AILA World Congress 2014
One world, many languages
Brisbane, 10-15 August 2014

The role of input properties on lexical development in foreign language acquisition: Transparency and frequency

Ada Valentini and Roberta Grassi
Università degli Studi di Bergamo, ITALY

Lexical development in SLA

- strong correlation between high frequency of lexical units in the input and early emergence in learners' repertoires (Laufer / Nation 1995);
- cognates appear early in learners' lexical repertoires, regardless of their frequency in the input (Horst / Collins 2006).

European VILLA project

European cross-linguistic Project VILLA *Varieties of Initial Learners in Language Acquisition: Controlled classroom input and elementary forms of linguistic organizatio*n

Dimroth, C., Rast, R., Starren, M. & Wątorek, M. (2013). "Methods for studying the learning of a new language under controlled input conditions: The VILLA project". *EUROSLA Yearbook* 13. Amsterdam: Benjamins, pp. 109-138.

Learner groups & source languages

Learner groups	Source languages				
	Italian	English	French	German	Dutch
Adult learners, meaning-based input	15	17	17	20	20
Adult learners, form-based input	**14**	18	19	-	20
Child learners, meaning-based input	-	(planned)	-	19	-

aims

- to investigate the role of transparency in Second Language lexical development on the basis of a translation task administered **before the beginning** of the VILLA experiment;
- to verify longitudinally the role of transparency and frequency on lexical development **in the VILLA experiment** on the basis of a picture production task administered at two separate intervals of the 14-hour Polish course.

Learner groups & source languages

Learner groups	**Source languages**				
	Italian	English	French	German	Dutch
Adult learners, meaning-based input	15	17	17	20	20
Adult learners, form-based input	**14**	18	19	-	20
Child learners, meaning-based input	-	(planned)	-	19	-

Translation Test (= TT)

5 groups of students heard 120 Polish words and were asked to write down their translation in their L1 (e.g. POL: *Brazy-lij-ka* 'Brasilian.F.NOM');

TT originally designed to identify Polish words immediately comprehensible (= transparent) to all groups of learners of different 5 L1s (correctly translated by 50% native speakers);

35 Italian native speakers without knowlegde of Slavic languages

Some results from Italian group

POLISH	ITALIAN	ACCURACY SCORES
ekonomista	*economista* 'economist'	100%
autobus	*autobus* 'bus'	53%
dialog	*dialog**o*** 'dialogue'	94%
banan	*banan**a*** 'banana'	35%

Levenshtein distance

Polish	Italian	Levenshtein distance	recogniza bility
autobus	*autobus*	0	+++
franzusem	*francese0*	3	++
kucharką	*cu0o0ca0*	5	+

0 = deletion

2 factors based on stress & syllable

1. Phonological identity of stressed syllable

PL *amerykanin* [amɛri'kanʲin] IT [ameri'ka:no]

'American' phonological identity

PL *kanadyjka* [kana'dɨjka] IT [kana'de:ze]

'Canadian' no phonological identity.

2 factors based on stress & syllable

Target and source language & stress

Polish: predictable stress (second last syllable)

Italian: non predictable stress; consider minimal pairs as:

an*cora* 'anchor' *an****co****ra* 'again'

Some results from Italian group

POLISH	ITALIAN		ACCURACY SCORES
ekonomista	*economista*	'economist'	100%
autobus	*autobus*	'bus'	53%
dialog	*dialog**o***	'dialogue'	94%
banan	*banan**a***	'banana'	35%

2 factors based on stress & syllable

2. Coincidence of stressed syllable

SAME STRESSED SYLLABLE

*pilote*m	PL	[pi'lɔtɛm]	IT [pi'lɔːta]
architektem	PL	[arçi'tɛktem]	IT [arki'tɛtto]
artystkq	PL	[ar'tɨstkɔw̃]	IT [ar'tista]

DIFFERENT STRESSED SYLLABLE

pilot	PL	['pilɔt]	IT [pi'lɔːta]
architekt	PL	[ar'çitɛkt]	IT [arki'tɛtto]
energia	PL	[ɛ'nɛrgja]	IT [ener'ʤiːa]

PHONOL. IDENT. STRESSED SYLL. & COINCIDENCE STRESSED SYLL.: 4 TYPES

Types of Polish words	Pol.	It.
IDENT_COINC	*amerykanin* [amɛri'kanʲin]	*americano* [ameri'kaːno]]
IDENT_NONCOINC	*dokument* [dɔ'kumɛnt]	*documento* [doku'mento]
NONIDENT_COINC	*kanadyjka* [kana'dɨjka]	*canadese* [kana'deːze]
NONIDENT_NONCOINC	*belgijka* [bɛl'giiˌka]	*belga* ['bɛlga]

repeated measures ANOVA

Independent variables:

- phonological identity or non identity of stressed syllable;
- coincidence or non-coincidence of stressed syllable in word pair;

Dependent variable: set of accuracy scores.

