z* 24.516

6.1. *Emmerick’s edition and translation*

516 [pha]ru padya [+]¹ –u –² rre ṣṣamana banāre
kye nā rro gratu hvāñāte śśākṣāvātā heḍā 6

⁵¹⁶ In many ways ... The king, the monks lament: “Who will preach instruction to us now, will give the *śikṣāpadas*?”.

⁵⁶ See the ed. and trans. by Böhtlingk/Rieu (1847: 173-174), where verses 927-933 contain words for: ‘55. *Jäger* (4 W[örter]). – 56. *Jagd* (5 W.). – 57. *Der mit Netzen jagt* (2 W.). – 58. *Jägernetz* (2 W.). – 59. *Strick* (7 W.). 60. – *Fischer* (3 W.). – 61. *Angel* (2 W.). – 62. *Fischernetz* (2 W.). – 63. *Fischkorb* (2 W.). – 64. *Vogelsteller* (2 W.). – 65. 66. *Fleischer* (4 W.). – 67. *Schlachthaus*. – 68. *Netz, Kette oder Käfig, um Vogel und vierfüßige Thiere zu halten*. – 69. *Schlinge* (2 W.). – 70. *Grube, um Wild zu fangen* (2 W.). – 71. *Falle* (2 W.). – 72. 73. *Mensch, der zu keiner Kaste gehört* (7 W.). – 74. 75. *Caṇḍāla* (10 W. [including *bukkasa-*])’ (my italics).

⁵⁷ See PED 461 s.v. *pukkusa* and Cone (2013: 99-100 s.v. *caṇḍāla*).

⁵⁸ See BHSD 223 s.v. *caṇḍālaka*.

6.2. *Revised edition and translation*

516 [pha]ru padya [hā]ruv[a]’ rre ṣṣamana banāre
kye nā rro gratu hvāñāte śśākṣāvātā heḍā 6

⁵¹⁶ In many ways the merchants (and) the king lament the (dead) monks: “Who will preach instruction to us now, will give the moral commandments?”⁵⁹.

This verse belongs to the same episode as the previously discussed one (513). If the rare verb *ban-* ‘to lament’ is regarded as only intransitive (SGS 93), the resulting translation ‘the monks lament’⁶⁰ is at variance with the fact that the monks are said to be all dead after their fight: Z 24.499 *śśūjāteye vaska āksuvīndā juvāre varī puṣṣo śśūjātu biśśā tvīṣṣe yanīndā* ‘they [i.e. the monks] begin to fight with one another. At once they all destroy each other’ (Emmerick). I suggest that the rare verb *ban-* ‘to lament’ is both transitive and intransitive just like English *to lament* and that *ṣṣamana* ‘monks’ is the object and the substantive lost in the gap and *rre* ‘king’ are the subject of the clause: ‘the ... (and) the king lament the [dead] monks’.

Since the presence of merchants as the king’s counsellors (cf. Skt. *śreṣṭhin-*) in this episode is indicated by Z 24.508 *karavātā naltsute hā hārvyau haṃtsa* ‘In the morning [after the monks’ slaughter] he [i.e. the king] went out with the merchants’ and 514 *kye vara hāruvānu ttā vāta* ‘Those of the merchants who were there’ (Emmerick), I based my Italian translation of the passage on the restoration [hār]u[va] rre ‘the merchants (and) the king’, which fits the metre perfectly. In doing so, I deviated from Leumann’s and Emmerick’s reading of the last damaged akṣara, under which they saw a subscript hook (–)⁶¹, and tentatively regarded the sign as a letter fragment detached from a contiguous, now lost folio and specularly stuck below verse 516, as sometimes happens. The suggested supplement [hār]u[va] rre – ◡ ◡ – would fit the 6-mora segment before the final cadence *ṣṣamana banāre* ◡ ◡ ◡ ◡ ◡ ◡ in this metre B hemistich⁶².

⁵⁹ Cf. Maggi (2004a: 1266, 1284 n. 169).

