



UNIVERSITY OF BERGAMO

School of Doctoral Studies

Doctoral Degree in Linguistics

XXIX Cycle

SSD: L-LIN/01

Joint PhD program with Université Paris VIII

**Utterance structure in the initial stages of Polish L2
acquisition: from semantics to case morphology**

Advisor

Chiar.mo Prof. Giuliano Bernini, Chiar.ma Prof.ssa Marzena Watorek

Doctoral Thesis

Jacopo SATURNO

Student ID 1013036

Academic year 2015/16

TABLE OF CONTENTS

FOREWORD.....	3
1. INTRODUCTION	6
1.1. Object of the study: Polish morphosyntax	6
1.2. Input: first exposure studies	13
1.3. The VILLA project: methodology	17
1.3.1. The target language	17
1.3.2. Learner selection.....	22
1.3.3. The input	23
1.3.4. The VILLA input as a variety of Polish	24
1.3.5. Data collection.....	37
1.3.6. The VILLA project: an interim summary	38
1.4. Research questions	41
2. THE REPETITION TEST	52
2.1. Introduction	52
2.2. The Repetition test: structure.....	56
2.2.1. Transcription of the repetition test.....	60
2.3. The repetition test: results.....	61
2.3.1. Overview of learner output	61
2.3.2. Repetition test: descriptive statistics	64
2.4. Repetition of -/e/	65
2.4.1. Repetition of -/e/: inferential statistics	69
2.5. A different perspective	70
2.5.1. Guessing in the Repetition test.....	73
2.5.2. Individual strategies of input processing: descriptive statistics.....	75
2.5.3. Individual strategies of input processing: inferential statistics.....	78
2.5.4. Repetition of -/e/: a comprehensive picture	79
2.6. Repetition of -/a/	82
2.7. Repetition test: discussion.....	88
3. THE COMPREHENSION TEST	97
3.1. The Comprehension test: structure	97
3.1.1. Comprehension test: target items	98
3.2. Comprehension test: results	100
3.2.1. Descriptive statistics	100
3.3. Guessing in the Comprehension test	102
3.3.1. Comprehension test: learners below chance level accuracy	103
3.3.2. Comprehension test: learners at chance level.....	106
3.3.3. Comprehension test: learners above chance.....	108
3.3.4. Comprehension test: a comprehensive picture	111
3.3.5. Comprehension test: inferential statistics.....	114

3.4.	Comprehension test: problematic scenarios	115
3.4.1.	Comprehension test: differential processing of OS word orders..	115
3.4.2.	Comprehension test: errors in the processing of SO targets	123
3.5.	Comprehension test: an interim summary	125
4.	AN OVERALL PICTURE OF MORPHOSYNTACTIC COMPETENCE.....	129
4.1.	The comprehension test as a disambiguator to the repetition test.....	129
4.2.	Correlating the repetition and comprehension tests: methodology ...	132
4.3.	Scenarios	133
4.3.1.	Effects of additional exposure to the input.....	143
4.1.	Inferential statistics	148
4.2.	Repetition in the absence of comprehension	149
4.3.	Summary	157
5.	LEARNER SEMI-SPONTANEOUS OUTPUT IN INTERACTION	161
5.1.	Introduction	161
5.2.	Methodology	166
5.3.	An overview of learner output	169
5.3.1.	Lexemes	172
5.4.	Same-word utterances	180
5.5.	Evaluating a syntactic principle of utterance organisation	183
5.6.	Principles of utterance formation and interpretation.....	186
5.7.	Correlating the structured tests with semi-spontaneous output.....	189
5.7.1.	Discussion	191
5.8.	Conclusion.....	195
6.	INPUT	198
6.1.	The VILLA methodology: input control.....	202
6.2.	Studying frequency effects in the VILLA project.....	206
6.2.1.	Form-function association	206
6.2.2.	Type frequency and input skewedness	213
6.2.3.	Token frequency	223
6.3.	Test target structures.....	231
6.3.1.	Repetition test.....	231
6.3.2.	Comprehension test input patterns.....	237
6.4.	Conclusion.....	241
7.	CONCLUSION	246
7.1.	An overall summary	246
7.2.	A provisional balance.....	254
	PRONUNCIATION GUIDE	257
	TRANSITIVE STRUCTURES IN THE INPUT - PATTERNS	260
	REFERENCES	265

FOREWORD

When I first joined the VILLA team, everything was ready to start. The input had been planned, the test designed, the schedule prepared in the greatest detail. My task at that time was to adopt this complex machinery to the two Italian editions, which were to conclude the data collection stage of the experiment. I can claim no credit as to the design and the organisation of this project, on which my whole research activity since then has been founded. Instead, I have a profound debt of gratitude to the large and enthusiastic team who made this work possible: Cecilia Andorno, Giuliano Bernini, Marina Chini, Christine Dimroth, Roberta Grassi, Henriette Hendricks, Heather Hilton, Johanna Hinz, Agnieszka Latos, Emma Marsden, Urszula Paprocka, Sebastian Piotrowski, Rebekah Rast, Leah Roberts, Ellie Shoemaker, Marianne Starren, Ada Valentini and Marzena Watorek, in addition to all the technicians and the student assistants who helped with collecting the data. And of course, Monika Krzempek, the Polish teacher, who for almost a year incessantly travelled across Europe delivering remarkably consistent input. Finally, special thanks go to the many learners who lent their time as well as their learner varieties to the project, trusting that the research team would make the best of it for the benefit of our understanding of the language faculty, which, although camouflaged as hard science, seems to me as the ultimate goal of humanism.

VILLA has always felt like a team, and accordingly this thesis was shaped by the contributions of all its members. Of course, I feel particularly indebted to my advisors, who guided me from my first, uncertain steps in research all the way to here. First, Giuliano Bernini: my interest in linguistics has a clear starting date which is unforgettably written in my mind. That is the day in the Autumn of 2009 when I first walked into his classes wondering what its mysterious title, "Linguistic Typology", might possibly refer to, and an hour later walked out of it knowing it was what I wanted to do. Not quite typology, it turned out, but it is then I learnt the spirit. Over the years of my MA and Phd

I received from him countless opportunities and a rock-solid foundation, as well as a certain way of thinking. All this work has been inspired by the wish to be up to it.

Marzena Watorek started inviting me to Paris to discuss an interesting point concerning the repetition test, just shortly after I began working on the VILLA data. Eventually she became my supervisor within a Co-tutelle agreement between the University of Bergamo and the University of Paris VIII. Among the most precious lessons I learnt from her is an almost personal relation with the data, along with a certain taste for complexity. This would prove precious to look at the same phenomenon from multiple perspectives in search of a coherent, overall picture.

Other people played a crucial role in the development of this study, too, although they may be largely unaware of that. Statistics have become an indispensable tool in applied linguistics studies, and accordingly my work makes ample use of it, though I only tried to use it as a tool for verifying things I should already know. Whenever possible, I always preferred to plot the data and let them speak for themselves. Although I performed all the analysis on my own and take full responsibility for both the methodology and the results, it is only thanks to the help of others that I can now say that. Sincere thanks to Roger Mundry, Marie Durand and Philippe Bonnet.

I am thankful to my family for supporting and, indeed, enduring me, just as I'm sorry I taught them to recognise the signs of the many failures I went through over these years.

Finally, although I always tried to do my best, I cannot forget that I have been privileged by incredible luck. It started in the middle of my MA degree, when as a student of Slavistics I decided out of no special reason to go on an Erasmus study stay abroad. The only (theoretically) Russian-speaking country available was Estonia, so I applied for it. In the same days, almost by chance, Slavic Linguistics professor Andrea Trovesi asked me after an exam why not go to Poland, instead, and learn another Slavic language. I thought "why not?", too, and that's how I changed my plans and first met Poland and its language. That

was the turning point in many respects. There I met my future wife, my reference native speaker, my Polish teacher for years, my safe haven. I took a first exposure course in Polish, as soon as I arrived, so that later on I could understand the VILLA learners better than they probably thought. I also discovered I really liked the language and started learning it avidly. When I came back to Bergamo, just in time to earn my MA degree, I found out that a new project was starting for which a person with a little Polish and some linguistics knowledge was needed. I had no idea about it when I left, I did not prepare for it, I can claim no credit. For this too I am ever so grateful.

1. INTRODUCTION

This work is devoted to the processing of morphosyntax in the initial stages of the acquisition of Polish¹ L2, with a focus on the role of input. The introductory chapter is mainly devoted to clarifying the terms contained in the preceding sentence: processing of morphosyntax, initial stages, Polish, input. In the next few pages, we will discuss the general methodological issues connected with our study, briefly review the relevant literature, and finally formulate general research questions useful for framing the more specific chapters that follow. The purpose of this section is to equip the reader with the basic tools necessary to proceed through the rest of this work with the greatest ease allowed by the matter at hand.

1.1. OBJECT OF THE STUDY: POLISH MORPHOSYNTAX

Morphosyntax is concerned with the way in which words change their form depending on the speaker intentions, the functions they carry within a clause and the presence of other words within or outside the same clause. Typical examples include verb inflection, case marking and agreement. We call *inflection* the very fact that words may take on different forms, as well as the sounds which distinguish them; *word-forms*, or *inflected forms* the various results of this process; and *paradigm* the inventory of possible forms for a given word. The size and complexity of a paradigm determines whether a language is more or less *morphologically complex*. In a language like Polish, paradigms are determined by the interaction of a number of semantic and grammatical categories, including gender, number, animacy, and case. This is the category which is of interest to us. Polish nouns, pronouns, adjectives and some numeral inflect for as many as six cases, namely nominative (NOM), genitive (GEN),

¹ Polish, belonging to the west-Slavic group, is the official language of Poland, spoken by

dative (DAT), accusative (ACC), instrumental (INS) and locative (LOC). We will treat what is generally called *grammatical, relational, core, abstract or functional* case, depending on the author and framework of reference (Haspelmath 2008). This expresses a core syntactic relation such as subject or object, assigned by the specific structural configuration, in contrast to the more specific semantic roles such as space relations expressed by what is called *semantic, adverbial, concrete, peripheral, or inherent* case. In this work we will be concerned with the expression of the core syntactic functions of subject (SUBJ) and object (OBJ), which in Polish correspond to the nominative (NOM) and the accusative (ACC) case respectively. Consider the sentence in (1):

- (1) *Filip* *ma* *kot-a*
 Filip.NOM has cat-ACC
 "Filip has a cat"

The word-forms *Filip* and *kota* are only two out of the many in which lexemes belonging to the same inflectional class or paradigm may occur depending on the syntactic function or semantic role performed in the sentence². Table 1 exemplifies the singular (SING) and plural (PLUR) paradigm of a masculine (M) and feminine (F) noun.

² Our use of the terms "syntactic functions", "subject" and "object" do not imply any reference to specific theoretical frameworks. Indeed, other categories such as the thematic roles "agent" and "patient" would also account for our data adequately. In fact, since the VILLA input does not include all the complexities of native varieties of Polish, the subject always coincides with the agent, and the object with the patient: departures from this patterns, such as passive structures, are not attested. In practical terms, then, the two sets of categories are equivalent. We will consistently adhere to "subject" and "object" for mere reasons of simplicity.

Table 1: example paradigms of Polish nouns

	SING		PLUR	
	M	F	M	F
NOM	<i>Strażak</i>	<i>kuchark-a</i>	<i>strażac-y</i>	<i>kuchark-i</i>
GEN	<i>strażak-a</i>	<i>kuchark-i</i>	<i>strażak-ów</i>	<i>kucharek-∅</i>
DAT	<i>strażak-owi</i>	<i>kucharc-e</i>	<i>strażak-om</i>	<i>kuchark-om</i>
ACC	<i>strażak-a</i>	<i>kuchark-ę</i>	<i>strażak-ów</i>	<i>kuchark-i</i>
INS	<i>strażaki-em</i>	<i>kuchark-ą</i>	<i>strażak-ami</i>	<i>kuchark-ami</i>
LOC	<i>strażak-u</i>	<i>kucharc-e</i>	<i>strażak-ach</i>	<i>kuchark-ach</i>
VOC	<i>strażak-u</i>	<i>kuchark-o</i>	<i>strażac-y</i>	<i>kuchark-i</i>
	"fireman"	"cook.F"	"firemen"	"cooks.F"

Thanks to case inflection, which links syntactic functions to the words performing them independently of their position in the utterance, Polish word order is theoretically free and can be manipulated for pragmatic purposes. Since in terms of information structure the utterance-initial position is usually reserved for the topic, the speaker may choose to topicalise the subject, as is the norm in unmarked sentences (1), or another constituent, such as the object. In the latter case, the subject will occupy the focus position, producing a syntactically marked OS structure with the same referential meaning as its SO alternative, but a different information structure (2).

- (2) *kot-a* *ma* *Filip*
 cat-ACC has Filip.NOM
 "Filip³ has a cat" or "it is Filip who has a cat"

Such complex morphological systems are well known to cause serious difficulties to language learners (see however Kempe & MacWhinney 1998 for a discussion). Throughout this study, we will refer to "morphosyntactic

³ Underlining indicates emphatic stress.

competence⁴ as to the learners' ability to manage case endings in order to express syntactic functions independently of word order, so as to veiculate pragmatically marked meaning if required. SLA studies in the acquisition of "morphologically complex" systems abound, though they are not necessarily comparable in terms of their object. It may not be legitimate, for instance, to parallel the acquisition of nominal and verbal inflectional systems: and even within the same sub-system, while it is clear that, say, Spanish, German and Russian are morphologically more complex than English or Chinese, still they are rather different from each other too. For these reasons, in the following bibliographical review we will only select the works concerning the acquisition of Polish, which, however, has been mainly treated as the object of L1 acquisition. Łuczyński (2002; 2004; 2010), for example, studied the development of nominal forms in the speech of two children aged two to three, that is, at the age in which the first morphological oppositions appear. The author shows that paradigms start off with three forms, namely nominative, accusative and vocative. However, because of frequent instances of syncretism, some of the corresponding functions are performed by the same form, e.g. *dom*, home.NOM/ACC as opposed to *chłopak*, boy.NOM vs. *chłopak-a*, boy-ACC. Similar findings will emerge from our own study of the VILLA input (chapter 6), highlighting the powerful link between semantics and syntactic functions. Smoczyńska (1972; 1985; 1997) observes that the first recognisable noun forms produced by young children are modelled on the nominative case. Later on a new phase kicks in, in which words appear in two forms, one of which is modelled on the nominative and the other simply contrasts with it. Similar observations were made for L2 acquisition (see below), and find a partial parallel in our own study as well. Further, Dziubalska-Kołączyk (1997), following Dressler & Karpf (1994), applies the terms *pre-* and *proto-morphology* to the acquisition of Polish L1. During the first stage, basic morphological operations, such as reduplication, are experimented by the young learner.

⁴ Our use of the term "competence" should not be interpreted as a reference to the opposition between competence and performance.

Proto-morphology marks the beginning of the morphological system of the language according to the principles of Natural Morphology (Dressler 1985; 1987; 2011; Wurzel 1989; Crocco Galeas 1998). The phase of morphology proper, finally, entails the final development and completion of the inflectional and derivation systems of the target language.

Dąbrowska (2004; 2005; 2006; Dąbrowska & Szczerbinski 2006; Dąbrowska & Tomasello 2008) analyses the development of the Polish case system arguing for usage-based strategies as opposed to the construction of abstract rules: due to the density of their neighbourhood (group of words sharing a certain feature), diminutive endings offered a particularly fertile research ground. The productivity of case endings is seen as the product of an interaction between phonological heterogeneity and type frequency, whereas regularity proved a poor predictor. The same methodology produced similar results when applied to adult speakers (Dąbrowska 2008). These findings will sound familiar when we turn to discuss form-function associations between case endings and syntactic functions in the VILLA input (section 6.2.1, p. 206).

In a similar vein, Olma (2007) focussed on the later stages of the acquisition of Polish inflectional morphology by studying the speech of a young girl. His work highlights the powerful effect of frequency on the order and rapidity with which different forms develop. Similar conclusions are drawn by Machowska (2006), who argues that while by the first year of school the acquisition of the case system is roughly complete, some forms of dative, instrumental and vocative still represent a source of difficulty for pupils due to their scarce frequency.

Finally, Krajewski (2005) shows that grammatical gender exists in the inflectional system from the very earliest stages of acquisition, although different children may develop different classifications. Given that gender is one of the factors which determines what paradigm nouns belong to, this finding is of significant importance for our discussion of the comprehension test (chapter 3).

Concerning the acquisition of Polish L2, one gets the impression that while conspicuous research addresses questions of teaching methodology, the actual data on acquisitional data are scarce, at least to our knowledge. In fact, this is not surprising, as Poland was not a destination of immigration until very recent times, and even then, on a very small scale. We are aware of a few works devoted to phonology (Smoczyński 1965; Hentschel 1990; Majewska-Tworek 2005; Magajewska 2008) and to cross-linguistic interference, with a specific focus on other Slavic languages (Pösingerová 1997; Ananieva 1997; Szapkina 1997; Pančiková 1997). Besides that, most research concerning Polish L2 was conducted within the VILLA project and will be reviewed in the appropriate section (1.3.6, p. 38).

Acquisitional data are much more common for Russian, an east-Slavic language which shares numerous structural characteristics with Polish. A systematic review of the research concerning the acquisition of this language is beyond the scope of our work: the interested reader is referred to Romanova and Gor (2016) for a recent review. However, an exception is in order here to report of a couple of studies which are directly relevant for our work. First, Kempe & MacWhinney (1998) compared the acquisition of the Russian L2 and German L2 case systems. They observed that while the Russian system is undoubtedly more complex, it is also mastered more rapidly by L2 learners: this is because its word-forms are also more univocal as to the case they correspond to, so that it is easier to identify the relations between form and meaning. In German, in contrast, wide-spread syncretism and lack of overt case marking make numerous inflectional types ambiguous as to their grammatical meaning. This study will be cited again and partially replicated in chapter 6, p. 198.

Other relevant studies were conducted within the framework of Processability Theory (Pienemann 1998). All show quite clearly that accusative case marking first emerges in syntactically unmarked SVO sentences: it can be argued, therefore, that it is linked not to the object function, but rather to the post-verbal position within the utterance: see the Russian L2 example in (3a), from Artoni & Magnani (2015:190). Indeed, this tendency has proved so widespread

as to generalise to contexts which do not require accusative marking, like in the German L2 examples provided in (3b) and (3c), from Baten (2011:490) and Diehl et al. (2000:235), respectively. Note that in the last example inflection is only visible on the determiner, which further witnesses to the difference in the target structures referred to by the same name in the literature.

- (3) a. *vilk-a prinës balerin-u*
 fork-NOM brought dancer-ACC
 "the dancer brought (the) fork"
- b. *nicht weit von hier befindet sich den Bahnhof*
 Not far from here find itself the-ACC
 station
 "the station is not far from here"
- c. *es ist ein-en Aprilfisch*
 it is an-ACC April's fool
 "it is an April's fool"

Only at more advanced developmental stages do learners acquire the ability to correctly case-mark the object constituent in syntactically marked structures like OVS, so as to manipulate word order for pragmatic purposes, if required. Regarding the development of inflectional paradigms, some authors signal a phase of non-nominative marking, in which a sort of mini-paradigm (Bittner, Dressler & Kilani-Schoch 2000) develops with only two forms: a basic one, typically modelled on the nominative case, and a marked, or non-basic one, as shown once again in examples from Slavic languages, specifically Russian L2 (4a: Artoni & Magnani 2015:188) and Serbian as a heritage language (4b: Di Biase, Bettoni & Medojević 2015).

- (4) a. *videla* *volk-e*
 saw-FEM wolf-NON.NOM
 "(she) saw a wolf"
- b. *onak su* *videli krevet-a*
 then have seen bed-NON.NOM
 "then (they) saw a bed"

This non-nominative form may be modelled on various target case endings, and is not necessarily produced consistently or systematically. It does, however, show that learners have at least noticed the morphological variability of the target and are trying to make sense of it, something which we will frequently observe in our work too.

1.2. INPUT: FIRST EXPOSURE STUDIES

We now turn to one of the main themes of our work, namely input. Its role is a crucial theme in Second Language Acquisition (SLA) studies, as it intuitively represents the main source of information allowing learners to structure and perfect their developing L2 system. Although this notion has been problematised and discussed at length in a number of works, starting with Krashen (1985) and then Gass (1997), Carroll (1999; 2001; 2005) Piske and Young-Scholten (2008) and many others, by input we generally mean any bit of the target language that learners are exposed to through any channel.

While everyone roughly agrees that input should play some role in SLA, the way in which it is processed and the factors that affect acquisition success the most are still heatedly debated questions. This is partly because input is extremely difficult to control in an experimental framework, as everyone's learning experience is to a certain extent unique. Consequently, the doubt remains that different acquisition outcomes may simply derive from input that differs in quantity or quality, rather than from the systematic, predictable

effects of such variables as frequency, transparency or salience, however operationalised.

Studies conducted with a longitudinal design (e.g. Perdue, 1993; Giacalone Ramat, 2003) typically resign from experimental control and rely on the assumption that learners will be exposed to "average" input, especially over sufficiently long periods of time. This approach, while methodologically the simplest, inevitably renders it problematic, if not impossible, to correlate learner output with the input received in a rigorous way. In a variant to this methodology, other researchers, like Collins et al. (2009), based their analysis on samples of input extracted from a relatively homogeneous context (in that case, classroom L2 instruction) and then generalised the findings to the rest of the input. While this might work if samples are collected with sufficient frequency and duration, it may cause significant distortion in the event of shorter experiments.

Moreover, if this approach may enjoy some ecological validity in the case of learners of intermediate and upper levels, it becomes problematic when attention is turned to another sensitive topic in SLA research, namely the very initial stages of L2 acquisition. In a rather circular manner, studying the strategies of input processing applied by learners confronted with a completely novel language may be illuminating with regard to the general mechanisms of input processing (Rast et al., 2011; Perdue, 2002); in order to do this, however, it is essential to have full control over the input, so as to correlate learner output with the relevant input parameters. Full control, in turn, derives from at least two requisites. First, the researcher input should be able to fully manipulate input, so as to formulate and pursue appropriate research questions. Second, the target language should be completely unknown to the learner, so that the experimental manipulation of the desired variables can highlight the effects of the input provided, and that alone: working with learners other than absolute beginners would imply that acquisitional results may be due either to the experimental input, or to their previous experience with the target language. Of course, participants may differ in other respects as well, such as

psychometric characteristics and encyclopedic knowledge: while some variability is unavoidable, it can be kept in check by an accurate selection of participants. This second requirement represents the rationale of first exposure studies, experiments in which learners are exposed to a completely new language for the first time in their life.

In short, we could identify the major challenges of this field in finding true beginner learners, on the one hand, and controlling for the input, on the other hand. Various methodologies have been developed to try and meet these objectives.

One possibility is to employ recorded speech in an "exotic" language, as indeed was done in a series of studies by Gullberg et al. (2010; 2012; Ristin-Kaufmann & Gullberg 2014). While this approach is ideal to make sure that different groups of learners are exposed to exactly the same input, the resulting language learning experience is hardly comparable to any real communicative situation.