Alpha level: .05.

Results of ANOVA

Significant main effects both for:

phonological identity of the accented syllable ($F_{1,33}$ = 216, p < .000, partial η^2 = .87, observed power = 1.0)

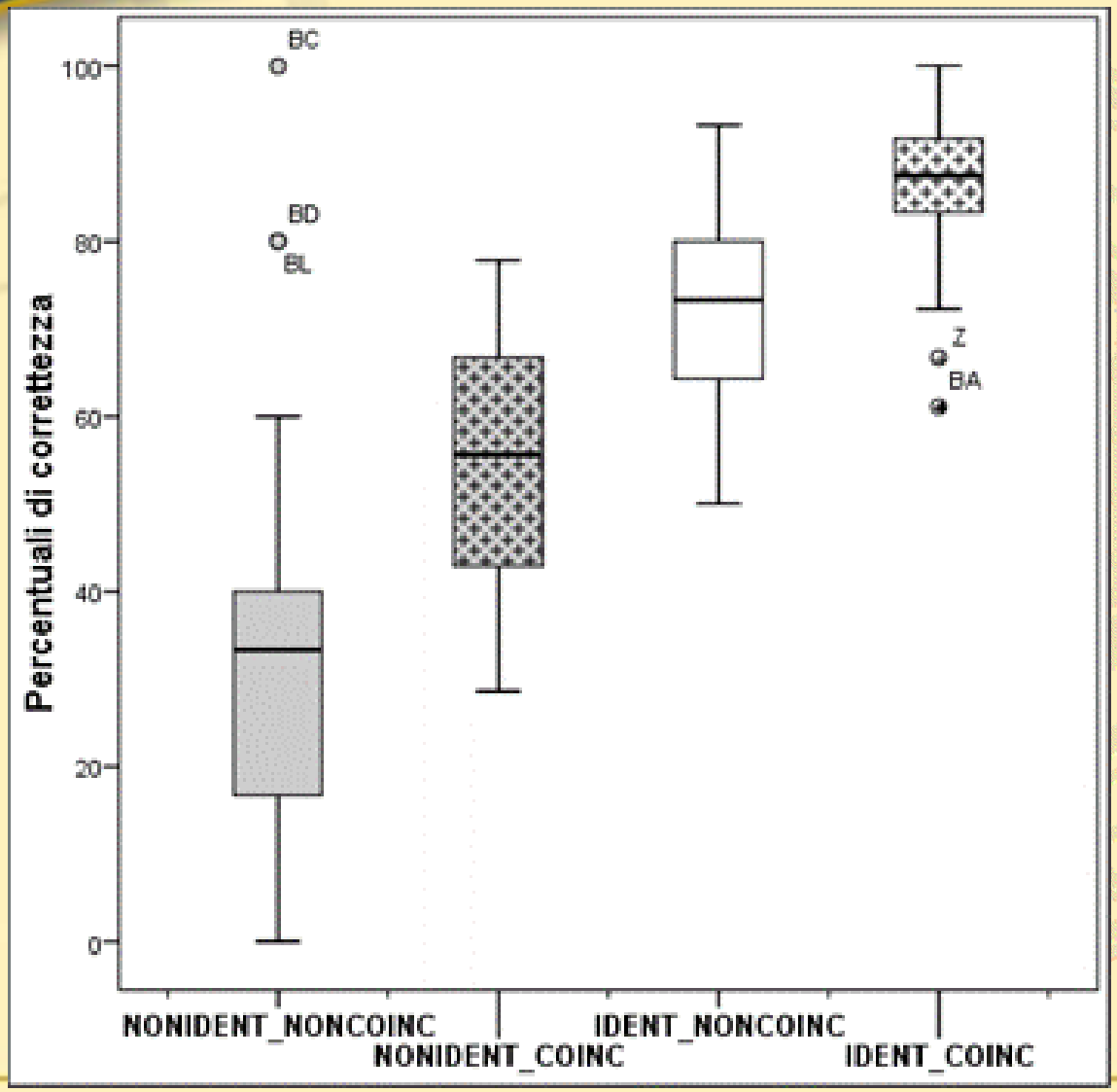
same or differ. stressed syllable ($F_{1,33}$= 46, p < .000, partial η^2 parziale = .58, observed power = 1.0).