⁶⁰ Cf. Leumann (1933-1936: 346: ‘Die Mönche jammern’), KT 6.224 (‘the *śramaṇas* lament’), and Dict. 268 (‘the *śramaṇa*-ascetics lament’).

⁶¹ The hook mostly continues Ir. intervocalic *-š- and is often combined with hiatus-bridging consonants like -t- or -v-. It possibly expressed breathiness (Emmerick 1992a: 158-165) and is usually transliterated as an apostrophe or an aleph sign.

⁶² Cf. Hitch (2014: 3-5).

Although closer inspection of the facsimile shows that there is no letter fragment stuck to this folio and that the sign is a fairly well preserved subscript hook⁶³, no other restoration seems to be possible than *[hā]ruv[a]*’ with hook instead of the normal nominative-accusative plural *hāruva* (Z 5.96 22.208 23.45 23.136). The unexpected addition of the hook may be accidental and be due to influence of similarly spelled words like, for instance, Z 2.174 13.106 23.16 *aruva*’ ‘medicine’ or words with the loc. pl. ending *-uvo*’, as is the case with such words with unetymological subscript hook as the loc. sg. Z 13.82 22.115 *āvuto*’, 6.24 *āvuvo*’ ‘village’ instead of normal *avuvo* (Z 13.90)⁶⁴ and the acc. sg. Z 23.122 *ṣṣuvo*’ ‘news’ instead of normal *ṣṣuvo* (Z 22.333)⁶⁵.

7. Z 24.650-652

7.1. Emmerick’s edition and translation

650 *[cī]tā nā ju ye vīri [asā]du o śśāru yīndā*
atī saṃtsera . hāmāte māstā vivāgā

651 *[+] rro tterā oṣye scā[tā] [mara]ta ye kāli .*
cī haṃgrīsīndi . balysā pūra hamālā . 51

652 *[vāna] balysā nīstā kye yuḍu yīndā pamā—*
tu tte bhikṣu-saṃggā . vasve māstā ṣṣahāne 52

⁶⁵⁰ If one does not do evil to anyone or (if one does) good, there will be very great *vipāka* for him in *saṃsāra*. ⁶⁵¹ (So) also at such an evil time here, in a period when the sons of the Buddha assemble in a group. ⁶⁵² (Apart from) the Buddha, there is no one who can measure the pure, great virtues of this Bhikṣusaṅgha.

7.2. Revised edition and translation

650 *[a]tā nā ju ye vīri [asā]du o śśāru yīndā*
atī saṃtsera . hāmāte māstā vivāgā

651 *[ṣṣai] rro tterā oṣye scā[tā] hama]taye kāli .*
cī haṃgrīsīndi . balysā pūra hamālā . 51

⁶³ See the facsimile in Vorob’ëv-Desjatovskij/Vorob’ëva-Desjatovskaja (1965: 293).

⁶⁴ Cf. Hitch (2015: 316 n. 33; 2016: 283).

⁶⁵ Cf. Hitch (2016: 270).

652 [väna] balysä ništā kye yuḍu yīndā pamā—
tu tte bhikṣu-saṃggā . vasve mäštā ṣṣahāne 52

⁶⁵⁰ (If) one does not do much evil to anyone or (if one does) good, there will be very great ripening for him in the cycle of birth.

⁶⁵¹ Even in such an evil age (as the present Kaliyuga), at the very time when the sons of the Buddha (i.e. the monks) assemble in a group, ⁶⁵² (apart from) the Buddha, there is no one who can measure the pure, great virtues of this community of monks⁶⁶.

This passage belongs to the conclusion of chapter 24 and of the entire Book of Zambasta. After narrating the rise of Buddhism and prophesying its future decline due precisely to the decay of the monks' morality, chapter 24 closes with an exhortation for the audience to follow the Buddha's teachings and rely on the community of monks, whose importance is said to be so invaluable that only the Buddha, not ordinary people, can fully appreciate it.

I follow Leumann in restoring [a]tā 'much' at the beginning of verse 650 and take the clause as a conditional clause with implied conjunction, because Emmerick's [cī]tā 'when, if' would make the hemistich hypermetrical with an initial 6-mora instead of 5-mora segment ([a]tā nā ju ye ~ ~ ~ ~ ~).