Further, artificial languages have attracted numerous researchers thanks to the possibility they offer to perfectly tune the target language to the desired research questions. In addition, it is clear that no learner could ever have any experience of such *ad hoc* varieties. For these reasons, numerous studies have made this choice (Reber 1967; Hulstijn 1997; Mueller 2006; Robinson 2010; Amato & MacDonald 2010; De Diego-Balaguer & Lopez-Barroso 2010; Endress & Bonatti 2007; Folia et al. 2010; Onnis, Waterfall & Edelman 2008; Saffran, Newport & Aslin 1996; Williams 2010; Zhang & Lantolf 2015). However, artificial languages typically lack the complexity and idiosyncrasies of natural languages, so that the ecological validity of such studies may be questioned.

In addition to the difficulty related to meeting the aforementioned objectives, two other typical challenges of this field should be mentioned. Firstly, as the means available to learners are usually too limited to produce any output, it is typically hard to analyse, or indeed even elicit utterances from initial learners (Grüter, Lieberman & Gualmini 2008). As a result, some studies on "early" L2 acquisition in fact refer to relatively developed learner varieties. For the same

reason, several studies concerned with the initial stages of acquisition have analysed learners' skills through receptive tests or highly structured tasks, rather than by observing actual learner output in a semi-communicative context. Unfortunately, this approach hardly allows researchers to look into the structure and development of early learner varieties in the qualitative fashion typical of longitudinal studies. Another methodological difficulty regards the fact that since most of the languages which are typically studied in this field of research are relatively common, it is usually hard to find learners who have never had any exposure to the target language, however minimal.

And finally, typical laboratory experiments only target a very short time of acquisition. Most of the time this is precisely the point: to show what little exposure learners require in order to acquire the target structure. Boyd et al. (2009), for instance, exposed adult participants to a novel syntactic construction and then found that they were successful in reproducing target regularities in their own output after only 3 minutes. Similarly, the studies by Gullberg and colleagues cited above show the promising, though obviously not conclusive results achieved by L2 learners after 7 minutes of exposure to a video. Further, Kittleson et al. (2010) demonstrate the ability of Norwegian L2 learners to segment running speech after only a few minutes of exposure. The review presented by De Diego-Balaguer and Lopez-Barroso (2010) also reports of successful speech segmentation within minutes of exposure to a novel L2 based on prosody. Positive results also come in the domain of morphosyntax. In a study on aptitude, VanPatten and Smith (2015) examine the acquisition of head-final word order and case marking in Japanese L2 by initial learners and found that indeed they proved sensitive to case-marking violations. Finally, Carroll and Widjaja (2013) show that adult English speakers exposed for the first time to Indonesian were able to learn all three constructions in which the L2 expresses number, namely bare noun phrases, reduplication and numeral + classifier constructions. To complete this necessarily inexhaustive review, the interested reader is referred to a recent book edited by Han and Rast (2014), which witnesses to the growing interest in this field of studies.

For our purposes, it will suffice to summarise this section as follows: it would be highly desirable to investigate the development of an L2 from its very onset, with a natural target language and whilst retaining full control over the input, ideally for a time span sufficient for observing the development of the L2 after the very first contact with the target language. Precisely such is the ambitious objective of the VILLA project, of which this work is part. Truly, some precedents exist: Rast's (2008) work anticipates most of the rationale and methodology of what would later become VILLA. In the next paragraph, we provide a brief description of this ground-breaking experiment.

1.3. THE VILLA PROJECT: METHODOLOGY⁵

As we have anticipated above, first exposure studies rest on two crucial requirements, namely a) that learners were never exposed to the target language, and b) that the input provided to all participants is identical in quantity and quality, as well as experimentally controlled or manipulated. In the following sections we will discuss how these requirements are met in the VILLA project.

1.3.1. The target language

Regarding the first point, we have already discussed the implications of the choice of the target language. The VILLA project chose to focus on a natural language presented orally, sufficiently rare in the countries where the experiment took place as to represent a completely novel target. In addition, this language had to possess features which could distinguish it from the native languages of the project and allow for appropriate research questions. The experiment took place in five countries, namely Germany, the Netherlands,

⁵ The description of the project provided here is strictly functional to the purposes of our work. For a more detailed and comprehensive overview of the experiment, see Dimroth et al. (2013)

France, the UK and Italy. On the basis of the characteristics of the languages of these countries with respect to morphosyntax, as well as to its scarce diffusion in those five countries, Polish was identified as the ideal candidate.

The L1 closest to Polish is certainly German, the only one to maintain a case system on nouns⁶. This allows for the manipulation of word order for pragmatic reasons, just as in Polish.

In German, however, case is mainly signalled on the determiner, while the inflectional paradigm is characterised by diffused syncretism, especially as far as feminine nouns are concerned (Table 2).

Table 2: German nominal inflection

		SING	
SING	NOM	<i>der Feuerwehrmann</i>	<i>die Köch-e</i>
	GEN	<i>des Feuerwehrmanns</i>	<i>der Köch-e</i>
	DAT	<i>dem Feuerwehrmann</i>	<i>den Köch-en</i>
	ACC	<i>den Feuerwehrmann</i>	<i>die Köch-e</i>
		"the fireman"	"the cook.F"
PLUR	NOM	<i>die Feuerwehrmänner</i>	<i>die Köchinn-en</i>
	GEN	<i>der Feuerwehrmänner</i>	<i>der Köchinn-en</i>
	DAT	<i>der Feuerwehrmännern</i>	<i>den Köchinn-en</i>
	ACC	<i>die Feuerwehrmänner</i>	<i>die Köchinn-en</i>
		"the firemen"	"the cooks.F"

There are some further notable differences between the two languages. In spite of determiners, the German paradigm still suffers from widespread syncretism and ambiguity. As a result, transitive sentences not involving masculine nouns are ambiguous as to subject and object assignment (5).

⁶ All the languages considered here distinguish case on personal pronouns.

- (5) *Das Tier, das die Katze jagt.*⁷
 the.NOM/ACC animal.NOM/ACC, which.NOM/ACC the.NOM/ACC
 cat.NOM/ACC chases
 "The animal which the cat chases" or "the animal which chases the cat."

In Polish, in contrast, the two alternative interpretations of the same German sentence are clearly distinguished by nominal inflection (6). Ambiguity would be further reduced if the first noun distinguished the nominative from the accusative word-form, as commonly happens in numerous Polish paradigms.

- (6) a. *zwierzę, które goni kotk-a.*
 animal.NOM/ACC which.NOM/ACC chases cat-NOM
 "the animal, which the cat chases"
- b. *zwierzę, które goni kotkę.*
 animal .NOM/ACC which.NOM/ACC chases cat-ACC
 "the animal, which chases the cat "

While these examples are not directly relevant to our work, as they involve neuter nouns and relative structures, both rare or absent in the VILLA input, they certainly make it clear that Polish has greater potential to manipulate word order than German does. In fact, morphosyntax alone would suffice to veiculate grammatical meaning, even in the absence of marked intonational profiles or other non-syntactic means to signal departures from the unmarked SO word order. This possibility was widely exploited in the structured tests described in chapters 2 and 3, perhaps to the extent that the situation recreated by the test lacks some ecological validity. Because of the potential ambiguity deriving from its syncretic nominal morphology, in contrast, German is more constrained with respect to pragmatics and discourse. Of course, it could be argued that Polish is sensitive to the same constraints, too, and that

⁷ Example from Rankin (2014:204).

consequently it will not necessarily exploit its theoretical potential to the full. By the same token, marked intonational profiles are commonly used as an accessory means to syntax in order to signal pragmatically marked meanings. The four remaining L1s, namely Dutch (Table 3), Italian (Table 4), French (Table 5) and English (Table 6) lack case altogether as far as nouns are concerned: in fact, these only present two forms, corresponding to singular and plural. Consequently, syntactic functions are assigned on the basis of the default SO word order. Departures from this pattern are possible, but require marked syntactic or phonological means, such as left and right dislocations, cleft sentences, and particular intonational contours. Note that all these possibilities are also available in languages with more complex morphology, like Polish and German.

Table 3: Dutch nominal inflection

SING		PLUR	
M	F	M	F
<i>de kok</i>	<i>de kokkin</i>	<i>de kok-s</i>	<i>de kokkin-s, kokkinn-en</i>
"the fireman"	"the cook.F"	"the firemen"	"the cooks.F"

Table 4: Italian nominal inflection

SING		PLUR	
F	M	F	M
<i>il pompier-e</i>	<i>la cuoca</i>	<i>i pompier-i</i>	<i>le cuoche</i>
"the fireman"	"the cook.F"	"the firemen"	"the cooks.F"

The cases of French and English are particularly extreme. Regarding the former, even the singular and plural forms of nouns are only distinguishable in the written variety, as they are completely homophonous in speech (Table 5). Again, the expression of this opposition relies on determiners.

Table 5: French nominal inflection

SING		PLUR	
M	F	M	F
<i>le pompier</i>	<i>la cuisinière</i>	<i>les pompiers</i>	<i>les cuisinières</i>
\lə pɔ̃.pje\	\la kɥi.zi.njɛʁ\	\le pɔ̃.pje\	\le kɥi.zi.njɛʁ\
"the fireman"	"the cook.F"	"the firemen"	"the cooks.F"

English mostly lacks grammatical gender on nouns, so that nouns only distinguish a singular and a plural form, while the determiner is invariable. Even when a noun is characterised in terms of intrinsic sex, the semantic equivalent of grammatical gender, this category is only visible through anaphoric reference.

Table 6: English nominal inflection

SING		PLUR	
the fireman	the cook	the firemen	the cook-s

Finally, we note that in contrast to Polish, all the L1s of the VILLA project have articles (Dryer 2013a).

Other exotic traits of Polish include its vocabulary, which is mostly of Slavic origin and thus highly opaque for the speakers of the L1s involved in the VILLA project. Lexical stress is fixed on the penultimate syllable, with the only exception of learned loanwords from Latin or Greek as well as elements to which clitics are attached⁸. Such rigid pattern is not found as such in any L1, although Italian nouns are also predominantly accented on the penultimate syllable and French does exhibit fixed stress, though on the last syllable of the intonation unit.

⁸ Note that even in these cases there is an increasing tendency to normalisation based on the dominant pattern.

1.3.2. Learner selection

Choosing an exotic language is only one of the steps required to build a first exposure study. It is also necessary to make sure that the learners never had any experience of it. To this purpose, learner selection in the VILLA project was a long and accurate process which took place in several steps. First, candidates were asked to fill in a questionnaire regarding their linguistic repertoire: anybody who had been exposed to Slavic languages was excluded at this stage. Whenever possible, learners who had studied morphologically complex languages such as Greek, Latin or even German were also excluded. The reason for this is that a linguistically naïve speaker of any of the VILLA L1s, with the sole exclusion of German native speakers, should not be aware of what grammatical case is and how it works. The explicit study of classical languages, in contrast, inevitably implies some familiarity with this category. There was the risk, in sum, that learners with that kind of experience could process Polish morphosyntax not just on the basis of the input provided during the course, as foreseen by the project research questions, but rather thanks to their explicit meta-linguistic competence. Hence the rationale for excluding this group of learners⁹.

The people who passed this stage took a "language sensitivity test", in which they heard languages in three different languages, namely Polish, Russian and Finnish, and were asked whether or not they thought the sentences were in Polish. This was done in order to make sure that somebody's "intuition" may be just too good and thus bias the results of the experiment.

Finally, all learners thus selected were asked to sign a contract stating that they would not seek additional information on Polish before and during the experiment. In the same spirit, they were asked not to take notes during classes. The rationale behind these requests was to make sure that everybody is exposed to exactly the same amount and quality of input, which could be

⁹ This was particularly difficult in Italy, as most secondary school students take at least a year of Latin. Despite the rather positive response to the call advertising the experiment, the number and range of applicants was simply insufficient to fully satisfy this particular requirement.

greatly affected by the degree of effort payed to taking notes or studying Polish grammar at home, for instance. Obviously, it was not possible to control for learners' attention during classes, but that sounds appears even in the most controlled experimental setting. In our case, moreover, it creates a nice feeling of realism and analogy with real language teaching contexts.

The learners whose acquisitional results will be discussed in this work attended the so-called meaning-based (MB) course, as opposed to the form-based (FB) version. In the former, input was presented with no focus on form (Sharwood-Smith 1993; Doughty & Williams 1998) and no corrective feedback. As it is not relevant for our analysis we will not pursue this opposition further, but see Dimroth et al. (2013) for the details.

Finally, to quantify the proportions of our study, the number of our learners by L1 is presented in Table 7.

Table 7: learners by L1, MB group

EN	FR	GE	IT	NL	TOT
16	17	20	17	20	90

1.3.3. The input

A first exposure study should provide input in such a way that it is identical for all learners and fully controlled to suit the research questions. An easy way to do so would be to use recordings, as indeed has been done in the past, but that would result in a highly artificial environment; moreover, it would be hard to envisage interaction and to carry out the experiment for a prolonged time, so as to study the effects of input exposure over more than some minutes.

For these reasons, the VILLA input was provided in the shape of a 14-hour Polish course taught by a native speaker specifically trained for that purpose. The same teacher worked in all editions of the VILLA project, moving across Europe to teach in the universities which took part in the initiative.

Input was planned in advance, and a scheme of the course schedule was prepared for the teacher to follow in all editions. The course thus had a very precise schedule, comprising the topics to treat, the vocabulary to introduce, the activities to perform, and, crucially, the frequency with which lexical items had to occur during classes. This particular parameter was checked in real time during classes by research assistants, who had the task of keeping track of frequencies for a specific set of lexical items so as to signal to the teacher, if necessary, if a particular item occurred too often or too rarely.

An *a posteriori* analysis on the input corpus showed that frequencies are indeed comparable across the editions of the VILLA project, which is a real credit to the teacher for managing to maintain consistency over ten different courses and a time span of almost a year.

1.3.4. *The VILLA input as a variety of Polish*

We just said that the VILLA project is innovative in that, among other things, it uses a natural language as input, instead of the artificial miniature languages commonly employed in this type of studies. It would be imprecise, however, to think that the input provided by the teacher could be a representative example of native varieties. This is quite natural if one takes into account the peculiar context in which the experiment took place, including its time span which was limited to 14 hours and the research questions regarding the role of input, which could only be answered by manipulating it. The resulting language variety at times could sound a little "odd" to a native speaker of Polish.

To start, the dramatic competence gap between the teacher and the total beginner learners naturally produced the register that Ferguson (1975) originally labelled *foreigner talk*. This particular language variety is supposed to facilitate communication between native speakers and less proficient L2 learners (Wesche 1994) by simplifying input along the following lines:

- a) slow-rated, emphatic speech;
- b) short, common, generic vocabulary;
- c) simple syntactic structures;
- d) simple meaning and concepts.

Such expedients to facilitate learners' access to meaning are common to the speech typically addressed at children, or *care-giver talk*, which includes higher pitch, hyper-articulation, enhanced intonation, clear pauses, short, simple, well-formed sentences with little subordination, frequent repetitions (see Kern et al. (2014) and the extensive bibliography cited thereby).

The general tendency to simplification characteristic of foreigner talk may lead to structures which would be ungrammatical in the target language, such as the omission of obligatory grammatical morphemes or the simplification of inflectional paradigms by using a single, uninflected word-form. This would clearly be unacceptable in a foreign language class, in which the teacher is responsible for providing learners with input which should be at least grammatically correct. *Teacher talk* (Larsen-Freeman & Long 1991:134–144) therefore represents a particular variety of foreigner talk, in which simplification still takes place, but within the boundaries of grammatical correctness and with the purpose of facilitating acquisition.

This framework is useful to contextualise some of the features of the VILLA input. As mentioned, the teacher was a native speaker and a professional teacher of Polish L2. Taking into account the initial level of proficiency of the learners, she produced extremely slow and hyperarticulated speech in an effort to make input more salient, i.e. more easily perceivable and segmentable. Vocabulary and grammatical structures were also deeply conditioned by the research questions of the project. First, lexical items were selected on the basis of their lexical transparency, one of the input parameters investigated through the structured tests. The frequency of individual lexemes was also manipulated accordingly, in order to statistically test the effects of all possible combinations of frequency and transparency. We acknowledge that these two parameters may

impact somehow on the accuracy of case marking as produced by learners in transitive sentences, most probably as a consequence of the different levels of attention needed to process them; however, we will not pursue this topic analytically. The particular contents of the VILLA course, tailored to provide sufficient material to perform the tests, conditioned the choice of vocabulary, so that normally infrequent words like *strażak* "fireman" and *pożar* "fire" are important items in the VILLA input.

Second, the choice was made to focus on a limited number of target structures, whose acquisition was later probed through the tests. This caused them to often be produced with unnatural frequency. To exemplify, in a previous paper (Saturno 2015a) we calculated the difference in the use of the copula verb *jest*, "is", in the VILLA corpus and in native varieties of spoken Polish. To this purpose we considered the Italian form-based input, consisting of 18,081 words. Of these, 2,109 (ratio: 0.12) are instantiations of *jest*. In contrast, a query run with the PELCRA search engine (Pęzik 2012a) on the spoken subcorpus (Pęzik 2012b) of the Polish National Corpus (Przepiórkowski et al. 2012) resulted in a ratio of 0.00005¹⁰. The striking difference is due to the fact that copular structures were among the linguistic features specifically targeted by the VILLA course, and thus occurred much more frequently in both teacher speech and classroom interaction.

The frequency of syntactic structures directly conditions the frequency of the inflected word-forms belonging to the paradigm of a word¹¹. Table 8 presents relative frequency data for the 12 word-forms of the lexemes *polak*¹² "Pole" and *polka*¹³ "Polish woman", extracted from three different sources: the database of internet text Monco (<http://monco.frazeo.pl/index>), the oral sub-corpus of the NKJP corpus, cited above, and the entire VILLA meaning-based input at 13:30

¹⁰ The corpus numbers 1,524,696,745 occurrences, of which *jest* accounts for 24,599 tokens.

¹¹ Note that whenever the same word-form corresponds to more than one case (e.g. *Polaku*, Pole-LOC and Pole-VOC), the same frequency ratio is assigned to all of them.

¹² Inflectional paradigm (cases provided in the same order as in column 1): SING: *polak*, *polaka*, *polakowi*, *polaka*, *polakiem*, *polaku*, *polaku*; PLUR: *polacy*, *polaków*, *polakom*, *polaków*, *polakami*, *polakach*, *polacy*.

¹³ Inflectional paradigm (cases provided in the same order as in column 1): SING: *polka*, *polki*, *polce*, *polkę*, *polką*, *polce*, *polko*; PLUR: *polki*, *polek*, *polkom*, *polki*, *polkami*, *polkach*, *polki*.

hours of exposure (T2 of the structured tests considered in this work). As the VILLA course includes a number of characters, each of whom is characterised in terms of nationality and profession, *polak* and *polka* may be considered as typical examples of frequent human nouns. Unlike most of their equivalents (e.g. *niemiec*, "German" or *listonosz*, "postman"), they should also be sufficiently common in average Polish L1 discourse to turn out with reasonable frequency in the other two corpora considered here. Forms with 0 frequency are shaded in light grey.

Table 8: relative frequency of two animate nouns by word-form

		MONCO		NKJP Oral		VILLA	
	case	<i>polak</i>	<i>polka</i>	<i>Polak</i>	<i>polka</i>	<i>polak</i>	<i>polka</i>
SINGULAR	NOM	11%	0%	24%	25%	30%	28%
	GEN	5%	74%	4%	44%	0%	0%
	DAT	1%	6%	0%	13%	0%	0%
	ACC	5%	5%	0%	3%	0%	0%
	INS	1%	8%	5%	13%	70%	72%
	LOC	0%	6%	1%	13%	0%	0%
	VOC	0%	4%	1%	0%	0%	0%
PLURAL	NOM	33%	74%	18%	44%	0%	0%
	GEN	41%	0%	14%	3%	0%	0%
	DAT	3%	0%	1%	0%	0%	0%
	ACC	41%	74%	1%	44%	0%	0%
	INS	3%	2%	1%	0%	0%	0%
	LOC	1%	0%	34%	0%	0%	0%
	VOC	33%	74%	18%	44%	0%	0%
	n.	1,507,650	88,790	192	32	82	145

Several observations are possible. First, only a limited number of word-forms are represented in the VILLA input: specifically, the entire plural paradigm is completely absent, which is generally the case with all nouns except only few

pluralia tantum, e.g. *drzwi*, "door". Second, even within the singular number, the VILLA input is much more restricted than the L1 equivalent, being limited to only a couple of forms: in this case, the nominative and instrumental case required by copular structures. The latter form, in particular, occurs in a construction which is specific to human nouns, in which the topic is instantiated by a personal pronoun: *ona jest polką*, "she is a Polish woman". Third, in the VILLA data case syncretism has a much lower impact, or none altogether, than it does in native varieties: consider the form *polki*, which in the L1 may correspond to the genitive singular or the nominative, accusative and vocative plural forms. Finally, to confirm the last two points identified by Ferguson (1975), namely simple vocabulary and simple concepts, native varieties show an obvious difference in the frequency of the masculine and feminine nouns, whereby the former, being semantically unmarked, overextends onto contexts in which gender is not relevant, indicating a generic person of Polish nationality. The term *polka*, in contrast, is only used when emphasis is placed on the fact that that particular person is a woman. Since none of this matters in the context of VILLA, frequency figures show no such trend: on the contrary, but for purely accidental reasons, the ratio is actually inverted in the corpus considered.

Table 9 presents a similar analysis on the basis of two inanimate nouns, namely *samochód*¹⁴, "car" and *kuchnia*¹⁵ "kitchen/cuisine".

¹⁴ Inflectional paradigm (cases provided in the same order as in column 1): SING: *samochód, samochodu, samochodowi, samochód, samochodem, samochodzie, samochodzie*; PLUR: *samochody, samochodów, samochodom, samochodami, samochody, samochodach, samochody*.

¹⁵ Inflectional paradigm (cases provided in the same order as in column 1): SING: *kuchnia, kuchni, kuchni, kuchnię, kuchnią, kuchni, kuchnio*; PLUR: *kuchnie, kuchni, kuchniom, kuchniami, kuchnie, kuchniach, kuchnie*.

Table 9: relative frequency of two inanimate nouns by word-form

		MONCO		NKJP Oral		VILLA	
	case	<i>samochód</i>	<i>kuchnia</i>	<i>samochód</i>	<i>kuchnia</i>	<i>samochód</i>	<i>kuchnia</i>
SINGULAR	NOM	23%	18%	42%	15%	99%	48%
	GEN	15%	65%	13%	58%	0%	0%
	DAT	0%	65%	0%	58%	0%	0%
	ACC	23%	8%	42%	13%	99%	52%
	INS	10%	5%	14%	9%	1%	0%
	LOC	4%	65%	9%	58%	0%	0%
	VOC	4%	0%	9%	0%	0%	0%
PLURAL	NOM	13%	3%	7%	4%	0%	0%
	GEN	17%	65%	4%	58%	0%	0%
	DAT	0%	0%	0%	0%	0%	0%
	ACC	3%	0%	2%	0%	0%	0%
	INS	13%	3%	7%	4%	0%	0%
	LOC	2%	0%	1%	1%	0%	0%
	VOC	13%	3%	7%	4%	0%	0%
	n.	1,315,800	151,246	1,001	369	111	48

The word *samochód* virtually occurs in a single form, which corresponds to both the nominative and the accusative case. Its semantics, though, causes it to always perform the function of object, at least in the VILLA corpus. Both forms are assigned the same frequency ratio because they could only be distinguished following syntactic disambiguation, which was not performed on the corpora performed here.