- no interaction found between 2 variables

Accuracy scores in 4 types of words

Types of words	% accuracy values		s. d.
NONIDENT_NONCOINC	34,6%	46%	24,5
NONIDENT_COINC	54,8%		15,0
IDENT_NONCOINC	74,3%	82,2%	10,7
IDENT_COINC	87%		8,6

Distribution of mean accuracy values



Accuracy scores in 4 types of words

Types of words	% accuracy values		s. d.
NONIDENT_NONCOINC	34,6%	46%	24,5
NONIDENT_COINC	54,8%		15,0
IDENT_NONCOINC	74,3%	82,2%	10,7
IDENT_COINC	87%		8,6

TIM | VILLA - Exposure & Testin | SquirrelMail 1.4.22

www.lili.uni-osnabrueck.de/villa/Main/ExposureTesting

VILLA Varieties of Initial Learners in Language Acquisition

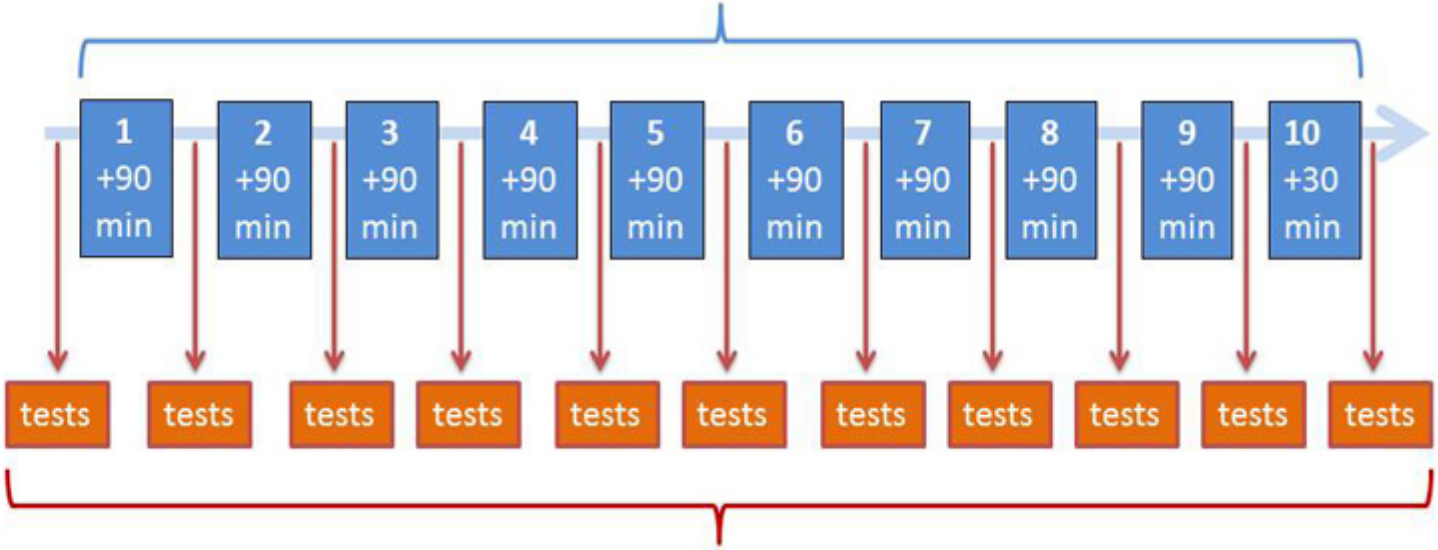
VILLA

Startseite
Steering Committee
Research Questions
Learners & Input Conditions
Exposure & Testing
Publications
Research Teams & Contact

Exposure & Testing

Groups of learners were exposed to 14 hours of Polish input that was monolingual and structured in such a way as to allow the testing of students at different linguistic levels. What makes this project unique is that the input was held constant under relatively natural and interactive exposure conditions, and that it was entirely audio and video recorded and transcribed.