I restore the adverbial collocation [ṣṣai] rro 'even' at the beginning of verse 651, as it fits this passage and occurs frequently in the text (Z 2.201 2.243 5.29 6.57 13.58 24.646 ṣṣai rro; 14.87 ṣṣai rro; 2.114 2.125 8.30 22.232 ṣṣei rro; 2.198 ṣṣai rru; 20.61 ṣṣai ro; 5.104 ṣṣai ru). In my Italian translation, I also suggested that what Leumann read [~ ~] ta ye and Emmerick restored as [mara]ta ye must be a single word because the remaining ye (not translated by Emmerick) can only be the indefinite pronoun 'one' and cannot be reconciled with the context. I presumed that the defective word may be an adjective accompanying the gen.-dat. *kāli* 'time' but offered no restoration. Here I propose to restore [hama]taye *kāli* 'at the time itself, at the very time': though the gen.-dat. sg. m. of the emphatic reflexive pronoun *hamata-* is *hamatye* in Z (2.229 24.453), OKh. *hamataye*, apparently from a secondary *hamataa-*, occurs in IOL Khot 36/3 (Kha. viii.1) a5, Catalogue 255-256 [ba]lysūñāvūysei hamataye āsā

⁶⁶ Cf. Maggi (2004a: 1267, 1284-1285 nn. 173-174).

[tc]e[ra] ‘the bodhi-seeker should himself do memorization of the teaching’ (also a3 *hamatayai*, i.e. *hamataye* + -ä 3 sg. enclitic pronoun).

It should also be noted that *kāli* ‘period’ is probably not just a synonym in apposition with the preceding *scā[tā]* ‘time’, as was translated by Emmerick. Rather, *kāli* generically means ‘a time’ here, whereas *tterä ośye scā[tā]* ‘in such an evil age’ has a specific reference and matches the Kali age (Kaliyuga), our decayed fourth and last age of the world, which is mentioned and compared with the Kṛtayuga, the first, golden age of the world, in Z 24.3-4:

- 3 *kho ye pyūṣḍe crrāmu ṣu vāta kṛyi yuggā satva*
crrāmu vātā kālā kho ttārā staurā hamāstä 3
4 *ne ne ju hā māñātā karā kho ye hāra daiyā*
kalā yuggī harbiśśu tterā stauru hamāstä 4

³ When one hears what beings were like in the Kṛtayuga, what that age was like, how very greatly has it changed! ⁴ It does not resemble at all how one sees things. So greatly has the Kaliyuga changed it all⁶⁷.

Mention of *tterä ośye scā[tā]* ‘such an evil age’ at the end of chapter 24—where the virtues of the Buddhist monks are indicated as a religious and moral counteragent—is a way to recall *kalā yugga-* ‘the Kaliyuga’ at chapter beginning and can be regarded as an instance of concentric concatenation.

Mauro Maggi
Università degli studi di Roma La Sapienza
Istituto italiano di studi orientali
Circonvallazione Tiburtina 4
I-00185 Roma
mauro.maggi@uniroma1.it

⁶⁷ This refers to the change to worse of the *uysnorāṇu padaṃgya* ‘the build of beings’ in the preceding verse 24.2. See Degener (1989: 204-205) on *padaṃgyā-* ‘build, physical appearance’ (from the past participle of *padīm-* : *padanda-* ‘to make, build’) and Skjærvø in Studies 3.86-88 on its further, secondary meanings ‘exposition, description; display, presentation’.

Conventions

<i>italics</i>	(or roman in quotations in italics) letters not clearly visible;
+++	illegible or partially preserved akşaras for which no reading can be suggested;
–	illegible or partially preserved consonant base or vowel mark for which no reading can be suggested;
[+++]	akşaras lost in a lacuna;
[–]	consonant base or vowel mark lost in a lacuna;
[abc]	restoration of a lacuna; translation of a lacuna;
<abc>	editor’s supplement where the manuscript has no lacuna;
{abc}	editor’s deletion;
*	emendation (the manuscript reading is given in parentheses or in a note);
—	blank in the manuscript between parts of a word;
(abc)	explanatory amplification in translations.