This said, it may come as a surprise that an inanimate word like *kuchnia* should distribute equally among its nominative and accusative form, despite its clearly inanimate semantics which should favour its association with the object function. However, there are two explanations for this phenomenon. First, *kuchnia* is a polysemic word, meaning both "kitchen" and "cuisine" (see examples (7) and (8) below). Since part of the VILLA course was devoted to a

decription of the various rooms of a home, this lexical items sometimes appears as the subject of copular constructions (7):

- (7) a. *gdzie znajduje się kuchni-a?*
 Where find REFL kitchen-NOM?
 "where is the kitchen?"
- b. *kuchni-a znajduje się na dole*
 kitchen-NOM find REFL downstairs
 "the kitchen is downstairs"

More generally, new lexical items were generally introduced in their citation form, which corresponds to the nominative case. It is not uncommon to encounter utterances like those exemplified in (8), in which the teacher first introduces the new lexical item and subsequently asks learners to repeat it aloud.

- (8) *TEA¹⁶: *proszę mówić kuchni-a*
 please say kitchen-NOM
 "please say «cuisine»"
- *TEA: *kuchni-a włosk-a*
 cuisine-NOM Italian-NOM
 "Italian cuisine"
- *TEA: *kuchni-a brazylijsk-a*
 cuisine-NOM Brazilian-NOM
 "Brazilian cuisine"

These two examples suggest that the semantics of individual lexical items have a very strong impact on the word-forms in which they occur. In chapter 6, p. 198 we will show that this is indeed the case.

¹⁶ When quoting two or more subsequent turns in the input, we include the line headers included in the original CHAT (MacWhinney 2000) transcripts (see chapter 6, p. 142): here, *TEA corresponds to "teacher",

We return to copular structures to illustrate another source of deviations from native varieties, namely pragmatics. Two main types of predicational copular clauses may be distinguished in the VILLA input¹⁷. In NOM-type structures¹⁸, the invariable pronoun *to* "this" is supplied independently of referent gender and number, in both affirmative and interrogative sentences. The complement appears in the nominative form, which is instantiated by the bare consonant stem (-C) in the case of masculine nouns and by the ending -*a* for feminine nouns. INS-type structures, in contrast, require the personal pronoun *on* "he" or *ona* "she"¹⁹, which also specify the gender of the corresponding referent. The noun is provided in the instrumental case, instantiated by the endings -*em* and -*q* on masculine and feminine nouns respectively. The two types are exemplified in Table 10.

Table 10: Polish copular structures comprised in the VILLA project

Interrogative sentence	Declarative sentence	Structure	Type
<i>kto to jest?</i> who.NOM this is?	<i>to jest listonosz</i> this is postman.NOM	to jest NP.NOM	NOM-type (M)
<i>kto to jest?</i> who.NOM this is?	<i>to jest dziewczyna</i> this is girl.NOM	to jest NP.NOM	NOM-type (F)
<i>kim on jest?</i> who.INS he is?	<i>on jest listonosz-em</i> he is postman.INS	PN jest NP.INS	INS-type (M)
<i>kim ona jest?</i> who.INS she is?	<i>ona jest dziewczyn-q</i> she is girl.INS	PN jest NP.INS	INS-type (F)

In native varieties of Polish, the two structures are pragmatically differentiated. NOM-type structures are mainly used deictically, *to* being used to indicate its referent ostensively, as in *a to, to kto to jest?* "and this [person], who is this [person]?". As a result, this kind of copular clauses is typically to introduce new referents in the discourse. It follows that in no case can the pronoun be

¹⁷ For a detailed description of the full range of Polish copular structures, see Bondaruk (2013).

¹⁸ The labels "NOM-type" and "INS-type" were originally introduced in Saturno (2015a).

¹⁹ The neuter pronoun *ono*, "it" does not appear in the VILLA input.

substituted by the corresponding noun: **a Karol, kto to jest?* "and Karol, who is this?". In contrast, the personal pronouns of INS-type structures typically refer to entities in an anaphoric manner and ask for additional details about them. The referent should already be part of the discourse: *a on, kim on jest?* "and he, who is he?". In this case pronouns and nouns are perfectly interchangeable, e.g. *a Karol, kim on jest?*, "and Karol, who is he [what's his job/nationality etc]?". In contrast, in the VILLA input the two structures are used more or less interchangeably in all contexts, so that they are not functionally differentiated. Example (9), extracted from the teacher's speech, shows that the two structures may be used to refer to the same entity in the same context. The teacher first asks who Karol is using a an INS-type copular structure, in which the referent appears in the nominative case and the interrogative pronoun is in the instrumental case. Using this structure in the question calls for the same structure in the response: effectively, in the two subsequent utterances the referent is instantiated by a person name and by a personal pronoun, but the predicate appears in the instrumental form, as pragmatically required.

- (9) *TEA: *A Karol, kim jest Karol?*
 and Karol.NOM who.INS is Karol.NOM?
 "And Karol, who is Karol?"
- *TEA: *Karol jest strażakiem.*
 Karol.NOM is fireman.INS
 "Karol is a fireman"
- *TEA: *On jest strażakiem.*
 he is fireman.INS
 "He is a fireman"
- *TEA: *To jest strażak.*
 this is fireman.NOM
 "This is a fireman"

However, in the last utterance, the teacher switches to a NOM-type copular structure, in which the noun appears in the nominative case and the referent is deictically instantiated by the invariable pronoun *to*, "this". While this structure is grammatically correct, of course, it would sound pragmatically inappropriate in a native variety of Polish.

This was evidently done intentionally for didactic purposes, in order to show learners that predications about referents may be expressed by two different syntactic constructions, which were tested in several tests. However, it is clear that the research questions concerning the acquisition of these structures were only formulated in terms of morphosyntax, neglecting pragmatics and information structure. Two negative consequences derive from this approach: first, the sequence in (9) would probably seem inappropriate to a native speaker of Polish; second, and worse, the two structures express exactly the same meaning, so that learners do not have any functional motivation to learn them both. Instead, they might use them interchangeably, focus on just one, or mix them up without any risk of misunderstanding.

Similar things also happen with respect to one of the features of Polish lying in the focus of the present work, namely word order. We have already anticipated that subject, object and verb may occur in whatever order. This does not mean, however, that different versions veiculate an identical meaning, at least from a pragmatic point of view (Siewierska 1993). While a detailed description of this phenomenon lies beyond the scope of our work, we may say that in general, the first position in the utterance is associated with the topic function, so that moving the object to that position from its canonical post-verbal position equals treating the subject as the focus and the object as the topic. Consider example (10), extracted from a web-page titled "how to buy chocolate"²⁰.

- (10) *Czekolad-ę* *lubi* *prawie* *każdy*.
Chocolate-ACC likes almost everyone-NOM

"almost everyone likes chocolate / chocolate, almost everyone likes it"

²⁰ <http://www.eko-market.pl/content.php/id/341>

As the topic of the web-page is precisely "chocolate", it is pragmatically appropriate for this introductory sentence to present this item in the topic position, although it performs the syntactic function of object. In contrast, later on in the text one encounters example (11):

- (11) *możesz*²¹ *wybrać* *czekolad-ę* *biał-ą*.
 can.2SG choose chocolate-ACC white-ACC
 "you can choose white chocolate"

Here, the focus is on the fact that people allergic to cacao can buy *white* chocolate as opposed to dark. As a consequence, this element occurs in the appropriate post-verbal focus position, while still performing the object function. As the text is written, intonation is of no help to veiculate information structure, but morphosyntax is sufficient.

The situation in the VILLA project looks a little different, as example (12) makes clear.

- (12) *TEA: *Filip* *wózek ciągnie*.
 Filip.NOM cart.ACC pulls
 "Filip pulls the cart"
 *TEA: *Filip* *ciągnie* *wózek tak* .
 Filip.NOM pulls cart.ACC yes.
 "Filip pulls the cart"
 *TEA: *wózek ciągnie* *Filip*.
 cart.ACC pulls Filip.NOM
 "Filip pulls the cart"

If these utterances were produced in the context of a conversation among native speakers, several facts would appear a little odd. First, the same

²¹ Note that Polish is a pro-drop language.

referential content is repeated three times, which means a lot even in a context of language teaching. Second, if we were to interpret these sentences on the basis of native pragmatics, we should conclude that the first one focusses the verb, as in "Filip doesn't push the cart, he *pulls* it"; the second, at least in the absence of specific intonational patterns, is unmarked; and in the third, the subject is focussed, as in "it's not Julia who pulls the cart, it is Filip". However, none of this was treated in the VILLA project.

In fact, the purpose of the teacher in this obvious manipulation of syntax was simply to show that Polish word order is free. This was functional to research questions concerning cross-linguistic interference, such as "will speakers of rigidly SO languages be able to recognise and, perhaps, exploit Polish free word order?". What is missing in this question is the purpose why learners should try and exploit this complex feature of the target language, that is, "in order to express pragmatically marked meaning". In any case, the teacher had to make sure that the learners were exposed to a sufficient number of structures with a word order other than SVO. Unfortunately, given the very limited scope and vocabulary range of the VILLA project, it was quite hard to do so in a pragmatically meaningful way. The important link between information structure and syntax had to be sacrificed for the sake of frequency. While such methodology certainly exposed learners to an appropriate number of examples of marked word order, it deprived such markedness of meaning. It may be licit to wonder, then, if learners even deemed it necessary to pay attention to such syntactic means, since meaning doesn't change: at most, they may have regarded it - provided that they noticed it, of course - as a syntactic curiosity of this exotic language. Speakers of languages which also allow the manipulation of word order, like German and, with different means, Italian, might even have found this apparently random use of syntactically marked structures a little odd. The same could be said about the copular structures discussed earlier. In both cases, from the point of view of form-function associations, the alternative structures presented in the input are not worth distinguishing, because they mean just the same. This said, the very school

context in which the project was carried out might have prompted students to pay attention to these details of the target language anyway, despite the difficulty to associate them to the corresponding meaning.

But besides its differences with native varieties, it would be inappropriate to state that the VILLA input differs significantly from Polish, or that it is not a natural language. Doing so would obliterate a great number of relevant points. First, the VILLA input retains numerous idiosyncrasies which are likely to create difficulties to the learners, and at the same time resemble the familiar difficulties they encountered when learning other natural languages, including their own L1. We already mentioned the case of *pluralia tantum*: these include several country names which recurred commonly throughout the course, such as *Włochy*, "Italy", *Niemcy*, "Germany", *Chiny*, "China". These elements constitute the only instances of plural forms in an otherwise strictly singular input. Further, the category of animacy is morphosyntactically relevant in the case of masculine nouns: specifically, the accusative case of animate nouns is syncretic with the genitive case, whereas in inanimate nouns it is syncretic with the nominative case (Table 11).

Table 11: animacy in the paradigm of Polish masculine nouns

NOM	ANIMATE	<i>strażak</i>	"fireman"
GEN		<i>strażak-a</i>	
ACC		<i>strażak-a</i>	
NOM	INANIMATE	<i>stół</i>	"table"
GEN		<i>stoł-u</i>	
ACC		<i>stół</i>	

However, the criterion which governs whether a noun should be classified as animate or inanimate is sometimes hard to grasp: in addition to human beings, animate nouns include some animals and, rather inexplicably, even such words as *hamburger*.

To summarise, then, the VILLA input represents a specific variety of Polish, one whose features perhaps are not uncommon in the context of L2 instruction. Thus, although the lexical repertoire is necessarily reduced, and syntax may be sometimes bent to the need of didactics, we can safely state that the VILLA input retains all the complexities and the idiosyncrasies of a natural language.

1.3.5. Data collection

The main tool to elicit L2 data is represented by structured tests, schematically listed in Table 12. Although the focus is on morphosyntax, several other levels of language were targeted too.

Table 12: VILLA project, linguistic tests

Test	Target
Phoneme Discrimination	phonology
Lexical Decision	vocabulary
Word Recognition	phonology
Grammaticality Judgement I	morphosyntax
Picture Production	morphosyntax
Written word order	syntax
Picture Verification (morpho-syntax)	morphosyntax
Sentence Imitation (morpho-syntax)	morphosyntax
Grammaticality Judgement II	morphosyntax
Cloze (Pronouns)	pro-drop
Route direction	free production: reference to space
Finite story (film retelling)	free production: narrative

In order to control for specific learner attitudes which may have an impact on the tests above, individual difference measures were also taken (Table 13). Although correlations with the results of the structured tests above have been

attempted (Watorek & Saturno 2016; Saturno forthcoming), these tasks will not be considered in the present work.

Table 13: VILLA project, psychometric tests

Language profile
Nonverbal intelligence
Personality (Big Five)
Memory span
Working memory span
Executive function (attention and inhibition)
cognitive style: perceptual preference
learning style
Motivation
metalinguistic awareness: grammatical inferencing
metacognitive: associative learning
phonological memory/ awareness

Finally, the VILLA project offers the exciting and innovative opportunity to monitor learner output during classes throughout the whole duration of the experiment. As we will widely take advantage of this possibility, the reader is referred to chapter 5, p. 161.

1.3.6. The VILLA project: an interim summary

While typical first exposure studies pursue specific research questions related to the minimal exposure required to identify regularities in the input, not necessarily linked to any meaning, the VILLA project offers the opportunity to do much more. Acquisition is observed from its very onset, like in laboratory studies, but for a much longer time: 14 hours. Input is fully controlled and thoroughly monitored, yet it is realistic in that it takes the shape of a language

course. Moreover, it is provided by a teacher in the flesh, rather than by some kind of recording. The numerous tests it comprises target numerous layers of language, a possibility which will turn very useful in this work.

Though research based on the VILLA project is still at a relatively early stage, in view of the immense amount of data produced by the experiment, several works have been published to date which are helpful to obtain a general picture of what breaking into Polish L2 at first exposure might mean. Below we briefly review some research based on the VILLA project as well as the pilot studies which preceded it (e.g. Rast et al. 2014) and largely shared its structure, though time of exposure and learner population were considerably smaller.

In addition to a few general papers (Dimroth et al. 2013; Rast 2015; Latos et al. 2016; Bernini forthcoming), research on the VILLA data so far has mainly focussed on morphosyntax and phonology, coherently with the number of tests devoted to this levels of the language. Regarding the former, Shoemaker and Rast (2013) investigated the input features which make words more or less likely to be extracted from the input stream at first exposure, discovering strong effects of sentence position and lexical transparency, but not frequency, contrary to the initial hypotheses. This study builds on previous work by Rast and Dommergues (2003) and Rast (2008), who reported of complex interactions among such factors as word length, word stress, phonemic distance, transparency, position and frequency as well as time of exposure. Shoemaker (2015) argues that learners improved significantly over time in their ability to discriminate sibilant sounds comprised in the rare and difficult inventory of Polish {s/z/ts/dz, ʃ/ʒ/tʃ/dʒ, ʒ/ʒ/tʃ/dʒ}. Finally, Bernini (2016) attempts a classification of learner varieties on the basis of their phonological component, in the same way that researchers working on the Basic Variety looked at morphosyntax. He argues that while widespread allophonic dispersion may be associated to the pre-basic variety, the reduction of such phonetic instability is related to the establishment of a pre-phonological structure which parallels the more or less stable basic word form of the Basic Variety. By the same token, the reduction of syllabic instability anticipates the establishment of a pre-

morphological structure as in post-basic varieties. Valentini and Grassi (2016) looked at the recognition rates of Polish words by initial listeners. As frequency plays no role in this context, the two authors discuss the effects of the position of stress and identity of the stressed syllable.

Regarding morphosyntax, Hinz et al. (2013) examined and correlated the Grammaticality Judgement and Question & Answer tests as performed by French and German learners. Their research questions concerned the role of the L1, test type, frequency and lexical transparency. While the initial superiority of the German group was quite evident, all other factors were shown to engage in interactions with each other, describing a rather complex picture of input processing. The same tests were also investigated by Rast and colleagues (2014) on the basis of one of the VILLA pilot studies, depicting an equally complex picture.

A couple of papers, including one by ourselves (Latos 2014; Saturno 2015b) are devoted to the effect of the version of input received, meaning-based or form-based (see 1.3.2, p. 22). The latter work regards the repetition test discussed in chapter 2 and is therefore of some interest to us here. While the meaning-based group, also considered in the rest of the present work, proved sensitive to case ending, word order and lexical transparency, only the first factor had an effect on the form-based group. The study concluded that the explicit presentation of the input somehow inhibited the learners' natural mechanisms to carefully and autonomously listen for regularities in the target input.

Among the topics to which our own work has been devoted we briefly signal copular structures (Saturno 2015a), transcription (Saturno 2013; 2015c), word formation (Saturno 2016) and more recently the effect of interaction on the processing of morphosyntax (Saturno in prep; in prep). Its bulk, however, has been devoted to the repetition test also discussed in this thesis (Saturno & Watorek forthcoming; Saturno 2014; 2015b; 2015d; forthcoming).

1.4. RESEARCH QUESTIONS

The purpose of this work is to analyse learner strategies of input processing in the early stages of second language acquisition, with a particular focus on morphosyntax. The object of our study are transitive sentences, that is, utterances in which both the subject and the object functions are realised. This general objective is pursued through two structured tests, namely a repetition (chapter 2) and a comprehension (chapter 3) task, as well as the elicitation of semi-spontaneous interaction in pairs or small groups (chapter 5).

We ask to what extent learners employ inflectional morphology to extract and encode the meaning²² of utterances in comprehension and production. If we assume that learners at least tried to process target items, this question is equivalent to asking what principles learners rely on to process target sentences, by which we simply mean what was paraphrased a little earlier: extract and encode the meaning of target sentences, that is, understanding what they mean and expressing what one wants to say. The answer to this question is comprised between two extremes - target-like morphosyntactic processing and basic-variety-like positional processing - but many nuances are possible depending on various factors.

Let us consider the case of a simple SVO sentence like (13).

- (13) *dziewczynk-a* *ciągnie* *portugalk-ę*
little girl.NOM pulls portuguese woman.ACC
"the little girl pulls the Portugues woman"

Such an utterance may be successfully interpreted by relying on at least two processing strategies. First, learners may perform native-like morphosyntactic analysis, i.e. derive syntactic functions²³ from case endings. This requires that

²² For the time being, this term is used with its common interpretation. See below for a discussion of its specialistic use in the literature.

²³ There is no evidence that learners process target sentences based on syntactic categories such as Subject and Object. Instead, it is equally possible that they might rely on semantic categories

case endings are categorised into paradigms depending on several features of the individual lexical items, such as inflectional class, gender, animacy and number, as the same ending - understood here as a sound or string of sounds - may veiculate different meaning in different inflectional paradigms. For instance, we have shown that the ending *-/a/* may encode, among other meanings, that of nominative of feminine nouns as well as accusative of masculine animate nouns. Thus, deriving meaning from inflectional endings requires the learner to know what inflectional class a given lexical item belongs to.

Alternatively, this utterance may be interpreted according to a positional principle: words are assigned syntactic functions depending on their relative position in the utterance, i.e. the order in which they appear in the string. A key is necessary to do so: various sources of data suggest that this may be the SO word order, which we therefore consider as the unmarked option.

There are several reasons for this preference. Firstly, it is the dominant constituent order in all L1s involved in the VILLA project, although with various degrees of rigidity and obligatoriness. Secondly, typological research suggests that this constituent order is by far more widespread among the languages of the world than OS orders (Dryer 2013b). The reasons for such biased ditribution, in turn, are generally believed to be of a cognitive nature (see discussion in Siewierska & Bakker 2008).

We can thus further reformulate our question as whether or not learners will be able to process the meaning of an utterance on the basis of inflectional morphology, rather than by relying on a default constituent order. In more general terms, this question is sometimes referred to in the literature as their ability to process form in addition to meaning. VanPatten's (1996) Principle 1,

such as Controller and Controllee, pragmatic notions like Topic and Focus, or thematic roles such as Agent and Patient, among others. Indeed, all these categories have been shown to have a prominent role in the structuring of learner utterances in early varieties such as the Basic Variety: the two principles "focus last" and especially "controller first", for instance, are usually compatible with SO word order. Since Subject and Object appear fully adequate to describe our data, however, we will refer to them for reasons of simplicity.

for instance, concerning "The Primacy of Meaning", generally says that learners usually attend to meaning before they process an utterance for form.

The word "meaning" appears a little ambiguous here, as grammar has its meaning, too: it simply is of a different nature compared to the lexicon. This point has been discussed by Carroll (2012), who points out that meaning may be distinguished between reference (to an entity), and sense (the relations that an element establishes with others in a text). She shows that person names, typically thought not to express sense, may be acquired after as little as 2 exposures. Similarly, Gullberg et al. (2012) show that after only 8 exposures, their Dutch learners were able to recognise Chinese sound strings; when accompanied by gestures, learners were also able to match them to the pictures which had appeared in the weather forecast constituting their input. Carroll (2013:134) warns, however, that this does not imply that learners acquired those words in terms of semantic representations (i.e., as language units possessing sense); rather, they might be more similar to names as in her own study. This is partly because, she argues, in the limited input provided in laboratory studies, unlike in reality, words and referents occur in a 1:1 ratio. In this respect, Vouloumanos (2008) targets the problem of having many potential relations between words and meanings and wonders whether statistical information may serve as a source of data. She shows that participants were indeed better at making associations in the case of 1:1 correspondence, though they were able to cope with less univocal form-function associations too.

What do we mean, then, by saying that learners process "meaning"? Essentially, it would seem, that they link a sound string, also known as a word, to its real-world referent. This appears rather straight-forward in the case of lexical items. Grammatical morphemes, in contrast, are used to establish relations within the sentence, and thus possess sense, but are not linked to any referent. In sum, it would seem that learning vocabulary items is not different from memorising person names, in itself, the only difference being the semantic extension the referents involved.