- 14 hours of input over 2 weeks (10 days)
- Audio and video recorded



25 language tests

The learners' performance was investigated over time at different levels of language (perception, comprehension, grammatical analysis and production) in order to find out more about the most elementary stages of linguistic organisation resulting from the

VILLA: fully controlled input conditions

- Same teacher (native speaker) in all 11 editions - 5 Countries
- Predetermined script (in two versions: «implicit» vs. «explicit»)
- Control on input

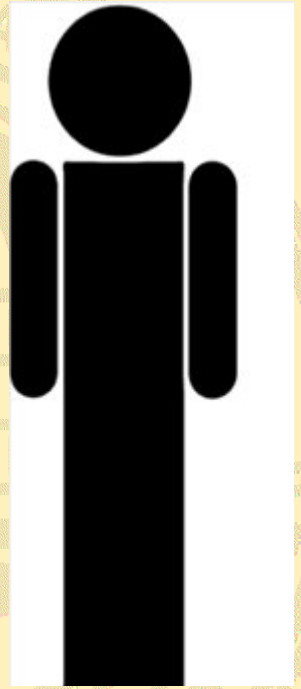
  (ex.): Italian Edition investigated here:
  - 1305 Types
  - 56321 Tokens
  - Type/Token ratio: 0,023

Picture Production Test (= PP)

- 32 items of which 24 = TT
- Oral questions + visual (non verbal) hints

STIMULUS (QUESTION)	EXPECTED ANSWER
Kim on jest? 'Who is he?'	*On jest Aktorem* (STRUM.M.) 'He is actor'
Kim ona jest? 'who is she?'	*Ona jest Aktorką* (STRUM.F.) 'She is actress'
Kto to jest ? 'who is that?'	*To jest Aktorka* (NOM.F.) 'That is an actress'

Picture Production Test (PP)



High Transparency - High Frequency

Target	L1 (Italian)	FREQUENCY at T1
In-ży-nier	‘in-ge-nie-re’ NOM. M.	2 types; 30 tokens
Fran-cuz	‘fran-ce-se’ NOM. M.	5 types; 73 tokens
Fo-to-gra-fem	‘fo-to-gra-fo’ STRUM. M.	4 types; 41 tokens
Bra-zy-lij-ka	‘bra-si-lia-na’ NOM. F.	3 types; 56 tokens
Por-tu-gal-ką	‘por-to-ghe-se’ STRUM. F.	3 types; 58 tokens
Ar-ty-stką	‘ar-ti-sta’ STRUM. F.	2 types; 34 tokens

High Frequency – Low Transparency

Target	L1 (Italian)	FREQUENCY at T1
Chiń-czyk	‘ci-ne-se’ NOM. M.	2 types; 48 tokens (24 chińczyk; 24 chińczykiem)
ku-char-ka	‘cuo-ca’ NOM. F.	3 types; 51 tokens (25 kucharka; 25 kucharką; 1 kucharzem)

Zero Frequency – High Transparency

Target	L1 (Italian)
In-for-ma-tyk	‘in-for-ma-ti-co’ NOM. M.
pi-lo-tem	‘pi-lo-ta’ STRUM. M.
Po-li-cjan-tką	‘po-li-ziot-ta’ STRUM. F.
Ka-na-dyj-ką	‘ca-na-de-se’ STRUM. F.
Au-stra-lij-czyk	‘au-stra-lia-no’ NOM. M.
Gre-kiem	‘gre-co’ STRUM.M.
Se-kre-tar-ka	‘se-gre-ta-ria’ NOM.F.

Research Questions

1. When
2. How

is word stress learnt?

Research Questions

1. When:

   Accuracy at T1 vs. T2

2. How:

   Lexical or grammatical learning ?