Abbreviations

Books

BHSD	Edgerton (1953: vol. 2)
Catalogue	Skjærvø (2002)
Dict.	Bailey (1979)
EWAia	Mayrhofer (1992-2001)
KBT	Bailey (1951)
KT 1-7	Bailey (1945-1985)
MW	Monier-Williams (1899)
PED	Rhys Davids/Stede (1925)
SGS	Emmerick (1968b)
Studies 1-3	Emmerick/Skjærvø (1982-1997).

Other

1/2/3	first/second/third person
acc.	accusative
f.	feminine
gen.-dat.	genitive-dative
instr.-abl.	instrumental-ablative

intr.	intransitive
Ir.	Iranian
LKh.	Late Khotanese
loc.	locative
m.	masculine
OInd.	Old Indian
OKh.	Old Khotanese
pl.	plural
prf.	perfect
sg.	singular
Skt.	Sanskrit.

Khotanese texts

Jātakastava	ed. and trans. Dresden (1955: Ch. 00274)
Jīvakapustaka	ed. KT 1.135-196 (Ch. ii.003); paragraph numbers according to Konow (1941) with added verse numbers according to Emmerick (1992b) and Chen (2005)
Mañjuśrīnairātmyāvātārasūtra	ed. KBT 113-135 (P 4099)
Saṅghāṭasūtra	ed. and trans. Canevascini (1993)
Siddhasāra	ed. KT 1.2-132 (Ch. ii.002), 5.315-324 (P 2892); paragraph numbers according to Emmerick (1980-1982)
Suvarṇabhāsottamasūtra	ed. and trans. Skjærvø (2004)
Z = Book of Zambasta	ed. and trans. by Emmerick (1968a).

References

- Bailey, Harold W., 1945-1985, *Khotanese texts*, Cambridge, Cambridge University Press: vol. 1, 1945; vol. 2, 1954; vol. 3, 1956 (vols. 1-3, 2nd ed. in one volume 1969; repr. 1980); vol. 4, *Saka texts from Khotan in the Hedin collection*, 1961 (repr. 1979); vol. 5, 1963 (repr. 1980); vol. 6, *Prolexis to the Book of Zambasta*, 1967; vol. 7, 1985.
- Bailey, Harold W., 1951, *Khotanese Buddhist texts*, London, Tailor's Foreign Press (2nd ed., Cambridge, Cambridge University Press, 1981).
- Bailey, Harold W., 1953, "Analecta Indoscythica, I". *Journal of the Royal Asiatic Society* 1953/3-4: 95-116.
- Bailey, Harold W., 1979, *Dictionary of Khotan Saka*, Cambridge, Cambridge University Press.