A sub-principle of VanPatten's "Primacy of meaning" principle cited above says that learners process content words in the input before anything else (P1a, The Primacy of Content Words Principle), as effectively proved in sentence repetition tests such as those performed by Klein (1986) and Rast (2008). Our own repetition test shows that nouns tend to be repeated more often and more accurately than verbs. While this fact could be explained in terms of perception, as verbs occur in the poorly-prominent utterance-medial position (Peters 1985), it could also be a consequence of the fact that it is harder to find a referent for verbs than it is for nouns as the former instantiate a predicate, i.e. they describe the relation of one or more participants (arguments) to a situation (Jezek 2016; Lyons 1977). It does not sound surprising, then, that learners should first pay attention to those elements in the input stream which they can link to some kind of referent, almost like names of known people or places in an otherwise unintelligible flow of sounds. Posing the problem in these terms is helpful to understand one of the most common findings of first exposure studies, especially with regard to L1 interference. Park (2013), for example, exposed Japanese and English L1 learners of various proficiency levels to written Korean L2 input. The author claims that beginner learners of both groups exhibited a form-oriented noticing behavior, guided by perceptual cues. In the more experienced group, however, Japanese learners adopted a more meaning-oriented approach, whereas English learners maintained their form-oriented approach strategy. In the same vein, Han and Peeverly 2007 asked absolute beginners of Norwegian to read a text, try and recall the main facts described in it, and then complete a fill-in-the-blank task and answer a questionnaire. Learners were reported to employ a form-based approach, supposedly contrary to the primacy of meaning principle.

While these and similar claims (e.g. Han & Sun 2014) are usually interpreted in terms of meaning-based or form-based processing, it seems that the findings have much in common with so-called cognate effects. For instance, while Carroll (2012a) indeed found that English L1 learners of German L2 easily associated person names with their referent, she also reports that when cognate

names sounded similar in the two language, learners would pronounce them according to the L1 phonology. Rast and Dommergues (2003) asked beginner French L1 learners of Polish L2 to repeat a set of sentences and try to translate them. They found that correctly translated words were mostly cognate between the two languages, and that these were also easier to repeat. All these studies suggest that when exposed to uncomprehensible input, learners will naturally try and grasp any familiar-sounding item before turning to processes of associative learning based on statistical learning, i.e. what one would call processing for form. Of course, if you ask participants to try and identify formal regularities in the input, they will do precisely that. But then, focussing on form when no familiar word can be identified is more similar to solving a logical problem than to processing speech: one looks for regularities and tries to make sense of them, usually by mere guessing.

The VILLA project is not quite the same as the laboratory experiments referred to above. It is a language course, during which learners quickly made progress and developed their competence in the L2, so that after the first classes they no longer were "initial learners" in the sense of many works in the field. At that point, VanPatten's (1996) principles become interesting because it is no longer obvious that learners cannot and will not process form under realistic conditions²⁴.

"The Lexical Preference Principle" (P1b) states that learners will tend to rely on lexical items as opposed to grammatical form in order to get meaning when both encode the same semantic information. This principle is particularly interesting as the word "preference" implies that learners finally have a choice as to what to process. For instance, when verb tense is expressed both by inflectional morphology and by adverbials, learners would typically focus on the latter (Bardovi-Harlig 1992a), even though they probably know how verbal morphology works in the target language. In real-time processing, though, they prefer to rely on cues that are more salient and usually quite reliable, like

²⁴ Indeed, these principles were not specifically formulated to account for first exposure acquisition, but for more advanced learner varieties, usually in the context of tutored instruction, in which learners could have the possibility to attend to form, at least theoretically.

adverbials. If we apply this logic to our case, learners may derive grammatical meaning, that is, syntactic functions, either through an analysis of inflectional morphology or on the basis of word order. While the latter is not a content word, of course, it is certainly more conspicuous than the poorly prominent segments which instantiate inflectional morphology. More importantly, word order is also the object of another principle, namely P2, "The First Noun Principle", according to which learners tend to process the first noun or pronoun they encounter in a sentence as the subject or agent. Here, too, it is not difficult to find empirical support, both within VanPatten's circle (Lee 1987; LoCoco 1987; VanPatten 1990) and beyond. Consider again the "focus last" and "controller first" principles identified for the Basic Variety, compatible in most cases with SO word order, as well as the studies conducted within the framework of Processability Theory cited in section 1.1, p. 6.

While processing for meaning justifies the regularities of learner behaviour, it can also motivate the departure from the unmarked word order strategy just referred to. Consider the following sub-principles:

- P2a, the Lexical Semantics Principle. If the semantics of the referents involved makes the interpretation based on SO word order unlikely, learners may decide not to rely on it. In sentences with the verbs bring and give, for example, SO word order is usually excluded when the first noun is instantiated by an inanimate entity, rather unlikely as a subject of those verbs (LoCoco 1987).
- P2b, the Event Probabilities Principle. Again, default word order may be modified if it does not fit the learner's real-world knowledge and common sense. See Houston (1997) and Marlovrh (2006) on the interpretation of OS strings involving known and unknown referents.
- P2c, the Contextual Constraint Principle. Default SO word order may be abandoned if the resulting interpretation does not seem likely in the present communicative context, as shown by VanPatten and Houston (1998).

Regrettably for the possibility to generalise the experimental results of the VILLA project, the "common sense" principles just cited are not applicable in the structured tests which will be used in our work, though it will be shown in the chapter devoted to semi-spontaneous interaction (5, p. 161) that they may be effectively employed in a more realistic communicative situation. For the time being, consider the typical target structure of the Repetition (14a) and Comprehension tests (14b):

- (14) a. *dziewczynk-a* *ciągnie* *portugalk-ę*
 little girl.NOM pulls portuguese woman.ACC
 "the little girl pulls the Portugues woman"
- b. *siostr-a* *pcha* *brat-a*
 sister-NOM pushes brother-ACC
 "the sister pushes (her) brother"

Both are pronounced with neutral intonation and only include animate human nouns, equally likely to exert control over the specified action (and in fact, each referent in turn appears in both the subject and object function). In general, due to the absence of any context, both seem to make little sense. From this, we can draw two conclusions: first, that the task is highly abstract and has no communicative value, its only objective being performing the exercise correctly; second, that the only resources available to the learners to do so are inflectional morphology and word order.

Exemplifying on sentence (14a) above, processing for meaning would consist in the steps listed below. We remind the reader that in VanPatten's (1996) terms, processing implies making form-meaning connections and parsing, that is, identifying the structure of the situation, typically "who does what to whom".

1. Segmenting speech, so as to identify known or deducible words: in our case, *dziewczynka*, *ciągnie*, *portugalka*. Note that all of them are cited as

- lemmas, that is, in their nominative citation form, rather than in other inflected word-forms;
2. Connecting form with meaning: *dziewczynka* > "little girl", *ciągnie* > "pull", *portugalka* > "Portuguese woman".
 3. Identifying relations:
 - a. since *ciągnie* is a transitive verb, the two nouns will instantiate its two main arguments, which in turn are associated with specific syntactic functions;
 - b. based on a positional principle, the first argument is instantiated by the first nouns, and the second argument by the second noun;
 4. Conclusion: "the little girl pulls the Portugues woman".

This conclusion is correct, because effectively the target has an SO constituent structure which allows for the use of a positional principle. This strategy certainly represents a guess, but not a random one: as we will show in the chapter devoted to input (6, p. 198), SO structures represent the vast majority of transitive utterances in the VILLA input. The same is true for native varieties of Polish, too (Dryer 2013b; 2013c). However, if the target structure had had OS word order, then the learner's processing would have led to the wrong conclusion.

Processing the same utterance for form, in contrast, would imply the following steps:

1. Segmenting speech: as above;
2. Connecting form with meaning: as above;
3. Identifying relations:
 - a. since *ciągnie* is a transitive verb, the two nouns will instantiate its two main arguments, or syntactic functions;
 - b. since *dziewczynka* occurs in the nominative case (i.e. ends in -a), it represents the subject of the verb. Conversely, *portugalka* is in the accusative case (i.e. ends in -e) and therefore represents its object;
4. Conclusion: as above.

Of course, in a real communicative situation this processing for form does not exclude the use of other clues, such as context, intonation and word order. But in the context of the VILLA structured tests it represents the only effective strategy to derive the meaning of OS targets. For this very reason, the manipulation of word order is going to serve us as a diagnostic tool for the learner's processing strategy.

We assume that the closer the learner variety gets to its target²⁵, the more learners will be able to rely on processing for form, if necessary. Indeed, this expectation seems confirmed in the existing literature. Lower proficiency adults primarily rely on lexical cues and other non-morphological means to express grammatical relations, such as word order in syntax and chronological order and adverbials in temporality (Klein & Perdue 1997, Bardovi-Harlig 2000; Lee 2002; Leiser 2004; Ellis & Sagarra 2010; Sagarra 2014 for a discussion). In contrast, higher proficiency learners behave more similarly to native speakers in that they rely on the cues which are most relevant in the particular language learnt (Giacalone Ramat 1992; Bardovi-Harlig 1992b; Skiba & Dittmar 1992; Dietrich, Klein & Noyau 1995; Starren 2001; Bardovi-Harlig 2000; Parodi, Schwartz & Clahsen 2006; Bordag & Pechmann 2007). Note that this does not imply "morphology": MacWhinney and colleagues (1984), for instance, have

²⁵ Note that native varieties may be considered as very advanced and automatised learner varieties, too, as made clear in Perdue (1993).

shown that while, effectively, subject-verb agreement is the most effective cue for retrieving the subject in Italian, word order is its best index in English. Advanced learners, as competent speakers, simply rely on the cues which are most widespread and reliable, independently of their nature.

Alongside these general tendencies, specific target structures may be more easily processed by speakers of a given L1 because they rest on identical processing mechanisms, which do not have to be devised and established anew (Bates & MacWhinney 1981; Ellis 2006a; Hopp 2013; Tokowicz & MacWhinney 2005; MacWhinney 2011). Effectively, some studies show that speakers of morphologically complex languages are more at ease when processing equally complex target languages. Ellis and Sagarra (2010a; 2010b; 2011) studied the acquisition of temporal reference in Latin after only one hour of exposure. While they found that focussing learner attention to verbs or adverbs orients their processing strategies towards that category, they also highlighted important L1 effects. Specifically, speakers of morphologically poor languages such as Chinese and English tended to rely more on lexical cues than did speakers of morphologically more complex L1s, such as Spanish and Russian. When paradigm complexity increased, however, all learners proved biased towards lexical cues. The same researchers (Sagarra & Ellis 2013; Sagarra 2014) eye-tracked the processing of Spanish L2 temporal reference by English L1 and Romanian L1 learners, discovering that the intermediate and advanced speakers of the more complex L1 are sensitive to tense incongruencies and tend to rely more heavily on verbs than their English equivalents, who mainly focus on lexical cues such as adverbials.

We can now summarise this bibliographical review and apply its main trends to the case of the VILLA project. Since our work adopted an explorative, descriptive approach, we will not formulate detailed and verifiable research questions here. However, we do have some general expectations, based on the research findings presented above. These are the following:

- a) It is quite likely that most learners will not be able to process transitive sentences for form at such early stages of acquisition. This should be evident in their high scores in the processing of SO targets, and minimal scores on their OS equivalents²⁶;
- b) Processing for form, signalled by above-chance scores on OS targets, might become more common as acquisition proceeds, as a result of the learner variety getting closer to its target. As the two structured tests were administered at two test times, we expect higher scores at T2 than at T1;
- c) Processing for form might be influenced by the degree of morphological complexity of the learners' L1. In this respect, we generally expect the German group to perform better than the other language groups;
- d) Finally, given the unnaturally abstract setting of the structured tests, we expect learners to rely on rather different processing strategies in that context than when observed in spontaneous interaction. The latter communicative situation, particularly, should allow them to rely on context, phonology, semantics and world knowledge.

We hope that this chapter provided the reader with all the tools required to go through the rest of this work. We then turn with no further hesitation to the first source of data, namely the Repetition test.

²⁶ This is only a general expectation, which could apply more or less directly to the comprehension test, less so to the repetition test. For details about the assumptions and the interpretation challenges of the latter, see chapter 2, p. 35.

2. THE REPETITION TEST

2.1. INTRODUCTION

Repetition tests, also called "elicited imitation" or "sentence repetition" tests, have been used successfully to investigate the implicit competence of a large range of populations, including L1 and L2 learners and patients affected by speech pathologies. Effectively, they offer numerous advantages to both language scientists and testers: if one accepts their rationale, then they appear relatively quick and easy to administer, require little equipment, and offer full control on targets (Van Moere 2012). This last point is certainly the most appealing to linguists, as it makes it possible to study linguistic structures which would otherwise take hours of spontaneous speech to observe, with no guarantee that they would turn out at all. In the present study, for instance, we will focus on the morphosyntactic opposition between the nominative and accusative case, arguing that only in utterances with OS structure can we be sure of its productive use. If we wanted to verify our claim using spontaneous data, we would have to wait for the learner to spontaneously produce an OS structure, which requires a certain degree of grammatical competence, certainly, but also the appropriate pragmatic context as well as the learner's intention to exploit it. In other words, even if the learner - or the native speaker, at that - can be thought to be equipped with the required grammatical and pragmatic competence, there is no warranty that he will use it even in the appropriate context, simply because the speaker is in control of his own speech. Things may be forced to a little extent, for instance by using tasks which make a specific structures particularly appropriate in a given context: for example, we could imagine a situation in which the object of a verb is topicalised in discourse, and therefore could appear in utterance-initial position, as in (15):

- (15) a. *kto* *gotuje kolacj-ę?*
 who.NOM cook dinner-ACC
 "who is cooking dinner?"
- b. *kolacj-ę* *gotuję ja.*
 dinner-ACC cook I.NOM
 "I am cooking dinner"

But this is just one out of many possibilities. The speaker might just emphasise the subject pronoun using intonation, thus maintaining the default SO word order; or indeed, he might produce a short answer like *ja*, "I (will)". In sum, spontaneous speech is undoubtedly the most ecologically valid measure of learner competence, so much so that various scholars, like Pienemann (1998) or Krashen (1981), have long advocated that research should exclusively, or at least primarily rely on data elicited in this manner; but it certainly may be less practical than other elicitation techniques.

The EI test would seem a good compromise between ecological validity and control over the target structures. Learners hear a target sentence and are asked to repeat it, the rationale being that one can only repeat the grammatical structures that are already part of their developing L2 grammar. With few exceptions, the test is administered orally, which should guarantee a speech rate close to that of spontaneous speech, and therefore, that the same implicit competence is accessed: test takers should not have the time to rely on explicit, declarative knowledge (Ellis 2005). What the researcher observes, therefore, is what the learner is really able to do in a realistic situation: target-like output if the structure has been acquired, alternative means to convey what the learner has understood of the target sentence, if it has not, or nothing at all, if the target is just too hard to process. What one should not expect to observe, it is argued, is a *verbatim* repetition of the target as a mere string of sounds. The idea is that people can remember meaning, but not form: the latter has to be produced anew in repetition. To this end, the design of the test becomes crucial. By relying on working memory (Baddeley 2003), in effect, humans can

remember short strings of sounds for a handful of seconds and then repeat them with reasonable accuracy (Sachs 1967), even without understanding their meaning: this is due to the phonological loop (Baddeley, Gathercole & Papagno 1998), the device which makes it possible to mentally store and rehearse a chain of sounds for some time before it fades. A properly designed EI test should inhibit this mechanism by engaging the test-taker in some distracting activity, preferably of a verbal nature: even some delay between the presentation of the stimuli and their repetition should inhibit exclusive reliance on phonological memory (Juffs & Harrington 2011). Under these conditions, the test becomes reconstructive in nature, meaning that participants have to listen to the targets, decode them, and then re-produce them, all steps being performed on the basis of the present developmental stage of the interlanguage. Indeed, it has been observed that test-takes may systematically produce ungrammatical structures or, conversely, correct ungrammatical targets: discussing the results of a test similar to that employed in this study, Håkansson (1989) for instance reports of a three-year old Swedish child consistently reproducing a NEG-AUX structure instead of the required AUX-NEG target, consistently with the principles of markedness. These reports witness that test-takers do not just repeat a string of sounds, but interpret and reproduce it "in their own way".

Not all researchers would agree with this rationale, and the relation of the EI test with working memory is certainly a complicated one. Many factors are thought to come into play, including the nature and length of the stimuli, the type of distractor, the target structure, the learner's proficiency level, and many more. Since a systematic discussion of this topic is not the aim of this chapter, the interested reader is referred to Vinther (2002) and Erlam (2006) for recent reviews. For our purposes, it is sufficient to say that some argue that the EI test has nothing to do with implicit competence, and only measure a learner's working memory capacity (Jessop, Suzuki & Tomita 2007), whereas others claim that working memory is only meagerly involved if at all (Okura & Lonsdale 2012).

A good example of this heated debate is the relatively recent controversy between Zhang and Lantolf (2015) and Pienemann (2015), who on the ground of a Language Learning special issue on developmental sequences disputed as to whether this task could reliably represent a learner's processing ability. Aiming to verify the Teachability Hypothesis (Pienemann 1984), Zhang and Lantolf exposed four English L1 learners of Chinese L2 to specially designed instruction. The results contradicted the initial hypotheses because learners were shown not only to be able to process structures deemed too advanced for them, following appropriate instruction, but also to skip developmental stages. Pienemann (2015) questioned Zhang and Lantolf's results on various grounds, among which their claim that they used the same elicitation methods and emergence criteria utilized in PT-inspired research. He argues instead that "data obtained through EI cannot be compared one to one with spontaneous speech production data. In terms of language processing, the two types of data tap into different psycholinguistic mechanisms" (2015:139). He cites in support a paper by himself, Keßler and Lenzing (2013), which demonstrated experimentally that learners of Swedish L2 systematically show better performance in repetition than spontaneous production. One key objective of that study was to differentiate formulaic echoes of teacher utterances and creative L2 production. Spontaneously produced structures were expected to be strictly in line with the L2 implicational hierarchy; in contrast, learners should have been able to repeat structures produced by the teacher, but beyond their current developmental stage, by treating them as unprocessed fixed formulæ. To verify these hypotheses, learners with various L1 background were exposed to a 30-minute one-to-one Swedish L2 lesson, whose purpose was to provide them with favourable conditions to produce formulaic speech by repeating teacher utterances. Following the lesson, the informants took part in four communicative tasks, regrettably not described in the paper, structured in such a way as to ensure the elicitation of sentences which had not been heard during the lesson, thus representing creative output. This is defined as "structures which are not copies of the previous utterance", as note 24 informs

us. The results show that learners were able to repeat V2 structures following teacher input, but could not produce them spontaneously. Instead, in the relevant context, namely adverb fronting, they only produced the ungrammatical *Adv-SVO structure, which suggests that they were not developmentally ready to process V2. These findings are interpreted as proof that structures beyond the current processability stage can indeed be repeated as formulaic items without being processed. Hence Pienemann's scepticism with regard to repetition tests. In their response to Pienemann's critique of their work, Lantolf and Zhang (2015) note that the method used to elicit repetition in the Pienemann et al. (2013) study is quite different from the typical EI test. Specifically, in that experiment learners repeated teacher utterances "straight after a stimulus sentence has been presented" (note 24 again), whereas in Zhang and Lantolf's (2015) they were first asked to perform a comprehension task. This is a sensitive point, as the design of repetition tests has been shown to have a direct and macroscopic impact on the kind of data it can produce. This heated exchange of opinions is useful to understand that the repetition test often produces output which is interpretable in radically different ways. To minimise this risk, methodological rigour is essential.

Having provided this brief introduction to the complex theoretical and methodological debate concerning the repetition test, we now turn to a description of the task employed in the VILLA project.

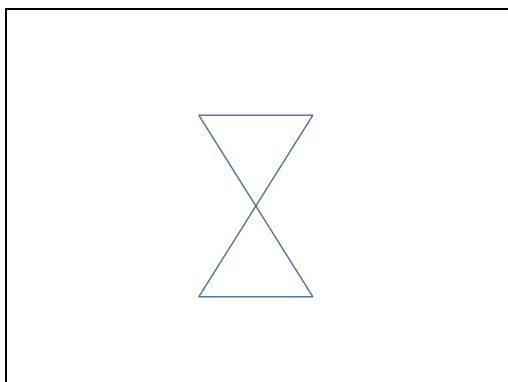
2.2. THE REPETITION TEST: STRUCTURE

The VILLA Repetition test is a highly structures test which learners took on two occasions, namely after 9 hs. of exposure to the input, to which we will refer as T1, and after 13:30 hs, which we label T2. It was administered individually on a computer screen; depending on the course edition, headphones or the integrated computer speakers were used.

The test works as follows. First, learners hear a short Polish transitive sentence, e.g. *dziewczynka ciągnie portugalkę*, "little girl-NOM pulls Portugues woman-ACC".

A simple geometric figure (Figure 1) subsequently appears on screen, and participants are required to draw it on a separate answer sheet.

Figure 1: Repetition test, distractors



The purpose of this task is to inhibit the learners' phonological memory, so as to make sure that they could not simply repeat a string of sounds, but rather had to process the target sentence in order to retrieve its meaning. However, this was not explained to the learners; as a result, some of them particularly focussed on this part of the test and took great care in drawing the figures as accurately as possible. Inevitably, this slowed down the pace of the test, and it is possible that it also negatively affected these learners' repetition performance as a result of the longer time elapsed between the presentation of the stimulus and their repetition.

In any case, once the figure has been drawn, the learners can repeat the target sentence as accurately as possible. Learner performances were not timed and no explicit time pressure was exerted.

Target sentences were 9 syllables long and had the same structure Noun - Verb - Noun. Throughout the test, the two nouns always appeared in association with each other²⁷. All targets were digitally recorded by the same female speakers. They were uttered with a slow speech rate and neutral intonation.

²⁷ One of the two nouns was classified as transparent (T), i.e. intuitively translatable, with some approximation, in every L1 of the VILLA project: e.g. *portugalka*, "Portuguese woman". The other noun was coded as non-transparent (NT), i.e. was completely opaque as to its meaning, e.g. *dziewczynka*, "little girl". Note that we will not consider this distinction in the rest of this work.

The latter point is extremely important, as prosodic contour are a powerful tool to express pragmatic meaning, in association with word order or simply instead of it, as exemplified in the preceding paragraph.

Each noun appeared in both the NOM and ACC case, instantiated by the endings *-/a/* and *-/e/ (<ę>)* respectively. As target sentences also varied with respect to constituent order, which could assume the values SVO or OVS²⁸, each pair of nouns appeared in four target sentences, which make it possible to isolate the parameters of case ending, word order and lexical transparency (Table 14). As there were 4 pairs of target nouns, the test included a total of 16 target sentences.

Table 14: Repetition test, target sentences

	SO	OS
NT - T	<i>dziewczynk-a woła portugalk-ę</i> little girl.NOM calls portuguese woman.ACC	<i>dziewczynk-ę woła portugalk-a</i> little girl.ACC calls portuguese woman.NOM
T - NT	<i>portugalk-/a/ woła dziewczynk-ę</i> portuguese woman.NOM calls little girl.ACC	<i>portugalk-ę woła dziewczynk-/a/</i> portuguese woman.ACC calls little girl.NOM

For the purposes of this study, target items are represented by each nominal ending taken in isolation, rather than by entire utterances. Each target item, therefore, may be described in terms of the three parameters 'target ending' (*-/a/* vs. *-/e/*), target sentence constituent order (SO vs. OS) and carrier word lexical transparency (T vs. NT). An example is presented in Table 15.

²⁸ As only the relative order of subject and object is relevant to our analysis, we will henceforth simplify SVO and OVS to SO and OS, respectively, unless explicitly stated otherwise.