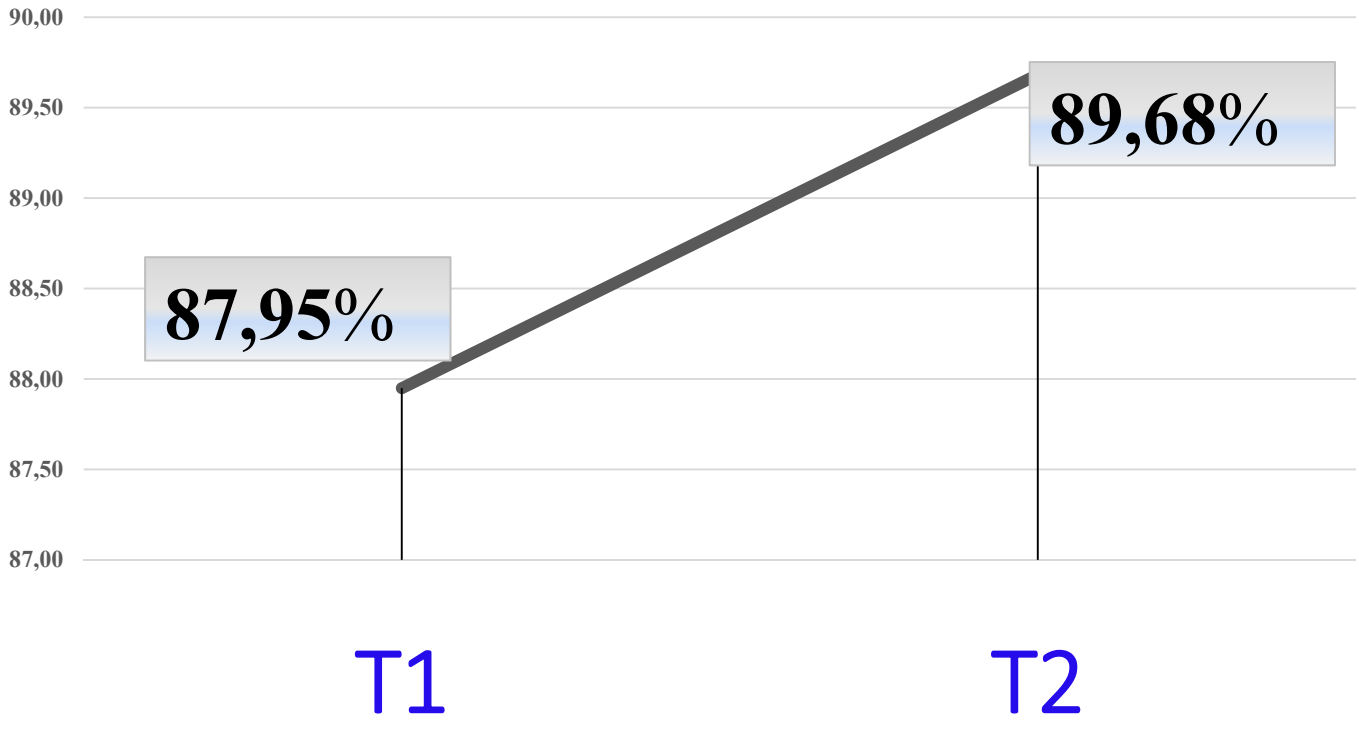
   Does diffusion follow:

   - Input Frequency?
   - Input Transparency?
   - Stressed syllable coincidence with L1?