- Böhtlingk, Otto / Rieu, Charles, 1847, *Hemākandra's Abhidhānakīntāmaṇi: ein systematisch angeordnetes synonymisches Lexicon*, St. Petersburg.
- Canevascini, Giotto, 1993, *The Khotanese Saṅghāṣasūtra: a critical edition*, Wiesbaden, Reichert.
- Chen, Ming 陳明, 2005, *Dunhuang chutu Hu yu yi dian "Qipo shu" yanjiu 敦煌出土胡語醫典 <<耆婆書>> 研究 = A study on Sanskrit text of Jīvaka-pustaka from Dunhuang*, Taipei, Xin Wenfeng Chuban Gongsi.
- Cone, Margaret, 2013, *A dictionary of Pali. 2: G-n*, Bristol, The Pali Text Society.
- Degener, Almuth, 1989, *Khotanische Suffixe*, Stuttgart, Steiner.
- Dresden, Mark J., 1955, *The Jātakastava or "Praise of the Buddha's former births": Indo-Scythian (Khotanese) text, English translation, grammatical notes and glossaries*, Philadelphia, The American Philosophical Society (= Transactions of the American Philosophical Society, n.s., 45/5): 397-508.
- Edgerton, Franklin. 1953. *Buddhist Hybrid Sanskrit grammar and dictionary*. Vol. 1, *Grammar*; vol. 2, *Dictionary*, New Haven, Yale University Press.
- Emmerick, Ronald E., 1966, "The nine new fragments from the Book of Zambasta". *Asia Major*, n.s., 12/2: 148-178.
- Emmerick, Ronald E., 1967, "The ten new folios of Khotanese". *Asia Major*, n.s., 13/1-2: 1-47.
- Emmerick, Ronald E., 1968a, *The Book of Zambasta: a Khotanese poem on Buddhism*, London, Oxford University Press.
- Emmerick, Ronald E., 1968b, *Saka grammatical studies*, London, Oxford University Press.
- Emmerick, Ronald E., 1980-1982, *The Siddhasāra of Ravigupta*, Wiesbaden, Steiner: vol. 1, *The Sanskrit text*, 1980; vol. 2, *The Tibetan version with facing English translation*, 1982.
- Emmerick, Ronald E., 1989a, "Khotanese and Tumshuqese". In: Rüdiger Schmitt (Hrsg.), *Compendium linguarum Iranicarum*, Wiesbaden, Reichert: 204-229.
- Emmerick, Ronald E., 1989b, "A Khotanese nightmare". In: *Études irano-aryennes offertes à Gilbert Lazard*, Paris: 95-98.
- Emmerick, Ronald E., 1992a, "The Dunhuang MS. Ch 00120: its importance for reconstructing the phonological system of Khotanese". In: Alfredo Cadonna (ed.), *Turfan and Tun-huang: the texts: encounter of civilizations on the Silk route*, Firenze, Olschki: 145-170.
- Emmerick, Ronald E., 1992b, "The Svastika antidote". *Journal of the European Āyurvedic Society* 2: 60-81.

- Emmerick, Ronald E. / Skjærvø, Prods O., 1982-1997, *Studies in the vocabulary of Khotanese*, Wien: vol. 1, 1982; vol. 2, 1987; vol. 3, ed. by Ronald E. Emmerick, contributed by Giotto Canevascini et al., 1997.
- Falk, Harry, 2010, "The 'split' collection of Kharoṣṭhī texts". *Annual report of the International Research Institute for Advanced Buddhology at Soka University* 14: 13-23.
- Falk, Harry, 2014, "A new Gāndhārī *Dharmapada* (Texts from the split collection 3)". *Annual report of the International Research Institute for Advanced Buddhology at Soka University* 18: 23-62.
- Hitch, Douglas A., 1990, "Old Khotanese synchronic umlaut". *Indo-Iranian journal* 33/3: 177-198.
- Hitch, Douglas A., 2014, "Meter in the Old Khotanese *Book of Zambasta*". *Ars metrica* 2014: 1-43.
- Hitch, Douglas A., 2015, "Contracted diphthongs in Old Khotanese". *Indo-Iranian journal* 58/4: 293-324.
- Hitch, Douglas A., 2016, "Contracted semivowels in Old Khotanese". *Indo-Iranian journal* 59/3: 259-294.
- Hoffmann, Karl, 1982, "Vedica". *Münchener Studien zur Sprachwissenschaft* 41: 61-94.
- Konow, Sten, 1939, "The late Professor Leumann's edition of a new Saka text, II". *Norsk tidsskrift for sprogvidenskap* 11: 5-84.
- Konow, Sten, 1941, *A medical text in Khotanese: Ch. ii 003 of the India Office Library with translation and vocabulary*, Oslo, Dybwad.
- Kuiper, Franciscus B. J., 1991, *Aryans in the Rigveda*, Amsterdam, Rodopi.
- Leslie, Julia, 1999, "The implications of the physical body: health, suffering, and karma". In: John Hinnells / Roy Porter (eds.), *Religion, health, and suffering*, London, Kegan Paul International: 23-45.
- Leumann, Ernst, 1933-1936, *Das nordarische (sakische) Lehrgedicht des Buddhismus: Text und Übersetzung*, aus dem Nachlaß hrsg. von Manu Leumann, Leipzig, Brockhaus.
- Maggi, Mauro, 2004a, "Il libro di Zambasta: capitoli 1-2, 5, 13-4, 24". In: Raniero Gnoli (a cura di), *La rivelazione del Buddha*, vol. 2, *Il Grande veicolo*, introd. di Claudio Cicuzza e Francesco Sferra con contributi di Mauro Maggi e Cristina Pecchia, Milano, Mondadori: 1193-1285.
- Maggi, Mauro, 2004b, "The manuscript T III S 16: its importance for the history of Khotanese literature". In: Desmond Durkin-Meisterernst et al. (eds.), *Turfan revisited: the first century of research into the arts and cultures of the Silk Road*, Berlin, Reimer: 184-190, 457.