Table 15: Repetition test, parameters of obligatory occurrences

<i>Kuchark-</i> cook	/e/ ACC	<i>woła</i> call 3SG	<i>Brazylijk-</i> Brazilian woman	/a/ NOM	
	-/e/ OS NT			-/a/ OS T	Target ending Constituent order Lexical transparency

"the Brazilian woman calls the cook"

The values assumed by the three parameters just discussed may combine in eight possible contexts (Table 16).

Table 16: Repetition test, combinations of parameters

target ending	-/a/	-/a/	-/a/	-/a/	-/e/	-/e/	-/e/	-/e/
constituent order	SO	SO	OS	OS	SO	SO	OS	OS
lexical transparency	T	NT	T	NT	T	NT	T	NT

Table 17 presents the paradigm of the feminine nouns in *-a*²⁹ employed in target sentences: shaded cells denote the cases not immediately relevant for the present study. Moreover, as no plural form was included in the test not the input, only singular forms are presented.

²⁹ Polish feminine nouns belong to two different inflectional classes, depending on whether their nominative ends in *-/a/*, like *żaba*, "frog", or in a consonant, like *noc*, "night". Only elements belonging to the former class are represented in the VILLA input.

Table 17: Paradigm of feminine nouns in *-a*

NOM	<i>dziewczynk-a</i>
GEN	<i>dziewczynk-i</i>
DAT	<i>dziewczync-e</i>
ACC	<i>dziewczynk-ę</i>
INS	<i>dziewczynk-ą</i>
LOC	<i>dziewczync-e</i>
VOC	<i>dziewczync-e</i>

The test also included 16 filler sentences: these were copula clauses with the structure 'NP (Neg) COP AP/PP', e.g. *Aleksander nie jest z Meksyku*, "Aleksander is not from Mexico". Finally, three warm-up sentences at the beginning of the test were administered to make sure that all learners had correctly understood the working of the test.

2.2.1. Transcription of the repetition test

Learner responses were digitally recorded in 44Khz .wav format using Zoom H4 digital recorders. These audio tracks were later recorded using a combination of the software ELAN (Brugman & Russell 2004) and CLAN (MacWhinney 2000).

Transcribing³⁰ early learning varieties is a particularly delicate operation (Saturno 2015c), as, on the one hand, they typically exhibit very conspicuous phonological deviation; on the other had, because of their fluid state, it is often impossible to distinguish what counts as simple phonological variability from contrasts reflecting an opposition in meaning (see also Bernini (2016) for a discussion of phonetics and phonology in the VILLA project). With this in mind, learner output was transcribed in broad IPA; in view of the specific object of this study, particular attention was paid to vocalic endings.

³⁰ The Italian, French and English corpora were transcribed by myself, the German corpus by Joanna Hinz and the Dutch corpus by Katarzyna Lozicka.

2.3. THE REPETITION TEST: RESULTS

2.3.1. Overview of learner output

The following transcripts³¹ present examples of correct (16) and incorrect (17) repetition of the case ending *-/e/* in identical target sentences, reported in (18) (see speaker and test time in the corresponding footnotes). As can be seen, in incorrect repetitions both nouns are marked as *-/a/*, on the model of the nominative case in the input, whereas in correct ones the two nouns are marked with different endings. Typically, each sentence presents one or no errors. Note that translations are not provided for the examples in (16) and (17) because we cannot be sure as to what the learner really meant. Similarly, the gloss only indicates the input word-form which is closer to learner output.

- (16) a. [artist-e pɔzdravja tumatʃk-a]³²
artist-ACC cheers interpreter-NOM
- b. [dʒɛvtʃjenka dʒɔnkje portugalk-e.]³³
little girl-NOM pulls Portuguese woman-ACC
- c. [brazilik-a vo'a kurark-e]³⁴
Brazilian woman-NOM calls cook-ACC

³¹ Orthography was normalised to standard IPA, although the original data employ various different formats as preferred by the individual transcribers. When lexical stress is not specified, it is assumed that it falls on the penultimate syllable. No effort has been made to accurately transcribe subtle phonetic contrasts in learner output, such as that between post-alveolar/retroflex {ʃ,ʒ,ʧ,ʤ} and pre-palatal {ç,ʒ,ʧ,ʤ} consonants, or that between high front /i/ and central /ɨ/ vowels. Regarding Polish so-called retroflex consonants, sometimes transcribed as {ʂ, ʐ, tʂ, sʐ}, we chose to transcribe them as {ʃ,ʒ,ʧ,ʤ} mainly for reasons of readability, although it can be argued that the notion of "retroflex" is quite slippery and corresponds to many rather different phonetic realisations (Hamann 2002a; 2002b; 2003; 2004; Zygis & Hamann 2003; Zygis 2003; Żygis & Padgett 2010).

³² 5104, T2.

³³ 5118, T1.

³⁴ 1119, T2.

- (17) a. [artistk-a posdravja tumatʃk-a]³⁵
 artist-NOM cheers interpreter-NOM
- b. [dʒevtʃinkn-a hm hm portugalk-a]³⁶
 little girl-NOM omission Portuguese woman-NOM
- c. [brazilik-a wave kukark-a]³⁷
 Brazilian woman-NOM calls cook-NOM
- (18) a. /artistk-e pozdravja twumatʃk-a/
 artist-ACC cheers interpreter-NOM
 "the interpreter cheers the artist"
- b. /dʒevtʃinka tɕɔŋnje portugalk-e/
 little girl-NOM pulls Portuguese woman-ACC
 "the little girl pulls the Portuguese woman"
- c. /brazilijk-a vɔwa kuxark-e/
 Brazilian woman-NOM calls cook-ACC
 "the Brazilian woman calls the cook"

The same may happen with the repetition of *-/a/* NOM, although, as will be detailed later on (section 2.6, p. 82), this is a much rarer event. The substitution of target *-/a/* with the competing ending *-/e/* produces a sentence in which both nouns are marked as *-/e/* (19a; target in 19b).

- (19) a. [brazilijk-ε vowa vowa kuxark-ε]³⁸
 Brazilian woman-ACC calls calls cook-ACC
- b. /brazilijk-e vɔwa kuxark-a/
 Brazilian woman-ACC calls calls cook-NOM
 "the cooks calls the Brazilian woman"

³⁵ 2117, T2.

³⁶ 4122, T1.

³⁷ 3118, T1.

³⁸ 4113, T1.

Occasionally, sentence with two errors may occur, in which case endings are swapped (20a; target in 20b). This kind of output, however, is decidedly rare.

- (20) a. [brazilijk-a vowa kuxark-e]³⁹
 Brazilian woman-NOM calls cook-ACC
- b. /brazilijk-e vɔwa kuxark-a/
 Brazilian woman-ACC calls cook-NOM
 "the cook calls the Brazilian woman"

The production of centralised vowels by some learners requires particular attention. Target items in which the ending produced was not clearly identifiable as either *-/a/* or *-/e/* were excluded from the analysis, as it was risky or outright impossible to establish which one the learner meant to produce. Altogether, 334 items were eliminated for this reason. The problem is particularly severe in the case of the German data, which account for 242 of all excluded items. This is because German has a phonological rule of word-final vowel centralisation, which causes vowels in that position to be pronounced as [ə]. This articulatory habit is often transferred to L2 output as well, as can be seen in (21). The respective targets are provided in (22).

- (21) a. [portugalk-ə tʃongje dʒevtʃink-a]⁴⁰
 Portugues woman-? pulls little girl-NOM
- b. [kuxarka vowa brazilijk-ə]⁴¹
 cook-NOM calls Brazilian woman-?

³⁹ 2104, T1.

⁴⁰ 4108, T2,.

⁴¹ 4106, T2.

- (22) a. /portugalk-e tɔŋnje dʒvtʃɪŋka/
 Portugues woman-ACC pulls little girl-NOM
 "the little girl pulls the Portuguese woman"
- b. /kuxark-a vɔwa brazilijk-e/
 cook-NOM calls Brazilian woman-ACC
 "the cook calls the Brazilian woman"

2.3.2. Repetition test: descriptive statistics⁴²

Mean scores are summarised in Table 18, aggregated by target ending, word order, test time and L1⁴³.

Table 18: Repetition test, mean accuracy scores

	-/a/				-/e/			
	OS		SO		OS		SO	
L1	T1	T2	T1	T2	T1	T2	T1	T2
FR	0.99	1.00	0.98	0.98	0.16	0.26	0.34	0.39
GE	0.99	0.98	0.98	0.96	0.57	0.74	0.62	0.69
IT	0.95	0.88	0.96	0.93	0.55	0.69	0.83	0.87
NL	0.88	0.9	0.92	0.90	0.33	0.39	0.48	0.53
EN	0.99	0.99	0.96	0.89	0.09	0.07	0.04	0.06
Mean	0.96	0.95	0.96	0.93	0.34	0.43	0.46	0.51

There is an obvious difference between the repetition accuracy of -/a/ and -/e/. The former is very close to 100% in all cases; the latter is much lower, with a

⁴² Unless otherwise stated, all quantitative analyses, statistics and graphics have been realised using R (R Core team 2014).

⁴³ The English data set for the Repetition test only comprises 16 learners, instead of 17 as in the Comprehension test (Chapter 3). This was deemed necessary because only the audio track at T2 is available for learner 3117. For this reason, we considered it safer to discard the learner in question from this as well as all subsequent analyses.

minimal score as low as as 9%. The significance of the ending required as a predictor of accuracy, thus, is rather clear.

For this reason, we proceed to separate the data set and analyse the processing of the endings *-/a/* and *-/e/* individually. There are two reasons for doing so: first, we assume that their processing may rely on very different strategies: in other words, that errors in the processing of *-/e/* are due to different reasons than errors in the processing of *-/a/*. The ending *-/a/*, modelled on the nominative case, appears the basic word form of lexical items, so that not only is it not hard to reproduce, but it tends to automatically overextend even where it is not due. Its counterpart *-/e/*, in contrast, is a marked ending which only acquires meaning in the context of a system of morphological oppositions, which is not found in most of our learners' L1s. As a result, we expect that it cannot be supplied automatically, at least at such early stage, and that on the contrary it requires a conscious effort on the side of the learner to produce.

The second reason is more technical, and is concerned with the scoring adopted in our data set. Recall that each case ending was considered as a separate target, whose correctness was decided based on whether or not the ending supplied by the learner matched that required by the stimulus sentence. As a result, we only have two values for our response: either correct (1) or wrong (0). As detailed above, all cases in which the learners produced an ending different from either *-/a/* or *-/e/*, or indeed failed to provide any ending at all, were eliminated from the analysis. Since the correctness of the response is based on the ending required by each target, *-/a/* and *-/e/* are in effect separate in the data set, so that incorrect processing of *-/e/* is not related to the processing accuracy of *-/a/*. For this reason we will report separate analysis for the two target endings, beginning with the latter.

2.4. REPETITION OF *-/e/*

Descriptive statistics for the repetition accuracy of *-/e/* are provided in Table 19 (OS targets) and Table 20 (SO targets).

Table 19: Repetition test descriptive statistics, -/e/ ending, OS targets

L1	T1 mean	T1 sd	T1 responses	T2 mean	T2 sd	T2 responses
FR	0.16	0.37	119	0.26	0.44	121
GE	0.57	0.50	110	0.74	0.44	111
IT	0.55	0.50	129	0.69	0.47	131
NL	0.33	0.47	155	0.39	0.49	157
EN	0.09	0.29	112	0.07	0.25	118

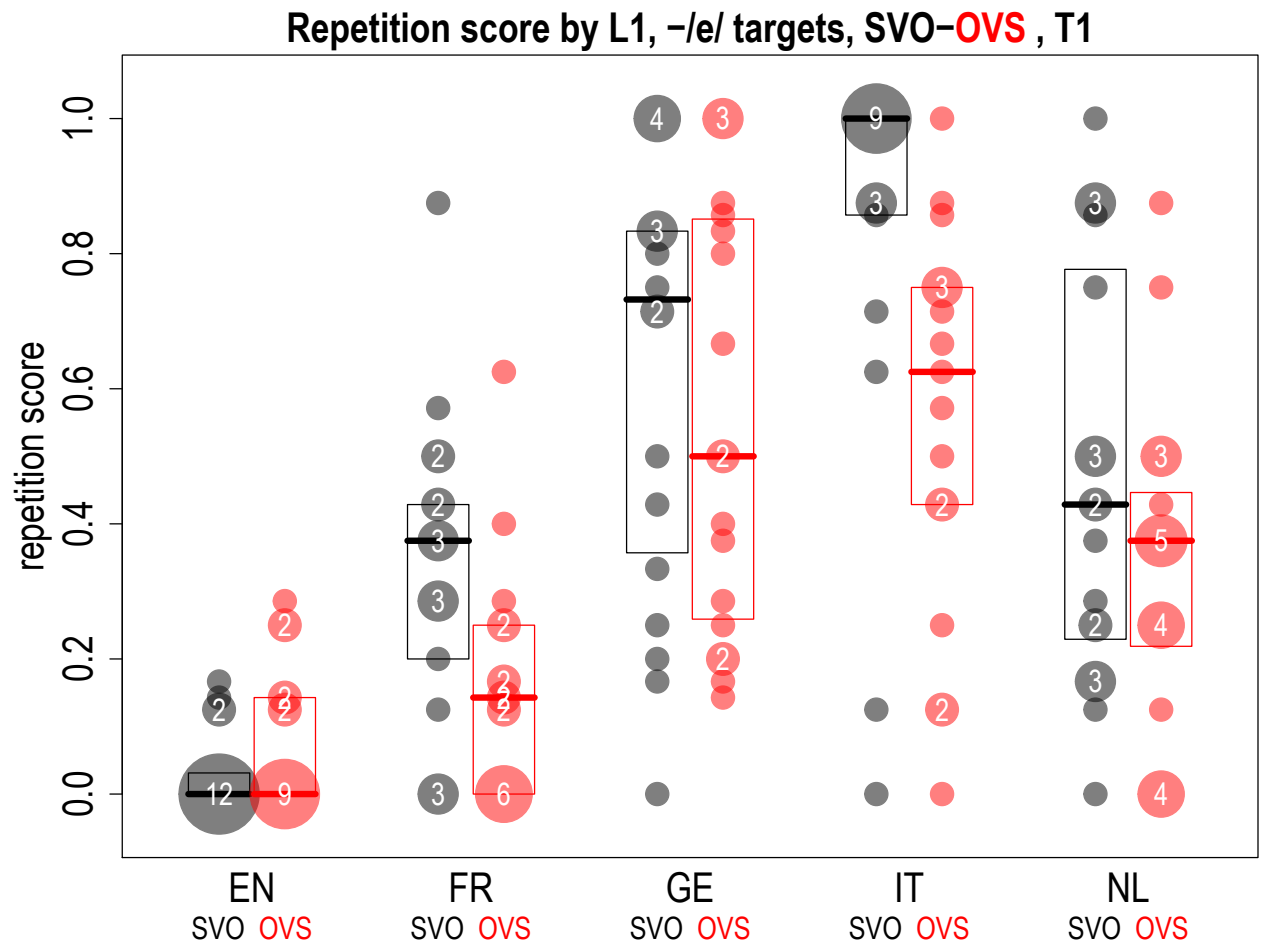
Table 20: repetition test L1 descriptive statistics, -/e/ ending, SO targets

L1	T1 mean	T1 sd	T1 responses	T2 mean	T2 sd	T2 responses
FR	0.34	0.48	123	0.39	0.49	126
GE	0.62	0.49	82	0.69	0.46	72
IT	0.83	0.37	126	0.87	0.34	127
NL	0.48	0.50	147	0.53	0.50	146
EN	0.04	0.19	108	0.06	0.24	118

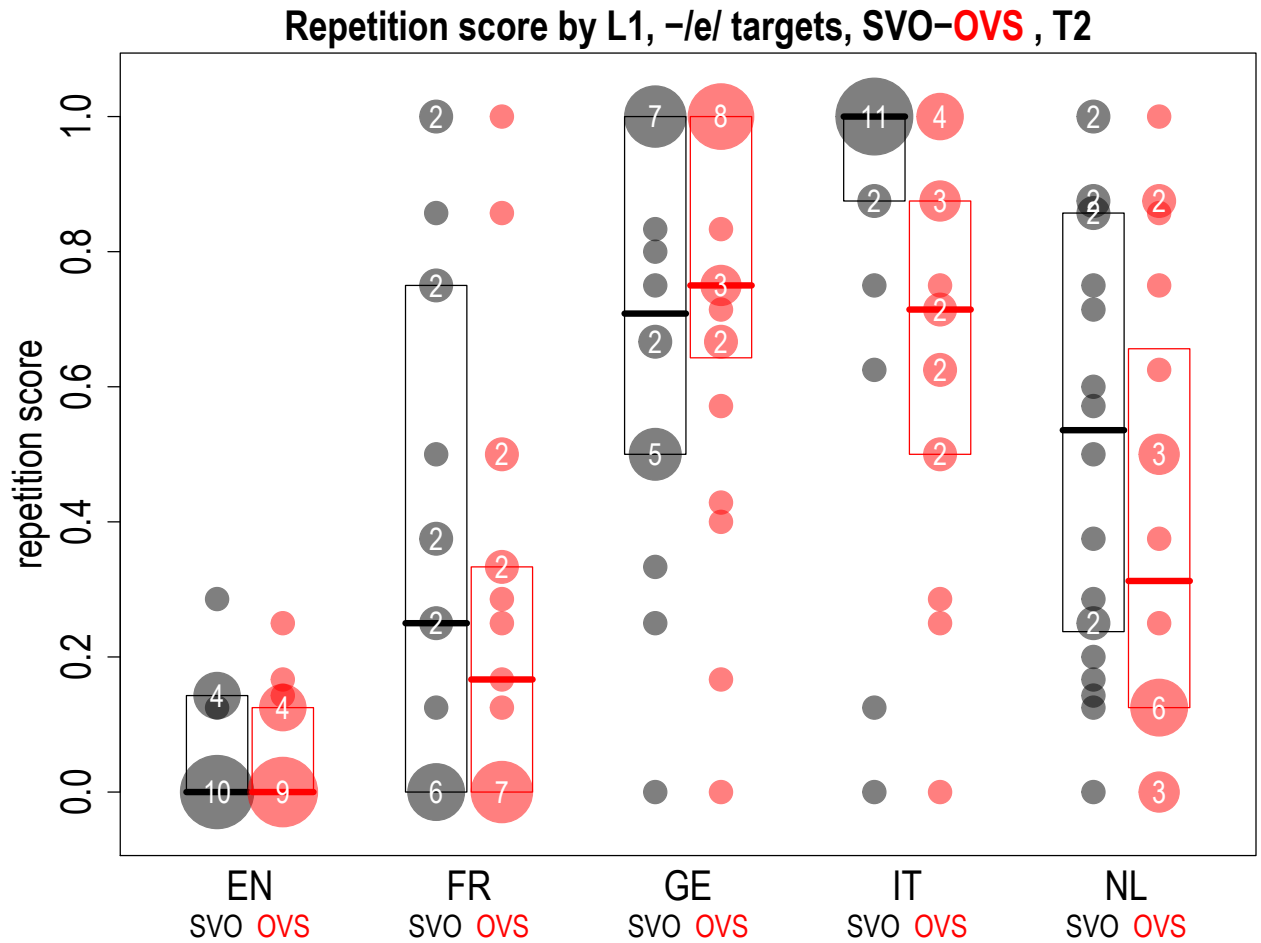
Mean scores inform us that SO targets are generally processed with greater accuracy than OS ones, although scores remain rather low - below 50% in most cases. L1 mean scores vary greatly, suggesting that there may be an important influence of the native language. Finally, standard variation is very high, which indicates that learners perform very differently from each other.

Information as to the performance of individual learners is presented graphically in the boxplots of Graph 1 (T1) and Graph 2 (T2).

Graph 1



Graph 2



The English L1 group consistently exhibits the poorest results, with most learners scoring exactly 0%. No learner in this group ever scored over about 30%. In contrast, the Italian group has the highest scores, followed by the German group and by the Dutch and French, somewhat behind. In all groups except the English and the French (at T1 only), at least one learner managed to reach 100% accuracy.

Individual variability within the L1 group is extremely high, with learners registering scores sometimes even as radically diverse as 0% and 100%. The only exception, again, is the English group, in which scores are consistently low.

2.4.1. Repetition of *-/e/*: inferential statistics

After this brief overview, we proceed with inferential statistics. We fit a generalised mixed linear model with binomial error structure and logit link function: fixed effects comprise the L1, word order and time as linear predictors, as well as their two-way interactions: L1:word order, L1:time, and time:word order. Random effects across participants include random intercepts as well as random slopes for word order⁴⁴. The rationale for including the interactions is as follows: we hypothesise that the learners' ability to correctly repeat *-/e/* is influenced by the word order of the target sentence, SO targets generally facilitating repetition, and OS targets hindering it. However, the extent of this word order effect is also influenced, sometimes positively, sometimes negatively, by the L1 of the learner. Finally, further exposure is generally beneficial to repetition, but the extent to which results improve between T1 and T2 is also determined by the learner L1: speakers of certain languages may improve more markedly than speakers of other languages. Finally, the effect of time is also constrained by word order: within the same L1 group, one of the two values of this factor shows greater improvement over time. The results of an ANOVA applied to the model produces are presented in Table 21.

⁴⁴ Ideally, a more complete model would also have included random slopes for time across participants, as well as a correlation between random slopes and intercept. However, this model failed to converge with the available data set. The model presented in the text is therefore the most complete which the authour could fit to the data.

Table 21: Repetition test linear model, ANOVA

Fixed effect	Df	Sum Sq	Mean Sq	F value
L1	4	92.32	23.08	23.08
Word order	1	39.41	39.41	39.41
Time	1	15.49	15.49	15.49
L1:Word order	4	30.49	7.62	7.62
L1:Time	4	2.90	0.72	0.72
Word order: Time	1	1.99	1.99	1.99

This was compared to an identical model, the only difference being the lack of interactions between L1, word order and time. ANOVA between the two models shows that they are significantly different ($\chi^2 = 37.29$, $df = 12$, $p < 0.01$). We therefore conclude that the hypothesised interaction is indeed relevant in accounting for the data.