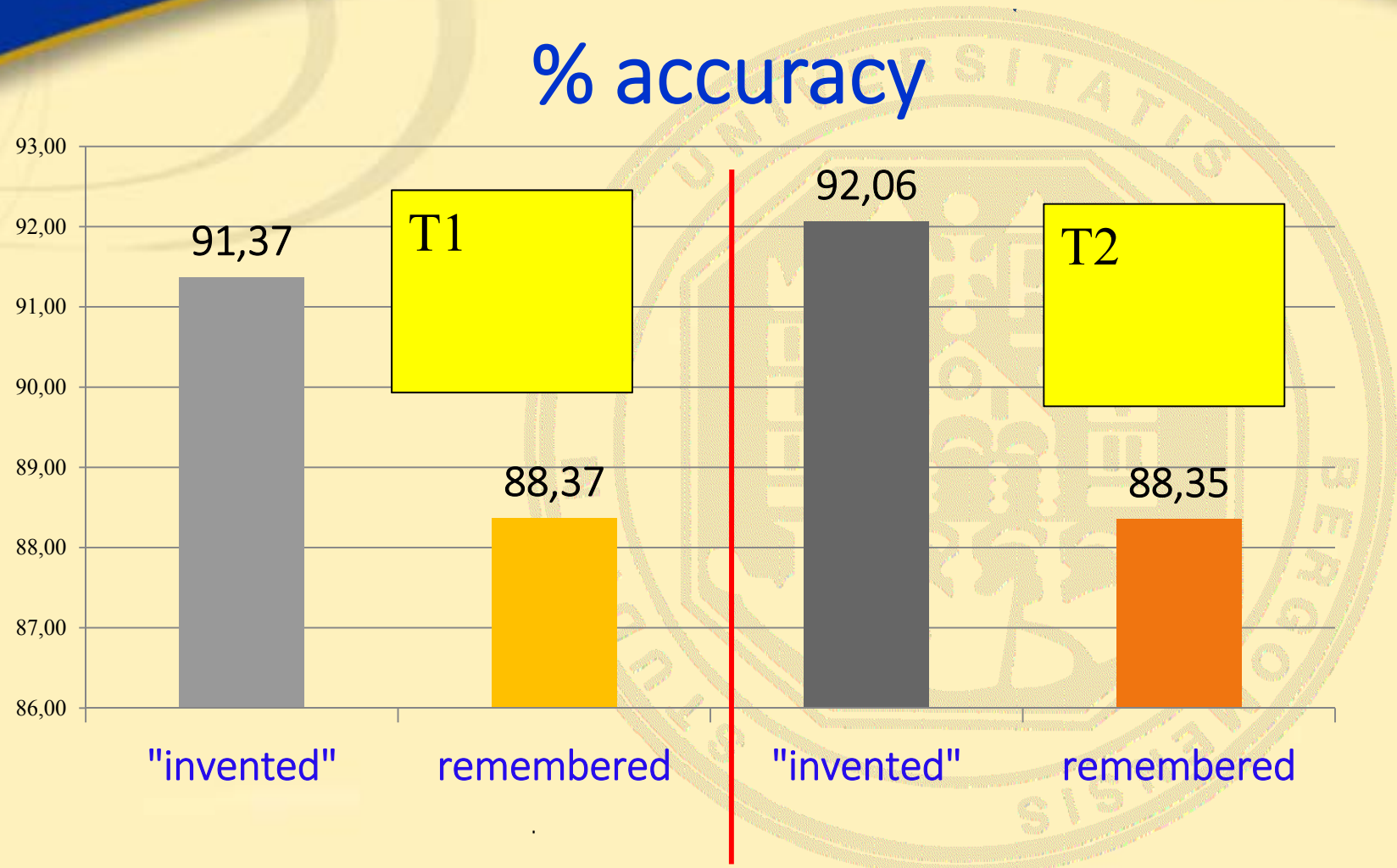
Word stress in VILLA (Italian Edition)

- word Stress in Polish course: NOT a part of the syllabus = not focussed in the input
- word stress in L1: unpredictable, contrastive stress
- word stress in the target language: predictable; general rule: second last syllable