- Maggi, Mauro, 2005, “Annotations on the *Book of Zambasta*, II: Khotanese *maṃkyā*”. *Bulletin of the Asia Institute*, n.s., 19: 103-109.
- Maggi, Mauro, 2009, “Annotations on the *Book of Zambasta*, I”. In: Desmond Durkin-Meisterernst / Christiane Reck / Dieter Weber (Hrsg.), *Literarische Stoffe und ihre Gestaltung in mitteliranischer Zeit: Kolloquium anlässlich des 70. Geburtstages von Werner Sundermann*, Wiesbaden, Harrassowitz: 153-171, pl. i.
- Maggi, Mauro, 2013, “Four more questions and answers from the Khotanese *Book of Vimalakīrti*”. In: Matteo De Chiara / Mauro Maggi / Giuliana Martini (eds.), *Buddhism among the Iranian peoples of Central Asia*, Wien, Verlag der Österreichischen Akademie der Wissenschaften, 131-151.
- Maggi, Mauro, 2017, “Annotations on the *Book of Zambasta*, IV: Ronald E. Emmerick’s notes”. In: Enrico Morano / Elio Provasi / Adriano V. Rossi (eds.), *Studia philologica Iranica in memory of Professor Gherardo Gnoli*, Roma, Scienze e lettere: 273-291.
- Maggi, Mauro, 2019a, “Annotations on the *Book of Zambasta*, V: Indian parallels to 2.139 and the musk of Khotan”. In: Almut Hintze / Desmond Durkin-Meisterernst / Claudius Naumann (eds.), *A thousand judgements: Festschrift for Maria Macuch*, Wiesbaden, Harrassowitz: 291-309.
- Maggi, Mauro, 2019b, “Annotations on the *Book of Zambasta*, VI: Ronald E. Emmerick’s **nvāta*- and Persian *navāxtan*”. In: Sabir Badalkhan / Gian Pietro Basello / Matteo De Chiara (eds.), *Iranian studies in honour of Adriano V. Rossi*, Napoli, Università degli studi di Napoli L’Orientale, 2019: 519-522.
- Maggi, Mauro, forthcoming, “Annotations on the *Book of Zambasta*, VII: chapter 14 revisited”. *Folia Orientalia* 57 (2020).
- Maggi, Mauro / Martini, Giuliana, 2014. “Annotations on the *Book of Zambasta*, III: chapter 18 no more”. *Scripta: an international journal of codicology and palaeography* 7: 139-158.
- Malalasekera, George Peiris, 1938, *Dictionary of Pāli proper names*, London, Murray, 2 vols.
- Martini, Giuliana, 2013, “Bodhisattva texts, ideologies, and rituals in Khotan in the fifth and sixth centuries”. In: Matteo De Chiara / Mauro Maggi / Giuliana Martini (eds.), *Buddhism among the Iranian peoples of Central Asia*, Wien, Verlag der Österreichischen Akademie der Wissenschaften, 13-69.
- Mayrhofer, Manfred, 1992-2001, *Etymologisches Wörterbuch des Altindoarischen*, Heidelberg, Winter: vol. 1, 1992; vol. 2, 1996; vol. 3, 2001.