2.5. A DIFFERENT PERSPECTIVE

These results, however traditional in their statistical approach, seem to us quite hard to interpret linguistically. We might take the influence of the L1 as an example. The analysis has shown that certain L1s seem to raise the average repetition accuracy when compared to other languages. Table 18, for instance, tell us that at T1, French speakers scored on average 0.16 on the repetition of /e/, whereas the Italians scored 0.55. From this we conclude that Italian L1 has a positive effect on processing accuracy. Now, the question is what it means to have an accuracy of 0.55 instead of 0.16. This is a mean value, which results by averaging the results of the whole group; Graph 1 and Graph 2 indicate that individual learner performance may vary greatly even within the same L1 group, ranging from 0 to 1 through a wide variety of intermediate scores, i.e. results which witness to neither complete success nor failure. What do such scores tell us regarding the real processing abilities of our learners? An accuracy score of 0.55 might indicate that learners correctly processes OS

sentences about half of the time; does it mean that they interpret the rest as SO? We should then picture these speakers as they make processing errors in oral interaction, guessing the syntactic structure of the utterances addressed to them on a near-chance basis. But then we are told that the probability of getting the meaning right is augmented by, say, 10% if they speak a particular L1. Those speakers now understand correctly about 65% of the sentences they are addressed, all things being equal. This is no longer a fifty-fifty ration, but it is still quite far from a well-defined strategy: we could perhaps argue that the learners are still guessing, with a slight bias towards the correct interpretation. Can we take this as an indication of greater success determined by the L1? Statistically speaking, if chance can be excluded, surely yes; in linguistic terms, however, it seems to us that there is little difference between 55% and 65% processing accuracy. The real question is whether or not an intermediate accuracy score on a structure test has a meaningful linguistic interpretation: we feel it does not. Our hypothetical learners are not applying a rule consistently: guessing with a slight bias is not a strategy, and research on even the earliest stages of acquisition has shown that there is always a structure in learner varieties, however initial. Whatever the principle, it is used consistently, until it is replaced by another one, possibly closer to the target variety. In sum, when we ask whether a learner can process L2 morpho-syntax we should not accept a percentage as a response, but we should strive for a yes or no answer. Either he does, or he does not, in which case he probably relies on some other principle, like a default constituent order. So we argue that the influence of the L1 should be felt not simply in the proportion between correct and wrong answers, which can produce any value between 0 and 1. Instead, it is realistic to think that a certain L1 might help its speakers to identify and consistently apply a given rule of the target language, shifting their processing behaviour from a provisional, creative strategy to a consistent principle, solidly founded on the input and closer to the rules of the target language. In short, the results should show that more learners rely on a target-like principle, rather than higher average accuracy.

The conclusion that learners are applying a systematic principle involves their reaching a threshold, as they may not perform with 100% accuracy on L2 structures. They might, however, process that structure correctly in most cases, save peculiar contingencies prevent it - noise, distraction, fatigue and the like. In the following section we will attempt to identify the role of our three parameters time, word order and L1 not simply in raising the average score, but in reaching a certain meaningful threshold allowing us to state that the learner is applying a consistent morphosyntactic principle.

To start, we postulate three scenarios:

- a) learners rely on an uninflected word-form of lexical item, with no morphological variation;
- b) learners have noticed some morphological variation in the input, but they cannot yet account for it with a productive rule. Therefore, these learners will supply the basic and inflected word-forms with no apparent regularity.
- c) learners regularly produce correctly inflected word-forms.

Scenario b) is what we might call chance performance, equivalent to guessing. With only two values to choose from (NOM *-/a/* and ACC *-/e/*), accuracy rates should be around 50%. Scenario a) is *below* chance: learners who behave in this way are not guessing, but applying a systematic principle, which, alas, is not compatible with the target language and thus produces accuracy rates tending to 0%. Specifically, this principle maintains that all feminine nouns, independently of their syntactic function, end in *-/a/*, and that syntactic functions are veiculated by the position of a noun in the utterance.

Finally, scenario c) is *above* chance: learners systematically apply a principle of case marking which is coherent with the rules of the target language.

All we need now is a statistical test which allows us to decide which scenario a learner belongs to. This is obtained by calculating the probability of observing a given result, in the form of a ratio between correct responses and total responses, on the basis of a statistical distribution whose mathematical

properties are known. We choose the binominal distribution, which describes the probability of obtaining either of two values (conventionally 0 and 1) out of a given number of trials, as in the throwing of a coin. Statistical tests based on this distribution allow us to answer such questions as "what is the probability of obtaining six heads if I throw my coin eight times?". If the probability is too small, conventionally below 5%, I may conclude that the coin I am throwing is not fair, that is, that something is biasing it towards a particular result in a consistent manner. In our experiment, the same question would sound as follows: "what is the probability that a learner, without applying a morphosyntactic principle, produced six instances of correct case-marking over eight trials"? Again, if the probability is too small, we should conclude that the learner is in fact applying the morphosyntactic principle. If not, on the basis of the same probability value, we can further tell whether he is relying on word order (below chance-level results) or merely guessing (chance-level results).

When applied to our context, this framework necessarily implies some simplification, as the outcome of the trial, being binary, may only assume two mutually excluding values: 0/1, true/false, correct/incorrect, *-a/-e*. Learner output may be more varied than this, as lexical items may take more than only two word-forms; the analysis presented in the preceding paragraph, however, showed that the great majority of learner-produced forms are instantiations of either *-a/* or *-e/*, so that the binary model does not appear so inadequate, after all. The test allows us to verify whether or not *-e* endings occur in such a proportion that their correct use cannot be due to chance alone, but on the contrary it must reflect a morphosyntactic strategy. For this reason, we consider *-e/* as the correct response, and everything else as a wrong one.

2.5.1. Guessing in the Repetition test

Effectively, the procedure we are going to apply consists in determining whether or not our learners developed a consistent morphosyntactic strategy,

the alternative being that they applied a positional principle based on default word order or that they performed the test by guessing. A word of caution is necessary here. It could be objected that the last scenario is not particularly realistic. Such strategy is not hard to imagine in the Comprehension test analysed later on in this chapter, in which the learner simply has to select either of two pictures. If the learners paid no attention at all to the target sentence, and only chose pictures through guessing, then each picture would have exactly 50% probability of being selected. The situation is different in the Repetition test, in which participants are required to actively produce output. In fact, one could argue that the two possible answers, namely *-/a/* and *-/e/*, do not have equal probability of being chosen. This is because *-/a/* is the unmarked, basic word-form of lexical items, and as such it should tend to spontaneously come up in all contexts, whether appropriately or not. Repeating *-/e/*, in contrast, requires a conscious effort on the side of the learner, and should therefore mirror a rational strategy. We will briefly show that we cannot always be so sure of this, and the same is true for repetitions of *-/a/*. But there is yet another objection against the possibility of guessing in the repetition test: this strategy implicitly relies on the assumptions that there are only two options between which one can choose. However, learners do not necessarily know that it is so, especially if one considers that the test also included numerous different syntactic structures as distractors. The options then would not be limited to two elements, each having 50% probabilities of being selected. For this reason, it should be unlikely that learners were just guessing.

While this all is true, in principle, the reality is a little different. The unmarkedness of *-/a/* certainly plays a role in explaining the extreme results scoring around 0% in the repetition of *-/e/*. The intermediate scores which we do observe, however, fit into this picture less well. Regarding the argument that if learners were guessing, then they would produce more than just two endings, learner output shows that the cases in which learners produce an ending other than *-/a/* or *-/e/*, excluding centralised *-/ə/*, are very rare. In

addition, target sentences only comprised *-/a/* or *-/e/* in target position. If one excludes the extreme, rather straight-forward case of participants scoring extremely low or extremely high, one could imagine that the average learner hears a sentence, notices - or knows - that the two nouns end one in *-/a/*, the other one in *-/e/*, but fails to understand the syntactic structure of the sentence. Effectively, this means that he does not understand, remember or re-construct which ending is attached to which noun. Two strategies are now possible: one could either repeat the unmarked ending *-/a/* on both nouns, or just guess. The first alternative seems to be more common, perhaps because the test exerted considerable pressure on the learners. The latter may explain results which are otherways a little trickier. Finally, it could also provide a context for the role of word order. One could think, for example, that the unusual structure of OS sentences makes it harder to remember which ending attaches to which noun, and thus augments the probability that the learner will resort to guessing.

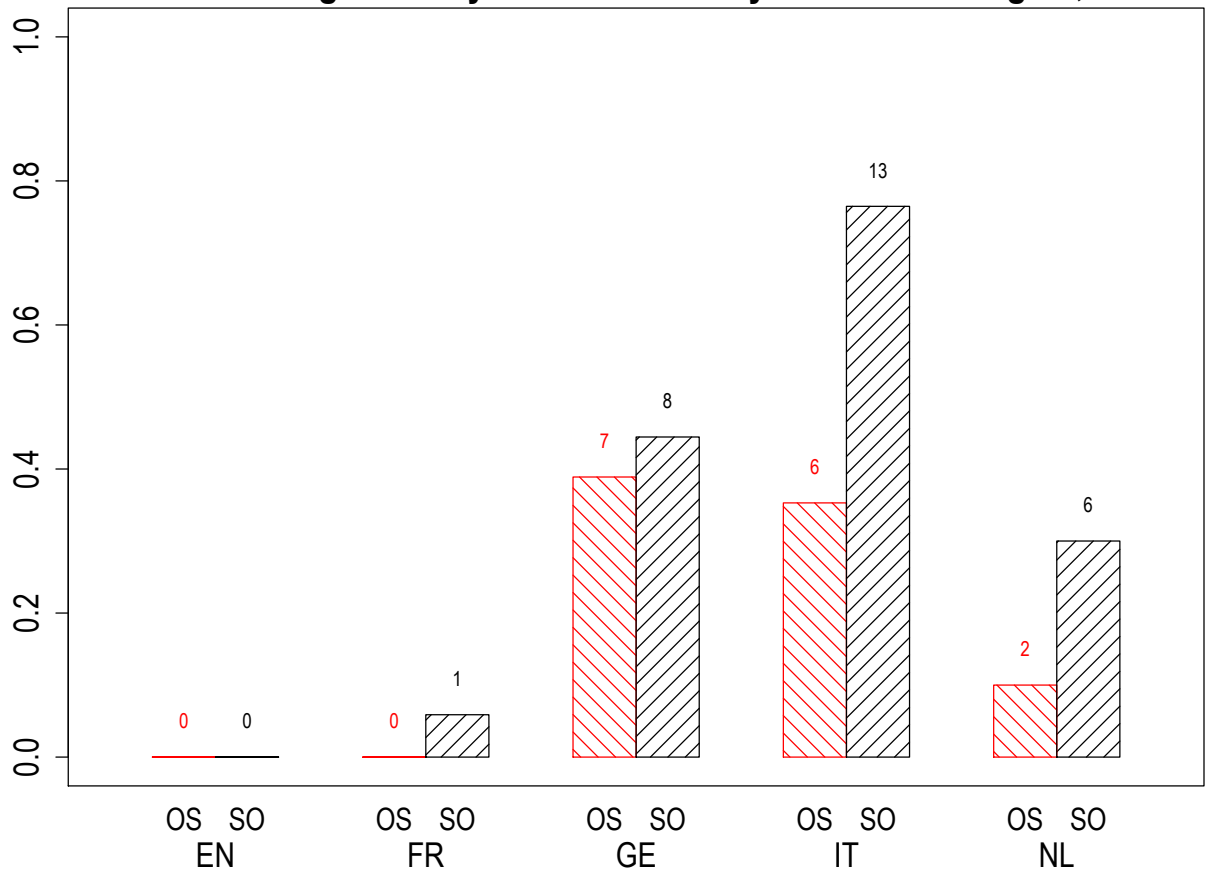
2.5.2. Individual strategies of input processing: descriptive statistics

After this brief introduction, we wil now attempt to establish whether the L1, word order and time impact on the learners' ability to perform above chance on the repetition test. Again, we first focus on the processing of the marked accusative ending *-/e/*, which is the most informative for our research questions on the development of morphosyntax.

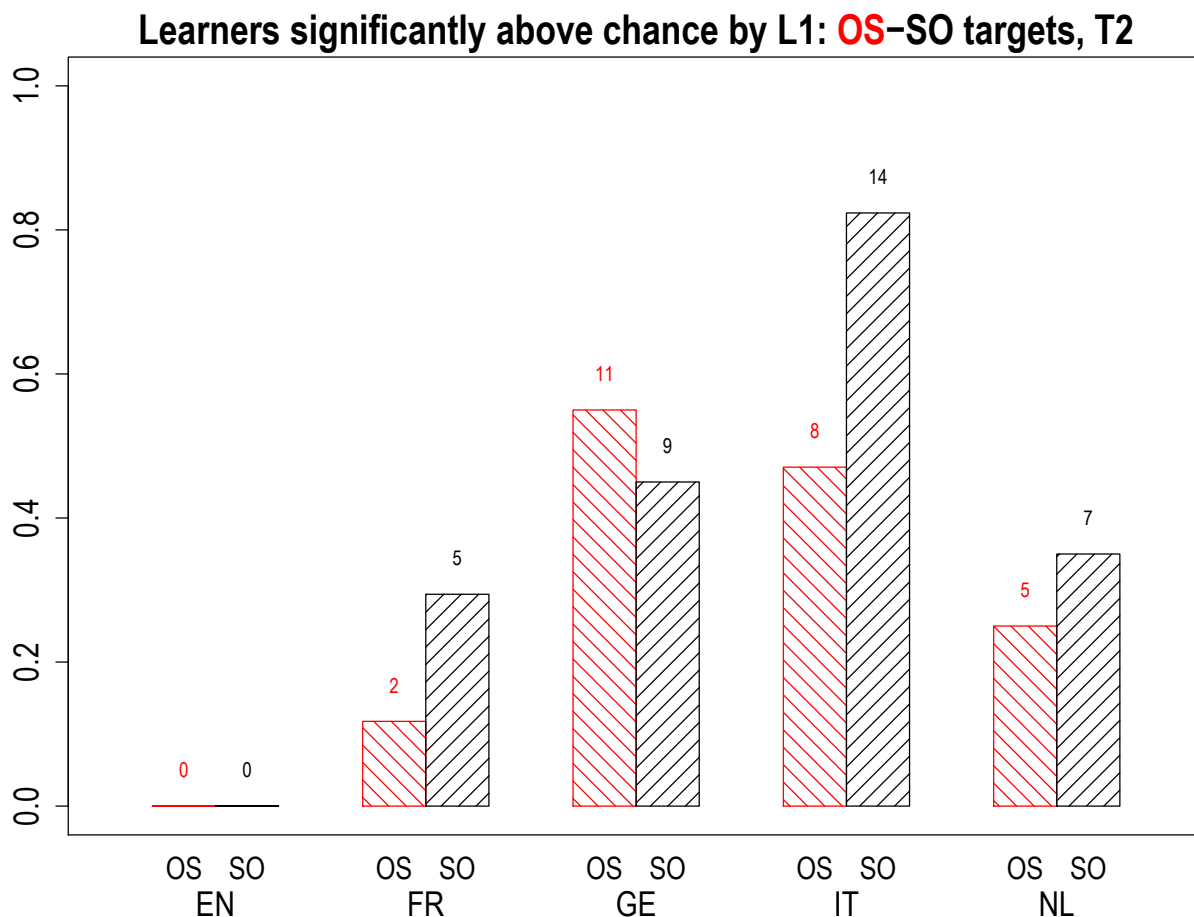
The probability that a learner obtained his set of responses under chance conditions is calculated based on a binomial distribution described by the number of correct and incorrect responses, with an arbitrary chance level set of 50%. The results are summarised in Graph 3 and Graph 4 for T1 and T2 respectively. The height of the bars indicate the proportion of learners performing above chance, whereas the digits above them specify the actual number.

Graph 3

Learners significantly above chance by L1: OS-SO targets, T1



Graph 4



As expected, learners tend to perform above chance more often on SO targets than on OS ones, the only minor exception being the German L1 group at T2. The difference between the two word order values is particularly noticeable in the Italian L1 group, whereas the others exhibit less dramatic effects. Regarding Time, not surprisingly scores improve from T1 to T2. As far as the effect of the L1 is concerned, the first and most noteworthy information is that no English L1 learner performs above chance at either test time. Besides that, a superficial analysis seems to tease apart two groups. One, characterised by lower scores, comprises the Dutch and French learners, whereas the other one, performing better, is made up of German and Italian learners. We now proceed to inferential statistics to verify our observations we have made.

2.5.3. Individual strategies of input processing: inferential statistics

We fit a generalised linear model with binomial error structure and logit link function. Again, we include Time, Word Order and L1 as linear predictors, as well as their two-way interaction, namely Time:Word Order, Time:L1 and L1:Word Order. The difference with the model we fitted in the previous section lies in that our response now is represented by the proportion of learners who perform above chance. This way our response is not provided by minor, often negligible differences in mean accuracy score, but in the number of learners who come to behave according to a rational morphosyntactic strategy. ANOVA applied to this model produced the results summarised in Table 22:

Table 22: Repetition test, individual processing strategies linear model: ANOVA

	Df	Deviance	Resid. Df	Resid. Dev	Pr(> Chi)
NULL	NA	NA	15	70.13	NA
L1	3	44.80	12	25.32	< 0.01
Word Order	1	8.53	11	16.78	< 0.01
Time	1	4.74	10	12.04	0.03
L1:Word Order	3	7.15	7	4.89	0.07
L1:Time	3	3.27	4	1.62	0.35
Word Order:Time	1	1.05	3	0.56	0.30

As can be seen, most of the variance is explained by the linear predictors L1 and Word order, whose interaction only slightly fails to reach significance, too. Time is also a significant predictor, as could be expected, but does not engage in significant interactions.

In sum, these results tell us that the ability to correctly repeat the case ending -/e/ depends primarily on one's native language, and secondarily on the word order of the target sentence. In addition, and rather unsurprisingly, the longer you are exposed to the language, the better you get at the task.

2.5.4. Repetition of -/e/: a comprehensive picture

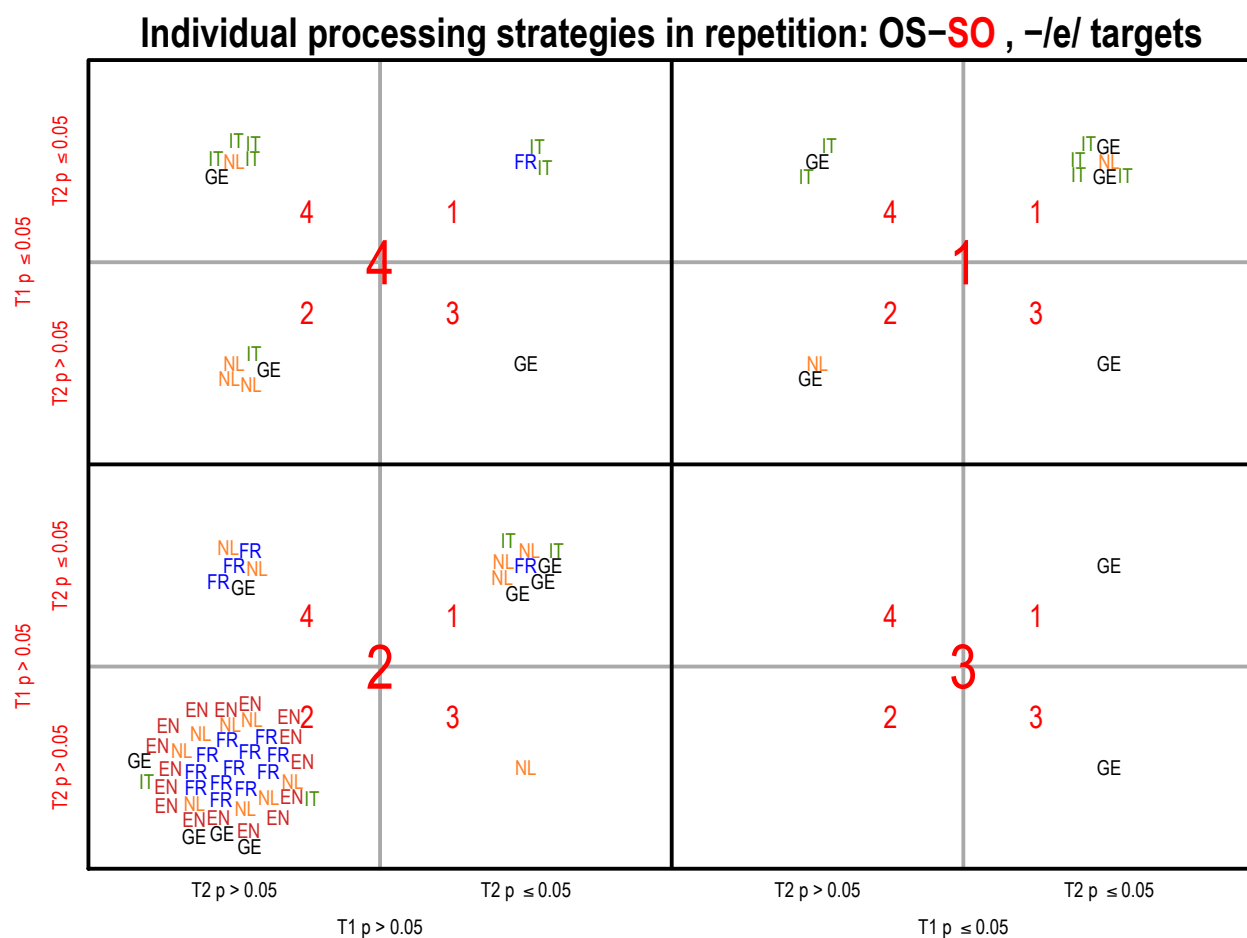
We now turn to the last piece of the complex mosaic we have been composing for the Repetition test. This is a global picture of the processing strategies developed by each learner, together with their evolution over time. Truly, we have shown that there are certain differences between T1 and T2, and SO and OS; however, even if we ignore mean scores and focus on the number of participants behaving significantly above chance under specific conditions, we have no information as to what each individual learner does at T1 *and* T2, or on SO *and* OS targets. To remedy, in Graph 5 each learner is synchronically described in terms of his performance on both OS and SO targets, at T1 and T2. The graph should be read as follows.

The main area is divided into four main squares, representing the situation at T1. Learners appear along the axes depending on whether their performance differs significantly from chance (upper half of the axis) or not (lower half). The horizontal and vertical axes represent OS and SO targets respectively.

Each main square is further divided into four smaller ones, which replicate the situation at T2 with the same logic. The combination of the square in which the learners lies at T1 and T2 (in this order) determines the *scenario* in which they fall. By scenario 3:1, for instance, we mean the large square 3 at T1 and the small square 1 at T2.

In determining learner processing strategies and their evolution over time, one should proceed as following. First, identify in which main square the learner is found. If, for example, a learner is in main square 1, that means that at T1 his output differed significantly from chance on both OS and SO targets. Then look at the smaller square in which the learner lies. If, in our example, it is square 3, then we know that at T2 the output of that same learner was still different from chance on OS targets, but no longer so on SO ones. There is only one learner exhibiting such surprising behaviour, belonging to the German L1 group.

Graph 5⁴⁵



The graph can be used to place SO and OS word orders in a hierarchy. At both test times, squares 1 and 2 represent extreme cases, in which both types of target are processed equally above or at chance level⁴⁶, respectively. Squares 3 and 4, in contrast, indicate a difference in the processing of the two word orders. It is not unexpected that square 3, in which learners behave above chance on OS, but not SO targets, only comprises 2 learners at T1. The opposite scenario, square 4, reunites 15 learners at T1. At T2, another 6 learners moved to square 4 at T2 from square 2 at T1 (scenario 2:4), indicating that they

⁴⁵ This graph was created using the statistical software R (R Core team 2014) and the packages *wordcloud* (Fellows 2014) and *extrafont* (Chang 2014)

⁴⁶ This includes below-chance accuracy, too, which corresponds to a positional principle. While we suspect that this, not guessing is indeed the most common processing strategy across learners who fail to develop a morphosyntactic principle, the statistical tests we used does not distinguish between the two.

improved on SO targets alone. Only one learner improves on OS alone (scenario 2:3). It seems, therefore, that if one value of word order is easier to process, or improves earlier than the other one, then in most cases it is SO.