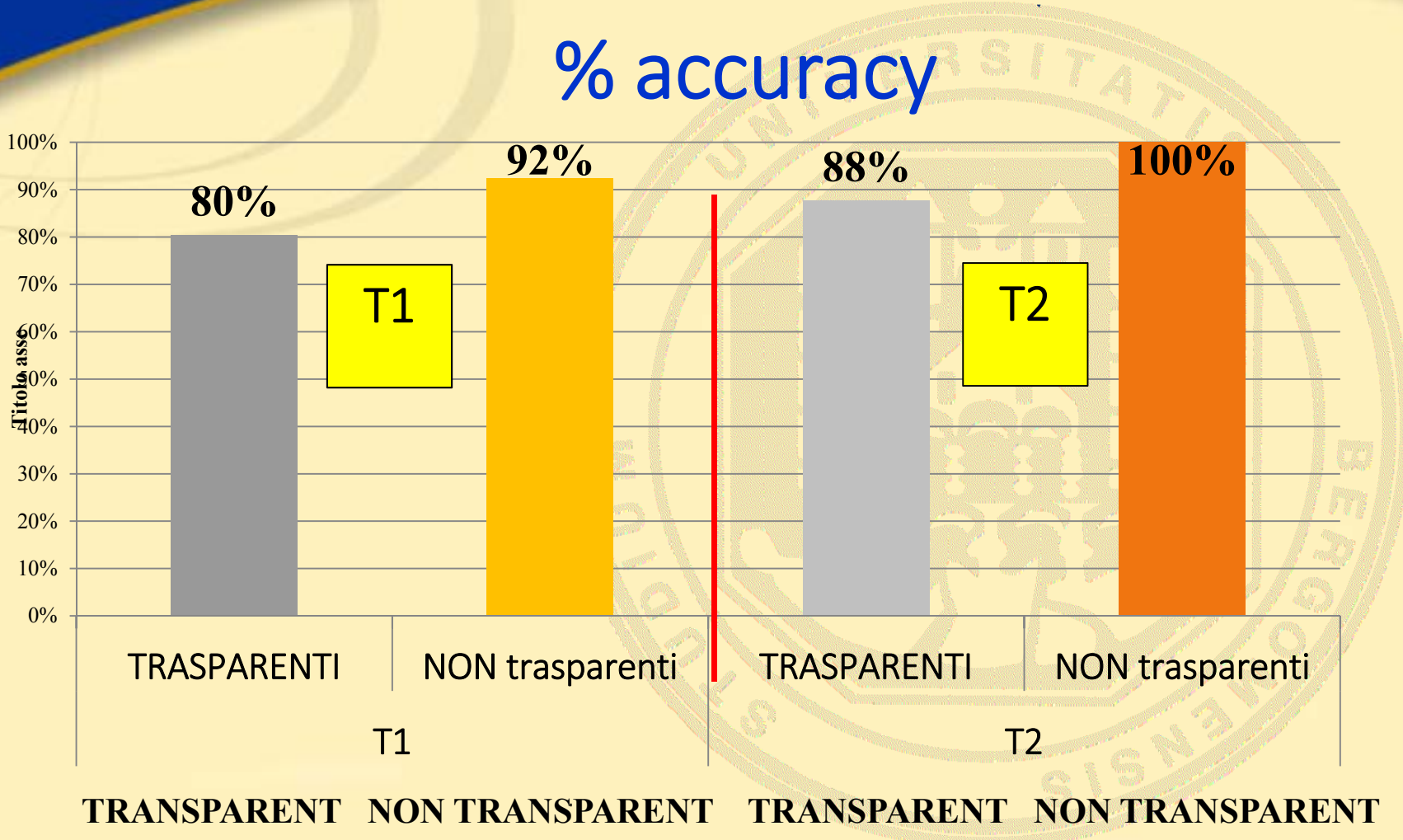
Accuracy at T1 vs. T2



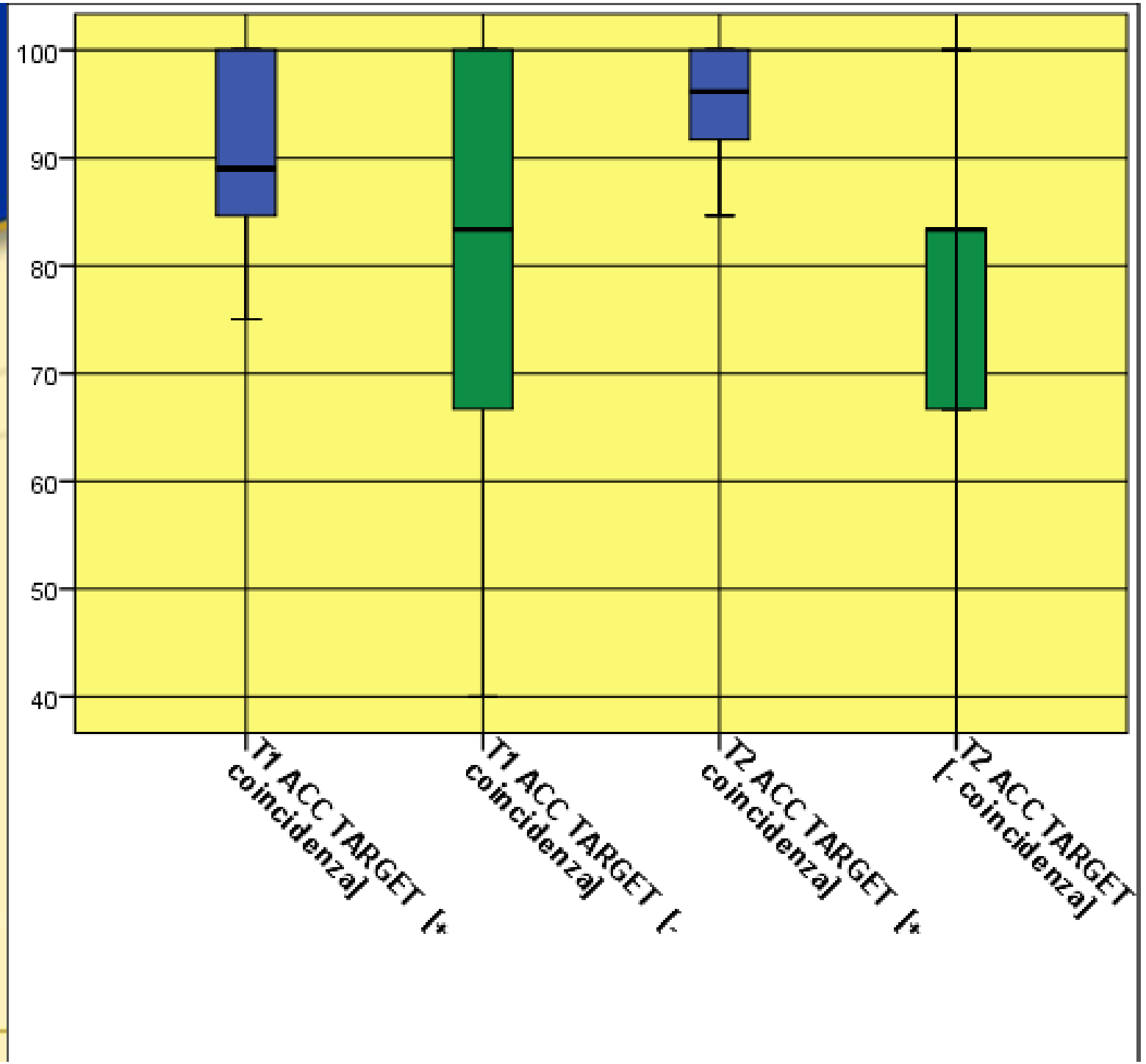
High- vs. Zero-frequency Items



trasparent vs. non transparent items



Paired
t-test
(α at .05)



Concluding remarks

- Learnability of fixed word stress
- Impact of +/-stress coincidence on recognition of so called «cognates»
- Opportunity to consider stress coincidence and stressed syllable phonological identity when:
  - developing tests/measuring lexical competence
  - L2 course syllabus structuring, exspecially in courses for beginners

VILLA project

A special thank to the members of the VILLA Project who designed the Translation Test and the Picture Production test

- Marzena Wątorek, CNRS and University Paris 8
- Rebekah Rast, CNRs and the American University of Paris
- Agnieszka Latos, CNRS

Thanks for your attention!

email: ada.valentini@unibg.it

roberta.grassi@unibg.it

Bibliography

- Bardel, C./ Gudmundson, A./Lindqvist, C., 2012, "Aspects of lexical sophistication in advanced learners' oral production. Vocabulary acquisition and use in L2 French and Italian" *Studies in Second Language Acquisition* 34: 269-290.
- Dimroth, C./Rast, R./Starren, M./Wątorek, M., 2013, "Methods for studying the learning of a new language under controlled input conditions: The VILLA project". EUROSLA Yearbook 13: *EUROSLA Yearbook* 13, 109-38. Amsterdam, Benjamins.