- Monier-Williams, Monier, 1899, *A Sanskrit-English dictionary*, Oxford, The Clarendon Press.
- Nattier, Jan, 1991, *Once upon a future time: studies in a Buddhist prophecy of decline*, Berkeley, Calif., Asian Humanities Press.
- Pye, Michael, 2003, *Skilful means: a concept in Mahayana Buddhism*, 2nd ed., London, Routledge.
- Régamey, Konstanty, 1938, *The Bhadramāyākāravāyākaraṇa: introduction, Tibetan text, translation and notes*, Warsaw, Nakład Towarzystwa Naukowego Warszawskiego.
- Rhys Davids, Thomas W. / Stede, William, 1921-1925, *The Pali Text Society's Pali-English dictionary*, Chipstead, The Pali Text Society.
- Skjærvø, Prods O., 2002, *Khotanese manuscripts from Chinese Turkestan in the British Library: a complete catalogue with texts and translations*, with contributions by Ursula Sims-Williams, London, The British Library.
- Skjærvø, Prods O., 2004, *This most excellent shine of gold, king of kings of sutras: the Khotanese Suvarṇabhāsottamasūtra*, Cambridge, Mass., Harvard University Press, 2 vols.
- Takata, Tokio 高田 時雄, 1988, *Tonkō shiryō ni yoru Chūgokugo shi no kenkyū: kyū, jū seiki no Kasei hōgan 敦煌資料所見中國語言史之研究: 九・十世紀の河西方言 = A historical study of the Chinese language based on Dunhuang materials: the Hexi 河西 dialect of the ninth and tenth centuries*, Tōkyō, Sōbunsha.
- Tremblay, Xavier, 2005, "Irano-Tocharica et Tocharo-Iranica". *Bulletin of the School of Oriental and African Studies* 68: 421-449.
- Turner, Ralph L., 1966, *A comparative dictionary of the Indo-Aryan languages*, London, Oxford University Press.
- Vogel, Claus, 1970, *The teachings of the six heretics according to the the Pravrajyāvastu of the Tibetan Mūlasarvāstivāda Vinaya: with an appendix containing an English translation of the pertinent sections in the Chinese Mūlasarvāstivāda Vinaya*, Wiesbaden, Steiner.
- Vorob'ëv-Desjatovskij, Vladimir S. / Vorob'ëva-Desjatovskaja, Margarita I., 1965, *Skazanie o Bhadre: novye listy sakskoj rukopisi "E": faksimile teksta, transkripcija, perevod, predislovie, vstupil'naja stat'ja, glossarij i prilozhenie*, Moskva, Nauka.
- Wujastyk, Dominik, 1999, "Miscarriages of justice: demonic vengeance in classical Indian medicine". In: John Hinnells / Roy Porter (eds.), *Religion, health, and suffering*, London, Kegan Paul International: 256-275.

Linguistica e Filologia

ELENCO DEI REVISORI PER I NUMERI 37, 38 E 39

Serenella Baggio	Patrizia Lendinara
Emanuele Banfi	Rita Librandi
Maria Patrizia Bologna	Luca Lorenzetti
Elisabetta Bonvino	Silvia Luraghi
Marina Buzzoni	Marco Mancini
Massimo Cerruti	Carla Marelo
Marina Chini	Giuliano Mion
Claudio Ciociola	Massimo Palermo
Vittoria Corazza	Filippo Pecorari
Dario Corno	Michele Prandi
Pierluigi Cuzzolin	Paolo Ramat
Paolo D'Achille	Adriano Rossi
Cécile Desoutter	Maria Grazia Saibene
Angela Ferrari	Stefania Spina
Livio Gaeta	Mario Squartini
Maria Gottardo	Federica Venier
Roberta Grassi	Letizia Vezzosi
Claudio Iacobini	Jacqueline Visconti
Elisabetta Jezek	Miriam Voghera
Francesco Lo Monaco	Alessandro Zironi

Finito di stampare
nel mese di dicembre 2020
Sestanteinc - Bergamo