It is worth noticing, however, that the vast majority of participants at T1 is found in scenario 2, indicating chance-level behaviour on both target types.

The graph also allows us to study the evolution of processing strategies in repetition over time, depending on the word order of the target sentence. The first observation we make is that for most learners, there is no evolution whatsoever. The bulk of the data set (42 learners on 88) can be found in scenario 2:2, which indicates chance behaviour under all conditions (OS and SO targets, T1 and T2 alike). This group includes all English L1 learners, most French, about a half of the Dutch, and only few Italians and Germans. Conversely, 7 learners can be found in scenario 1:1, which indicates the presence of a morpho-syntactic processing strategy all the way from T1 to T2 on both OS and SO targets. Finally, the 6 learners in scenario 4:4 were able to process SO, but not OS targets at T1, and this does not change with time.

Then there are learners who improve from T1 to T2. Some change towards more target-like processing strategies: this is the case of scenarios 2:4, 2:3 and 2:1, in which one finds learners who improved on SO, OS, or both target types, respectively. Some seem to move away from the target variety: learners in scenario 4:2 processed SO targets above chance at T1, but no longer do so at T2. Other surprising, but rare cases can be found in scenarios 1:3, 1:2 and 1:4. These learners were able to process everything at T1, but got worse at T2. There might be various explanations for this rare and apparently illogical behaviour. Most probably, these learners had a rather border-line score in the first place, so that only slightly different scores determined their being on either side of the threshold. It could also be that they were just tired or bored, or had stayed out until late the night before⁴⁷; perhaps their headphones did not work very well, or the traffic was particularly loud at the time when they

⁴⁷ At least one such case was indeed documented!

took the test. This all, of course, we cannot control, but we cannot exclude either.

If we suppose that environmental conditions were acceptable, though, and that learners were in a reasonably lucid state, then we could try and find an explanation fitting into the picture of linguistic development. We could argue that by the second test time, learners were paying attention to something else as part of their input processing. For instance, one could hypothesise that at T1 their high repetition scores were due to good listening skills, which, contrary to the claims of most of the relevant literature, did not correspond to morpho-syntactic abilities. In other words, these learners were good at segmenting, remembering (perhaps through quiet rehearsing) and reproducing sounds, but these had no particular linguistic meaning to them. In sum, they were not processing target sentences for meaning, but simply repeating them. By T2, perhaps, they started processing meaning too, according to the strategies typically employed by initial learner varieties. That is to say, they were no longer parroting sounds, but finally repeating meanings in their own way. In this last case, it could well be that an initial learner, when trying to understand and to say something, will not pay attention to inflectional endings, but focus on word order according to the principles of utterance organisation identified in the pre-basic and basic variety.

2.6. REPETITION OF -/a/

The data set concerning the repetition of -/a/ is too heavily unbalanced to perform reliable statistical tests. This is because of the evident ceiling effects observed in all language groups: recall that -/a/ is the basic word-form of the lexical items present in the test, so that learners not only have no difficulty in producing it, but also greatly overextend it to targets requiring -/e/. However, apart from the slightly different distribution across L1s, it is interesting to observe that some learners failed to repeat -/a/ in all cases in which it was required. Since we excluded from the analysis any output different from either

-/a/ or -/e/, failure to repeat -/a/ necessarily means that -/e/ was produced. As we have argued so far, this ending is marked and proves difficult for learners to repeat; indeed, some never got to process it correctly throughout the whole duration of the experiment.

Table 23 and Table 24, as well as Graph 6 as Graph 7, allow us to obtain a more concrete picture of the magnitude of the phenomenon. Two observations can be made: the first one is that accuracy on -/a/ repetition is not 100%, as one would expect. The second observation is that accuracy scores *decrease* over time. No doubt, the two phenomena must be connected, and one may add the additional fact that accuracy scores for the repetition of -/e/ *increase* over time⁴⁸.

Table 23: repetition test L1 group scores, -/a/ ending, OVS targets

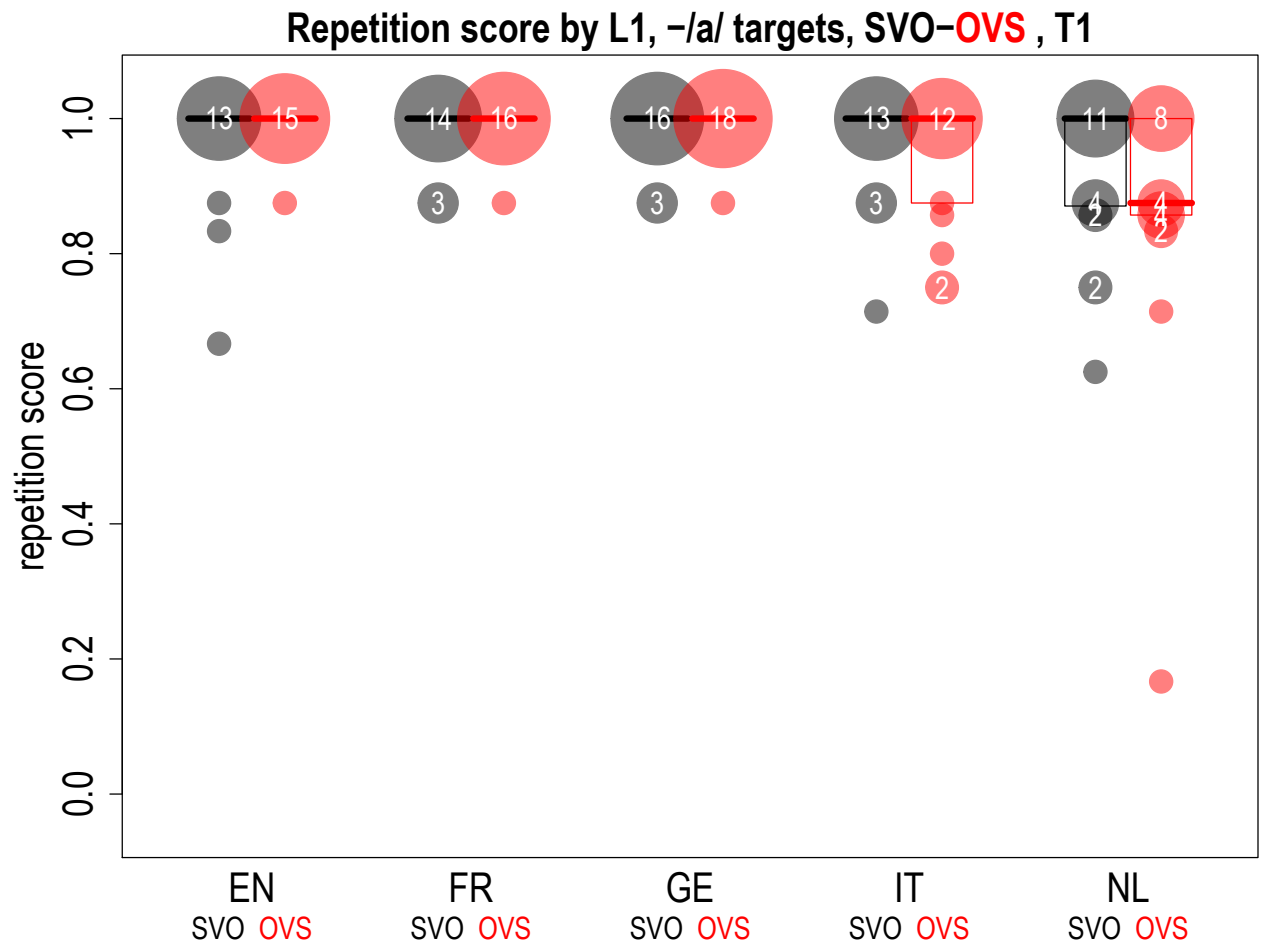
L1	T1 mean	T1 sd	T1 responses	T2 mean	T2 sd	T2 responses
FR	0.99	0.09	130	1.00	0.00	133
GE	0.99	0.08	141	0.98	0.14	155
IT	0.95	0.23	128	0.88	0.32	128
NL	0.88	0.32	145	0.90	0.30	150
EN	0.99	0.09	116	0.99	0.09	119

Table 24: repetition test L1 group scores, -/a/ ending, SVO targets

L1	T1 mean	T1 sd	T1 responses	T2 mean	T2 sd	T2 responses
FR	0.98	0.15	127	0.98	0.15	132
GE	0.98	0.14	145	0.96	0.19	155
IT	0.96	0.19	133	0.93	0.26	134
NL	0.92	0.28	154	0.90	0.30	158
EN	0.96	0.20	101	0.89	0.31	119

⁴⁸ Note that the two things do not imply each other: increasing accuracy of -/e/ is not the other side of decreasing accuracy of -/a/. Scores were computed on different subsets of data, identified by the ending required by the target.

Graph 6



Graph 7

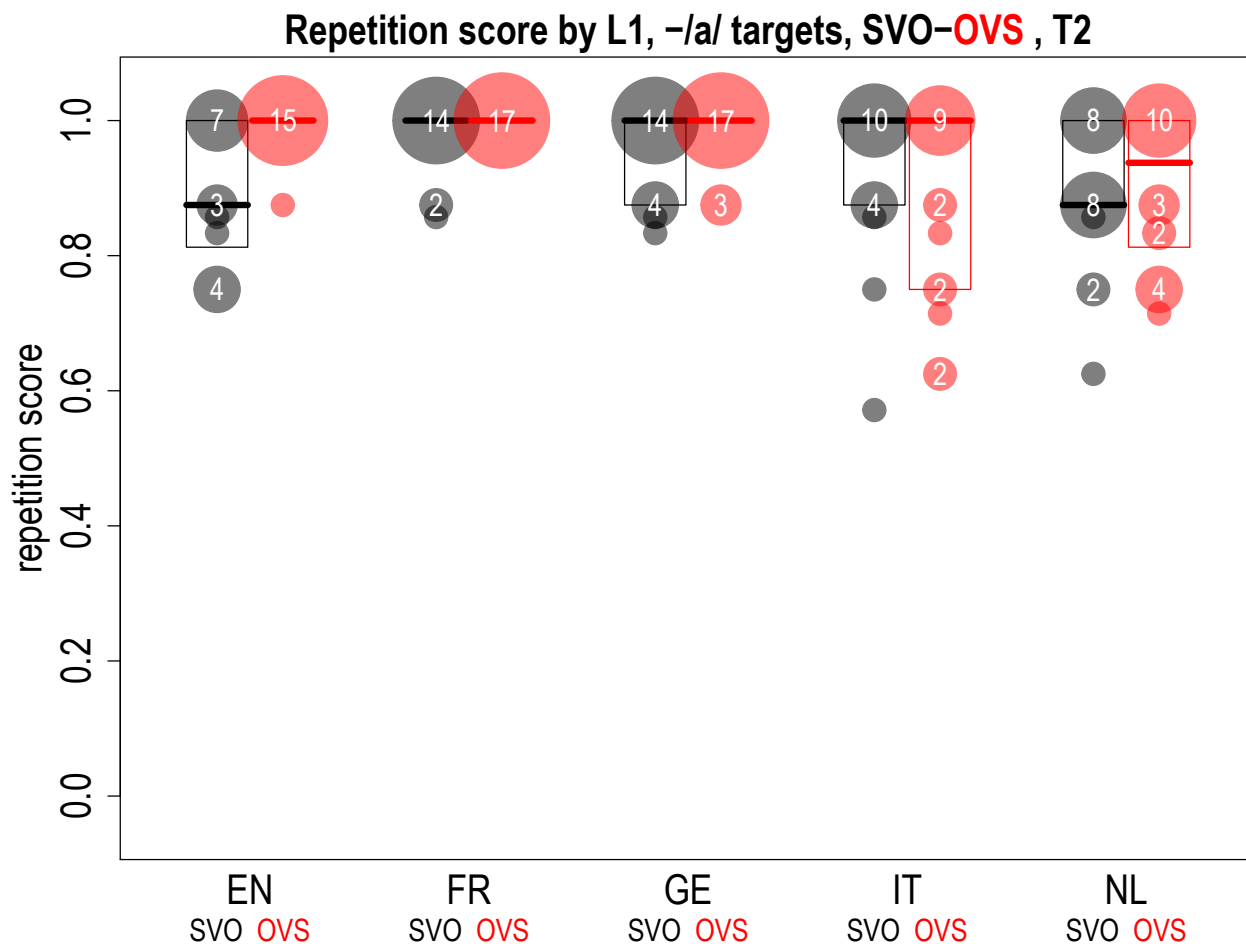


Table 25 lists the learners whose probability of correctly repeating -/a/ do not differ from chance⁴⁹. The learners it comprises do not appear to correctly repeat target -/a/ with a frequency which we could call systematic; the results are compatible with a strategy of guessing. The table describes each observation in terms of participants, L1, word order and time: for each relevant combination, then, it provides the number of correct responses, the number of trials, the resulting mean accuracy, and the probability that this output may be observed under chance conditions, that is, if the learner was guessing. These values are provided for both endings, with the following rationale: if the learner truly was guessing, then we should observe a chance result for both -

⁴⁹ Again, this probability is computed for each test time based on a binomial distribution described by the number of correct response, the number of trials, and a 0.5 chance threshold.

/a/ and -/e/ targets, as these are the only two alternative answers between which one can chose. Conversely, p values below 0.05 indicate that the learner is so consistent in repeating targets correctly that it is unlikely that he could have done so by mere guessing. Finally, p values above 0.95 would have a somewhat peculiar interpretation: the learner did establish the rule, only this is the exact reverse of what happens in the target language. In other words, the learner managed to associate case endings and syntactic function, but infortunately associated -/a/ with the accusative case and -/e/ with the nominative case. Naturally, it should not be too easy to draw such a conclusion on the basis of the input.

Table 25: Repetition test, repetition of -/a/ at chance level

Subject	L1	WO	Time	SI_correct		SI_trials		SI_mean		SI_p	
				-/a/	-/e/	-/a/	-/e/	-/a/	-/e/	-/a/	-/e/
2101	NL	SO	2	5	1	8	7	0.63	0.14	0.14	0.94
2102	NL	OS	2	5	0	7	8	0.71	0.00	0.06	1.00
2104	NL	OS	1	1	2	6	8	0.17	0.25	0.89	0.86
2108	NL	SO	1	5	8	8	8	0.63	1.00	0.14	0.00
2118	NL	OS	1	5	3	7	6	0.71	0.50	0.06	0.34
3119	EN	SO	1	4	0	6	7	0.67	0.00	0.11	0.99
5105	IT	SO	1	5	7	7	8	0.71	0.88	0.06	0.00
5105	IT	SO	2	4	7	7	7	0.57	1.00	0.23	0.00
5106	IT	OS	2	5	4	7	8	0.71	0.50	0.06	0.36
5114	IT	OS	2	5	3	8	6	0.63	0.50	0.14	0.34
5115	IT	OS	2	5	6	8	8	0.63	0.75	0.14	0.04

A few comments can be made. First, learners belong to only three L1 groups, the vast majority being speakers of Dutch or Italian. One (5105) appears in the table twice. Word order and test time, in contrast, are fairly varied.

Some learners (2118, 5105, 5115, highlighted in red) fail to reach significance on -/a/ repetition, but curiously do on -/e/ repetition. This result, however

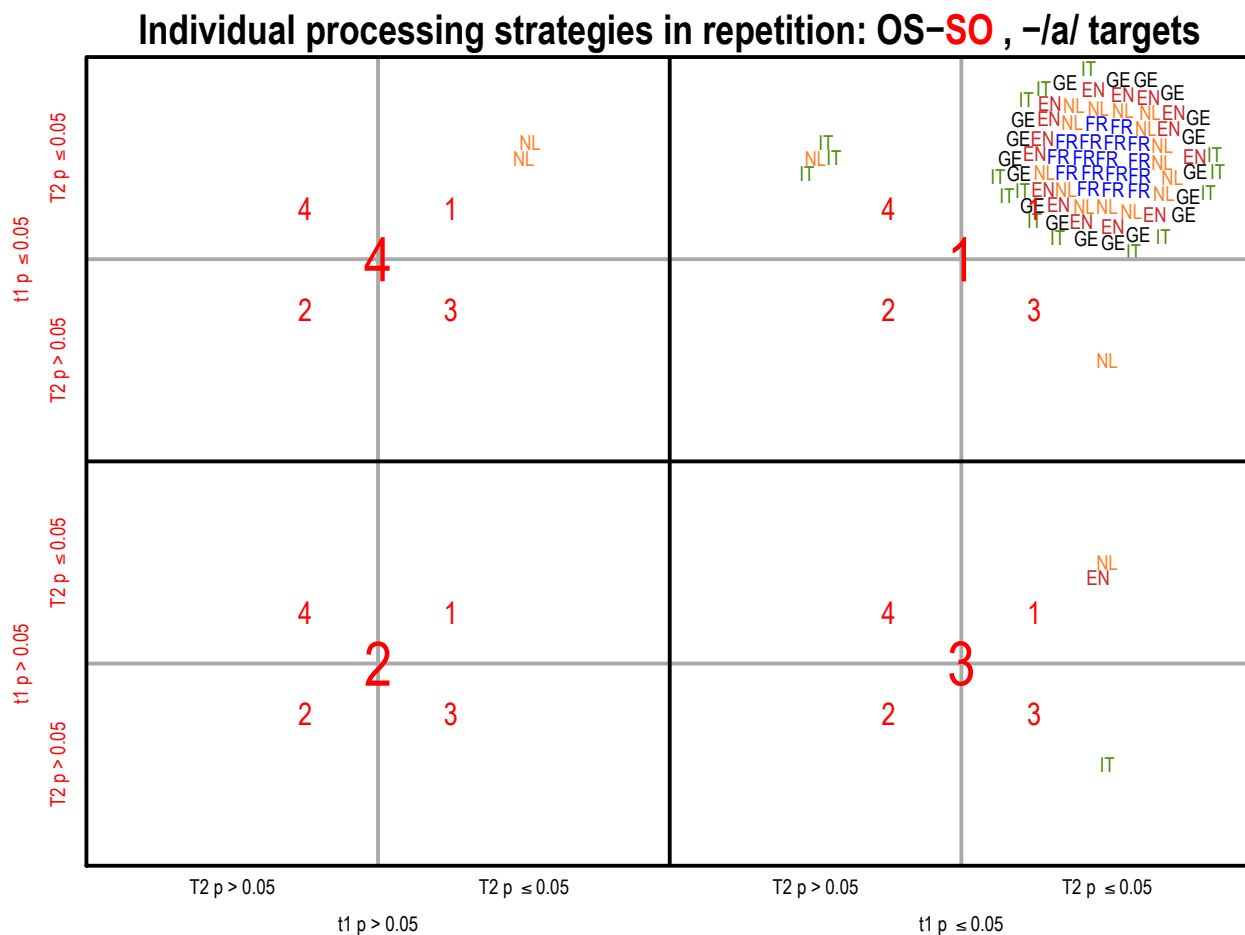
unexpected, probably witnesses to one of the main motors of change in the interlanguage, namely, fear for errors. The learners have noticed, either from the input or from the test items themselves, that Polish words sometimes end in *-/e/*. However, they failed to grasp the regularity that governs this eccentric behaviour; at the same time, they know that they tend to supply *-/a/* in all contexts, and can guess that sometimes this must be incorrect. For fear of this error, then, they make the opposite one, that is, they provide the marked form *-/e/* slightly more often than it is required.

All other learners perform at chance level⁵⁰ on both *-/a/* and *-/e/* targets, so that we can safely conclude that they performed the test by guessing.

We can project these findings against the bigger picture of *-/a/* processing, along the lines of section 2.5.4 (p. 79). Graph 8 plots learners according to their performance on the repetition of *-/a/* in OS and SO targets at T1 and T2.

⁵⁰ Some of the participants (2102, 2118, 5105, 5106) only marginally fail to reach significance on *-/a/* repetition. Typically, their p value is 0.06, their mean 0.71, and the correct/total ratio is 5/7. Interestingly, all of them appear to have missed a trial, which means that they supplied an ending other than *-/a/* or *-/e/*: this may witness to a certain creativity on their side, which is an indication of system restructuring. The 0.05 chance threshold is set arbitrarily with the purpose of indicating a reasonably small probability, and one could argue that 0.06, although undoubtedly greater, is not so far greater. That is the problem with all thresholds: setting a new one will simply shift the problem. For reasons of consistency, we will only signal this subtlety, and keep considering the learners in question as not performing chance.

Graph 8⁵¹



There is not much to say. The vast majority of learners lie in scenario 1:1, corresponding to scores above chance on OS and SO targets alike at both T1 and T2. The few data-points in other scenarios correspond to the 10 learners we just discussed above.

This concludes the presentation of the results yielded by the Repetition test.

2.7. REPETITION TEST: DISCUSSION

Once the distribution of the data is reasonably clear, we can finally turn to interpreting it. We start with a general remark on the nature of the target

⁵¹ This graph was created using the statistical software R (R Core team 2014) and the packages *wordcloud* (Fellows 2014) and *extrafont* (Chang 2014).

structure we chose to investigate: case marking poses radically different challenges from other target structures which only affect grammatical correctness, like, say, verb placement as in Håkansson (1989) or Schimke (2011). Said simply, the position of a verb in the sentence does not change its meaning, whereas case marking has the power to completely reverse the meaning of the utterance, all things being equal. Truly, it could be objected that all things are rarely equal: alongside the obvious use of prosody, semantics provides key information as to the thematic role that referents are most likely to perform; context, on the other hand, can make a particular interpretation of an utterance much more likely than all competing ones. The investigation of case marking, which is only possible through word order manipulation, inevitably requires a certain degree of artificiality.

This said, the fact is that the repetition test only provides us with the final output produced by the learner, but tells us nothing as to the processes which led to it. Crucially, in the absence of a comprehension or translation test, we have no way to know what learners meant to say. To this we should add that the output of the repetition test is the product of at least two complex processes, namely comprehension and production: it follows that errors may lie at either level, and that both should be tested, with all that follows. In other words, when we observe the learners processing of our main diagnostic tool, namely OS targets, we do not know if they interpreted them as such, and if they tried to express the same pragmatically marked meaning. We will try and solve this problem in the next two chapters.

For the time being, though, we should note that this uncertainty is also detrimental to the rationale of our methodology. We always said that word order manipulation is the tool which allows us to probe a learner's morphosyntactic competence, the hypothesis being that SO targets, being unmarked, will be easier to process than their marked OS counterpart. As shown in the Introduction, this claim is indeed founded on existing research showing that learners first case-mark objects in SO sentences, and only later learn to do the same in OS ones.