- Ferreri, S., 2005, *L'alfabetizzazione lessicale. Studi di linguistica educativ*a, Roma, Aracne.
- Ford, M.A./Maslen-Wilson, W./ Davies, M.H., 2003, "Morphology and frequency: Contrasting methodologies". In Baayen H., Schreuder R. (eds.) *Morphological structure in language processing*. Berlin, De Gruyter, 89-124.
- Goedemans, R./van der Hulst, H., 2011, "Fixed stress locations" In: Dryer, M.S./ Haspelmath, M. (eds.) *The World Atlas of Language Structures Online*, Munich, Max Planck Digital Library, cap. 14 (http://wals.info/chapter/14).
- Laufer, B./Nation, P., 1995, "Vocabulary size and use: Lexical richness in L2 written production" *Applied Linguistics* 16, 307-322.
- Moeller, R., 2011, "Wann sind Kognaten erkennbar? Ähnlichkeit und synchrone Transparenz von Kognatenbeziehungen in der germanischen Interkomprehension". *Linguistik online* 46, 2/11, 79-101.
- Rast, R., 2008, Foreign Language Input: Initial Processing, Bristol, Multilingual Matters.
- Rubach, J./ Booij, G., 1990, "Syllable structure assignment in Polish". In: Ewan, C.J./ Kaisse, E.M. (eds.) *Phonology 7*, Cambridge, CUP, 121-158.
- Tafel, K. et al., 2009, *Slavische Intercomprension*. Tuebingen, Narr Verlag.