While this idea may be intuitive if one considers spontaneous speech, in which the learner is intentionally saying something meaningful, applying it to the repetition test implies several assumptions. For learners to find OS targets harder, it is necessary that they can recognise it, when they come across it: which is not obvious. Recall that when the object is processed incorrectly, it receives a basic ending in *-/a/*, just like the noun performing the subject function. As both nouns are now marked by an identical ending, what matters to express meaning is their relative position: the first one is the subject, the second is the object. To express the meaning of the OS target using a positional principle, the learner would need to swap the two nouns: yet we showed that this only happens in one occasion in the whole corpus. There are only two conclusions we can draw from these facts. The first is that learners understand the OS structure of the target, but since they cannot express it using inflectional morphology, they simply renounce to express it at all and produce an utterance which looks like an SO structure. While this may be the case, given that tests are always a stressful situation, we find it hard that not a single learner commented on this difficulty.

The other explanation, which sounds more natural, is that learners producing non-target-like case-marking could *not* identify the OS structure of the target, and either interpreted it as an SO structure, or simply identified the lexical items involved, at most. This explanation sounds sensible, but then a question comes up: if word order, that is, syntax is not involved in processing, then why are OS repetition scores consistently lower than SO ones? If test-takers do not associate case endings to the corresponding syntactic functions, it does not even make sense to speak of word order. Since *-/a/* and *-/e/* do not correspond to subject and object, but are merely two segments, what difference should it make to learners which one comes first in the target?

The answer may come from a field of studies which have little in common with syntax, namely those on perceptual prominence. Let us consider the nature of the repetition test. It is quite evident that in order to repeat a target sentence, whether as a mere string of sounds or as the result of de-coding and re-

encoding, that sentence has to be perceived first. As Gaonac'h (1991) concludes after a review of the available psycho-linguistic literature, perception is a level of crucial importance for all subsequent stages, namely noticing, storing, and ultimately productive use, although this level of input processing is more often postulated than explicitly addressed. Questions of perception are closely related to questions of saliency, however understood: as a preliminary operationalization, we may adhere to Peters' (1985:1030) interpretation that only salient stretches of sound constitute reasonable candidates for extraction from the input string, extraction in turn being defined as the recognising and remembering of language elements. This view is projected against the wider picture of child language acquisition by Slobin (1985:1164):

“on the most basic level, accessibility of linguistic material can be defined in terms of ‘perceptibility’. That is to say, the only linguistic material that can figure in language making are stretches of speech that attract the child’s ‘attention’ to a sufficient degree to be noticed and held in memory”.

This is achieved through the ‘extract’ operational principle which Peters (1985:1065) formulates as “extract whatever salient chunk of speech you can”. Data on earlier repetition test studies show that indeed, perception may be a relevant factor in at least explaining the results of this task. Gallimore and Tharp (1981) state that the accessibility of linguistic elements depends on their position in the utterance according to the hierarchy initial > final > medial. Peters (1985) claims that utterance-initial and utterance-final positions are maximally prominent and accessible for segmentation and storage, whereas utterance-internal positions are harder to access. Slobin (1985:1166) formulates for L1 acquisition the operating principles ‘attention: end of unit’ and ‘attention: beginning of unit’. More recently, VanPatten (2000:300) proposed his operating principles P4 (learners first process elements in sentence/utterance initial position) and P4a (learners process elements in final

position before elements in medial position). Finally, and most relevantly for the present work, Rast (2008:151) found that the accuracy of word repetitions in initial Polish L2 is affected by word position (utterance initial and final vs. medial) independently of the time of exposure (0, 4 and 8 hrs).

The studies cited so far typically considered the perceptual prominence of entire words or structures. Instead, we will now try to apply this reasoning to inflectional morphemes. Before we start, let us remind the reader that we consider the nominative ending *-/a/* as the unmarked option, for several reasons: it is more frequent, it appears in a wider variety of constructions, it represents the citation form of lexical items, and more. The accusative *-/e/*, in contrast, appears to be the marked option because it only marks the object in transitive structures, and as a result is much less frequent and was not introduced until later in the course. Now, in terms of perceptual prominence, it is evident that in SVO sentences the disfavoured accusative ending *-/e/* occurs in utterance-final position, thus receiving maximal prominence (23a). In OVS sentences, in contrast, this element always occurs in utterance-medial position, which might make it harder to perceive and consequently reproduce (23b)

- (23) a. *nauczycielk-/a/* *pcha* *studentk-/e/*.
 teacher-NOM pushes student-ACC
- b. *studentk-/e/* *pcha* *nauczycielk-/a/*.
 student-ACC pushes teacher-NOM

Thus, error distribution may be accounted for more accurately by hypothesising that learners are more successful at reproducing target structures if these are more retrievable from a perceptual point of view. In SVO sentences, the disfavoured infrequent ending is in the maximally prominent utterance-final position and stands the best chances of being noticed and processed. Higher error rate in OVS sentences, in contrast, is a consequence of the reduced perceptual prominence of the non-default case ending in utterance-internal position. In this condition, learners can only rely on very weak phonetic clues

to retrieve and reproduce the correct target ending. Indeed, in such contexts the data show a significant tendency to provide the default word-form in *-/a/*. For the time being we do not have sufficient elements to solve the question definitively. While the picture will hopefully become clearer in the next two chapters, we need to be aware that more than one utterly different explanation may be available for the same phenomenon.

There are a few other observations we would like to make. The range of case endings produced by the learners, for instance, seems to be unexpectedly restricted. Recall from the introduction that developmental studies suggest that learners first go through a NOM/non-NOM opposition, as shown once again in examples from Slavic languages, namely Russian L2 (24a: Artoni & Magnani 2015:188) and Serbian as a heritage language (24b: Di Biase, Bettoni & Medojević 2015:209). Only later does this generic opposition stabilise as NOM/ACC.

- (24) a. *videla* *volk-e*
 saw-FEM wolf-NON.NOM
 "(she) saw a wolf"
- b. *onak su* *videli krevet-a*
 then have seen bed-NON.NOM
 "then (they) saw (a/the) bed"

This does not necessarily mean that the VILLA learners acquired case marking better and more quickly than in untutored SLA. The analysis of spontaneous production carried out by Bernini (2016) and Dimroth (submitted) shows that utterance structure simultaneously reflects a variety of principles which in spontaneous SLA are typical of different developmental stages: the pre-basic, pragmatic structure "focus last"; the basic, semantic "controller first"; and the post-basic, SVO syntactic organisation. Interpreting such mixture of apparently anachronistic principles as a consequence of the particular VILLA received, we proposed the label "Instructed Basic Variety". Bernini in particular correlates

the structural properties of the interlanguage with its phonology, arguing that while random phonological variability, or rather tolerance towards allophonic variation is typical of pre-basic varieties, "la fixation d'une forme de base du [...] mot dans la variété basique réduit la gamme de variation (allo-)phonique [...], en fondant la possibilité d'oppositions phonémiques". In this respect, he also observes that while several lexical items are relatively stable in their phonological form, others show considerable variability, both in their supposed target and in their phonetic structure, e.g. {çpi, ʃpi, spi} for target /çpi/, "sleeps". Even when the various tokens produced by a learner seem to be mappable onto specific target forms, and thus to reflect the input to a certain extent, their use nonetheless witnesses to a lack of functional differentiation, coherently with information available on spontaneous SLA (1993). Moreover, Bernini suggests four factors which may have an influence in determining the phonological variability of lexical items in initial SLA, namely a) frequency, b) the number and structure of syllables, c) the number of different word-forms present in the input, and d) semantics. While b) and d) are intrinsic to the lexical items, a) and c) depend on the input. This wealth of data only makes it harder to make sense of our EI output, as even the correct repetition of case endings may indicate a post-basic syntactic utterance organisation just as well as pre-basic, random phonological variability. What we can say for sure is that learners picked up these alternative endings from the input, thus showing some sensitivity from it. Moreover, in the vast majority of case the endings in question belong to the appropriate paradigm. In fact, if we exclude the instances of centralised vowels mainly produced by the German learners, we only had to exclude only a handful of endings which did not match either -/a/ or -/e/.

Finally, one could hypothesise that EI probes different types of competence depending on the test taker's proficiency level: if indeed targets are filtered through the learner's grammatical system, we should expect more proficient learners do to better at this test because their grammatical system helps them to overcome mnemonic constraints, for instance through what is commonly

known as "chunking" (1956), that is, the ability to group more than words into a constituent and treat that as a unit. If a learner's linguistic system is not sufficiently developed, in contrast, the target will sound more similar to a chain of nonce syllables. Okura and Lonsdale (2012), for instance, show that EI scores significantly correlate with participants' scores on a general English placement test, but not with WM scores, and that the lowest-scoring students were unable to repeat anything beyond their WM capacity.

We conclude by pointing out a few methodological faults which emerged in the course of this work, in order to build more accurate and efficient tools in the future. To start, the distractor employed in the test may not be the most appropriate. Drawing a simple geometrical figure does not necessarily inhibit WM, as required by the EI rationale. It is possible that the difference between the groups considered in this work, which we tried to attribute to WM, were in fact the results of a different personal approach to the test. Learners in scenario "c", who failed, might have taken their time to draw the picture, allowing their WM to fade. In contrast, those in scenario "e", who passed the EI test, but not the corresponding comprehension task, might have tried to complete the distracting phase as quickly as possible, mentally rehearsing the target. A more appropriate distractor should replace the content of their WM with new material, for instance by asking to perform simple calculations, read a sentence, answer a question, count from one to ten and backwards, and so on. Only then could one be sure that repetition really involves the re-coding of previously comprehended meaning.

This leads us to a further point, namely assessing target comprehension. In a standard situation, the researcher can observe learner output and correlate it to the meaning of the target sentence. There is no guarantee, however, that learner comprehension is target-like: the output may be deviant following a failure not at the level of production, but at the level of comprehension. A comprehension test, or, even better, a translation test applied to the test items

of the EI test would partially solve the problem. Alternatively, comprehension can be assessed through an appropriate distractor, as indeed is sometimes done. Such methodology also helps to make sure that learners aim at meaning and avoid focussing on form, in full accordance with the rationale of the test.

Timing the test could usefully bring the test closer to the context of time-constrained spontaneous speech, while at the same time minimising the chances that learners might focus on form. In fact, timed tasks are generally deemed more appropriate for accessing implicit competence (Ellis 2005).

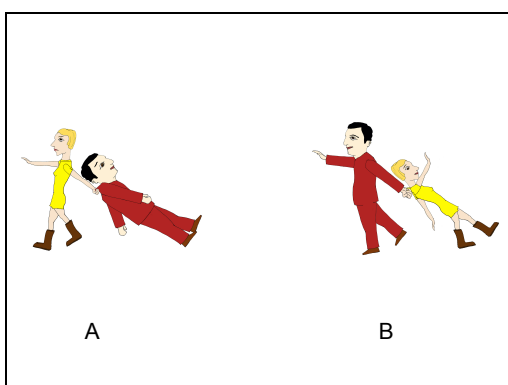
In spite of these methodological faults, which constitute precious experience for future improvement, the VILLA EI test showed how lively the debate is about the EI test, resulting no doubt from its usefulness, on the one hand, and from the many still not fully answered questions as to its mysterious mechanisms, on the other hand. We hope that this study brought a little contribution to this fertile debate, while at the same time showing the linguistic problem in all its intriguing complexity.

3. THE COMPREHENSION TEST

3.1. THE COMPREHENSION TEST: STRUCTURE

In the Comprehension test (*Picture Verification* in the VILLA project), learners hear short Polish transitive sentences and subsequently see two pictures on screen. These depict the same two referents in the same situation, but they differ in the thematic roles of the two referents: in other words, the same referent has the role of agent in one picture and of patient in the other one (Figure 2).

Figure 2: Comprehension test: alternative descriptions of the target utterance



The learners' task is to select the picture which in their opinion best depicts the target sentence they heard. Responses were marked in pen on an answer sheet, as shown in Figure 3. The data thus obtained were digitalised manually in spreadsheet format and then further manipulated and analysed with R (R Core team 2014).

Figure 3: Comprehension test, example of answer sheet

CS A

STUDENTE: 5101

RISCALDAMENTO

N°	A	B
1		x
2	x	x
3	x	

TEST

N°	A	B
1	x	
2		x
3	x	
4	x	
5		x
6		x
7	x	
8	x	
9	x	
10	x	
11	x	
12	x	
13		x
14		x
15		x
16		x
17	x	
18	x	
19		x
20		x
21		x
22	x	
23		x
24		x

VILLA Italia – Settembre 2012

3.1.1. Comprehension test: target items

The test comprises 24 target sentences, in addition to three warm-up sentences to make sure that learners had correctly understood its structure. The test was administered collectively in a classroom: target sentences were played aloud through loudspeakers and pictures were projected on screen with a beamer. Learners took the test twice, namely after 9 hours (T1) and 13:30 hours (T2) of exposure to the input. Note that this is consistent with the timing of the Sentence Repetition test described in the preceding chapter: the two tasks

probe different aspects of the learner's developing competence in the L2 after identical exposure to the input.

Target sentences have the structure NP - Verb - NP (Table 26). Only two nouns were used for this test, namely *brat*, "brother", and *siostra*, "sister". The verbs were the same employed in the SI test, namely *ciagnie*, "pulls", *pcha*, "pushes", *pozdrawia*, "greet", and *wola*, "calls". Each noun appeared in both its NOM and ACC form; further, target sentence varied in constituent order (SVO, OVS, OSV), each occurring in eight target sentences.

Table 26: Comprehension test, example of target sentences with the verb *wola*

SVO	<i>siostr-a wola brat-a</i> sister-NOM calls brother-ACC	<i>brat wola siostr-ę</i> brother.NOM calls sister.ACC
OSV	<i>siostr-a brat-a wola</i> sister-ACC brother.NOM calls	<i>siostr-ę brat wola</i> sister-ACC calls brother.NOM
OVS	<i>brat-a wola siostr-a</i> brother.ACC calls sister-NOM	<i>siostr-ę wola brat</i> sister-NOM brother-ACC calls

Table 27 presents the paradigm of the two target nouns employed in the test: shaded cells denote the cases not immediately relevant for the present study. *Brat* follows the declension of masculine animate nouns, *siostra* that of feminine nouns in *-a*.

Table 27: Comprehension test, paradigm of the target nouns

	<i>brat</i> "brother"	<i>siostra</i> "sister"
NOM	<i>brat</i>	<i>siostr-a</i>
GEN	<i>brat-a</i>	<i>siostr-y</i>
DAT	<i>brat-u</i>	<i>siostrz-e</i>
ACC	<i>brat-a</i>	<i>siostr-ę</i>
INSTR	<i>brat-em</i>	<i>siostr-q</i>
LOC	<i>braci-e</i>	<i>siostrz-e</i>
VOC	<i>braci-e</i>	<i>siostrz-e</i>

As can be seen, the ACC case of masculine nouns like *brat* is characterised by the ending *-a*, which also occurs in the NOM case of feminine nouns like *siostra*. This observation will be of some relevance in our subsequent analysis of the data.

3.2. COMPREHENSION TEST: RESULTS

3.2.1. Descriptive statistics

Table 28, Table 29 and Table 30 below provide descriptive statistics relative to learner scores on SVO, OSV and OVS targets at T1 and T2.

Table 28: descriptive statistics, SVO targets

L1	n.	mean - T1	sd - T1	responses - T1	mean - T2	sd - T2	responses - T2
EN	17	0.98	0.15	136	0.99	0.08	136
FR	17	0.93	0.25	136	0.86	0.34	136
GE	20	0.87	0.34	160	0.87	0.33	160
IT	17	0.97	0.17	136	1.00	0.00	136
NL	20	0.99	0.08	140	0.86	0.34	140

Table 29: descriptive statistics, OSV targets

L1	n.	mean - T1	sd - T1	responses - T1	mean - T2	sd - T2	responses - T2
EN	17	0.35	0.48	136	0.52	0.50	136
FR	17	0.43	0.50	136	0.72	0.44	136
GE	20	0.71	0.45	160	0.80	0.39	160
IT	17	0.51	0.50	136	0.71	0.45	136
NL	20	0.36	0.48	140	0.55	0.49	140

Table 30: descriptive statistics, OVS targets

L1	n.	mean - T1	sd - T1	responses - T1	mean - T2	sd - T2	responses - T2
EN	17	0.09	0.28	136	0.08	0.28	136
FR	17	0.29	0.45	136	0.55	0.49	136
GE	20	0.64	0.48	160	0.75	0.43	160
IT	17	0.45	0.50	136	0.66	0.47	136
NL	20	0.30	0.46	140	0.43	0.49	140

A few observations can be made based on these descriptive statistics alone. First, as expected, SO scores are much higher than their OS equivalents in all cases. Curiously, though, we find surprising ratios of standard deviation, pointing to the fact that some learners actually made several errors on SVO targets too, which runs contrary to our hypotheses. These cases will be discussed in detail in section 3.4.2, p. 123.

Regularities are also observed in the difference between the two OS constituent orders. Specifically, accuracy on OSV targets is higher in all cases, the only exception being the German group at T2. The English group stands out too in this respect, in that the difference between the two word order is particularly extreme, and the standard deviation on OVS targets is much lower than in the other L1 groups. Combined, these two pieces of information indicate that compared to OSV targets, English learners perform much worse on OVS ones than the other L1 groups do, and that all learners of this group do so in a rather uniform manner, which seems to lay the condition to suspect an L1

effect. However, a difference in means does not necessarily mean a difference in processing strategies, if it is not statistically significant. The topic is treated in more detail in section 3.4.1, p. 115.

3.3. GUESSING IN THE COMPREHENSION TEST

Based on the discussion of the Repetition test (section 2.5, p. 70), we do not deem useful to produce inferential statistics based on learners' mean scores. Instead, we will calculate the likelihood that learners might have performed the Comprehension test significantly *above* chance, that is, having identified the rule which governs the associations of case endings to syntactic functions.

The comprehension test probes the learner's use of case endings by manipulating word order. On SO targets, meaning can be derived based on both a positional and a morphosyntactic principle. In the former, syntactic functions are assigned to nouns on the basis of their relative position in the utterance, so that the first noun will be treated as agent, the second one as patient. In the latter, syntactic functions are derived from case endings, so that *-/a/* corresponds to agent, *-/e/* to patient.

On OS targets, only the morphosyntactic principle will work, as the agent of the utterance no longer occurs in first position. By studying learners' scores, thus, we can make deductions as to the probability that they applied a morphosyntactic strategy when performing the test. Specifically, we can make reasonable claims as to their use of such principle, if the probability of obtaining their score by mere guessing are so low that we can safely enough discard this possibility, the typical threshold being 5%. On 8 trials, that means six correct responses or more, an unlikely result under chance conditions or a positional principle.

Before that, however, we would like to spend some words on those learners who do *not* perform significantly above chance. This encompasses two scenarios: learners performing *at chance level*, and learners performing significantly *below chance*. We will discuss the latter first.

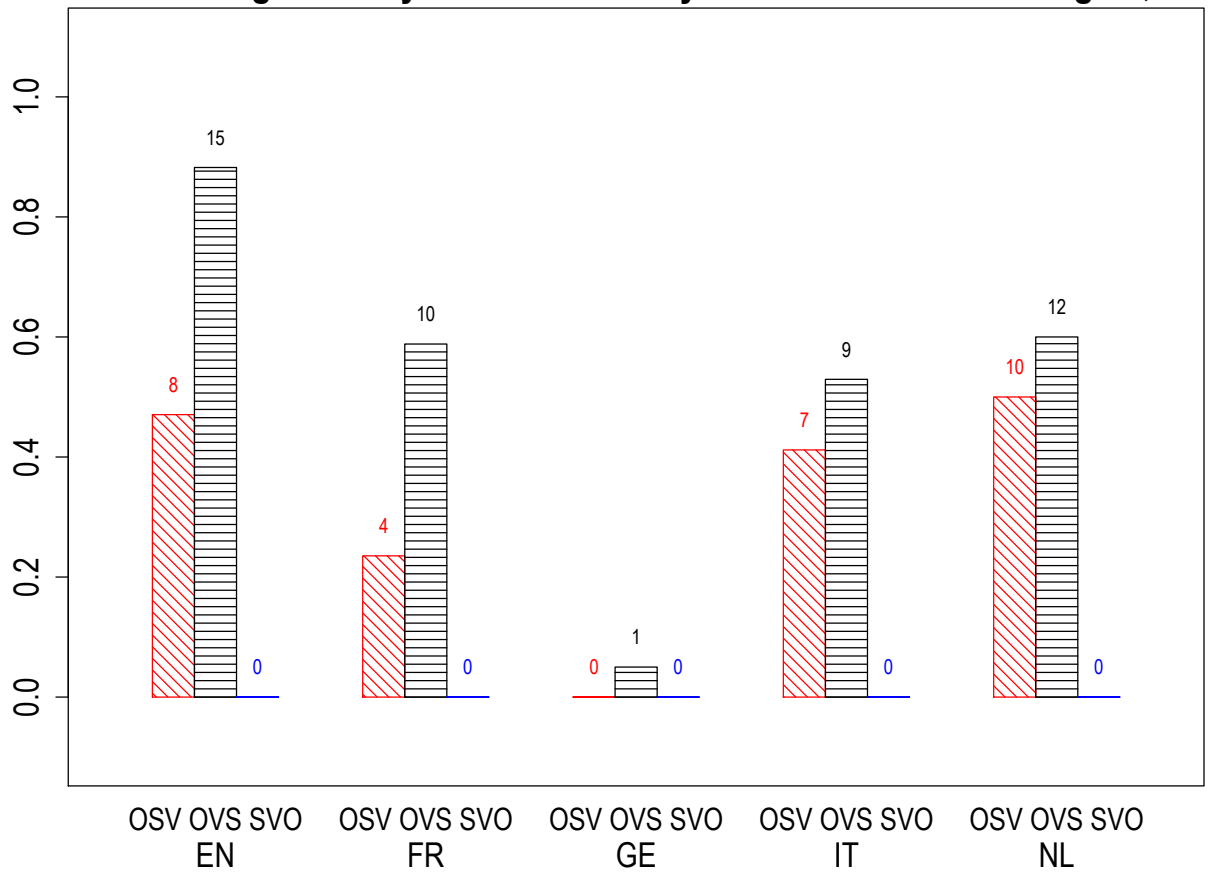
3.3.1. *Comprehension test: learners below chance level accuracy*

This scenario is really the reverse of learners performing significantly above chance: their scores are so extreme that they must be due to some consistent strategy. The difference lies in the fact that the scores of these learners are extremely *low*, rather than high⁵². In other words, the learner must have been guided by a systematic principle, which alas proved to be incorrect. In the case of OS targets, that may have been the positional principle, indicating that the learner systematically interpreted these utterances as agent-first. In the case of SO targets, in contrast, the unlikely interpretation would be either that the learner considered all utterances to be patient-first, or otherwise applied an inverse morphosyntactic principle in which NOM endings correspond to the object function, and ACC endings correspond to the subject. Graph 9 and Graph 10 present the relevant data at T1 and T2 respectively. The height of the bars indicate the proportion of learners on the L1 group, whereas the digits above them specify the actual number.

⁵² Technically, that means that the probability of that result in the lower tail of the distribution is below 5%.

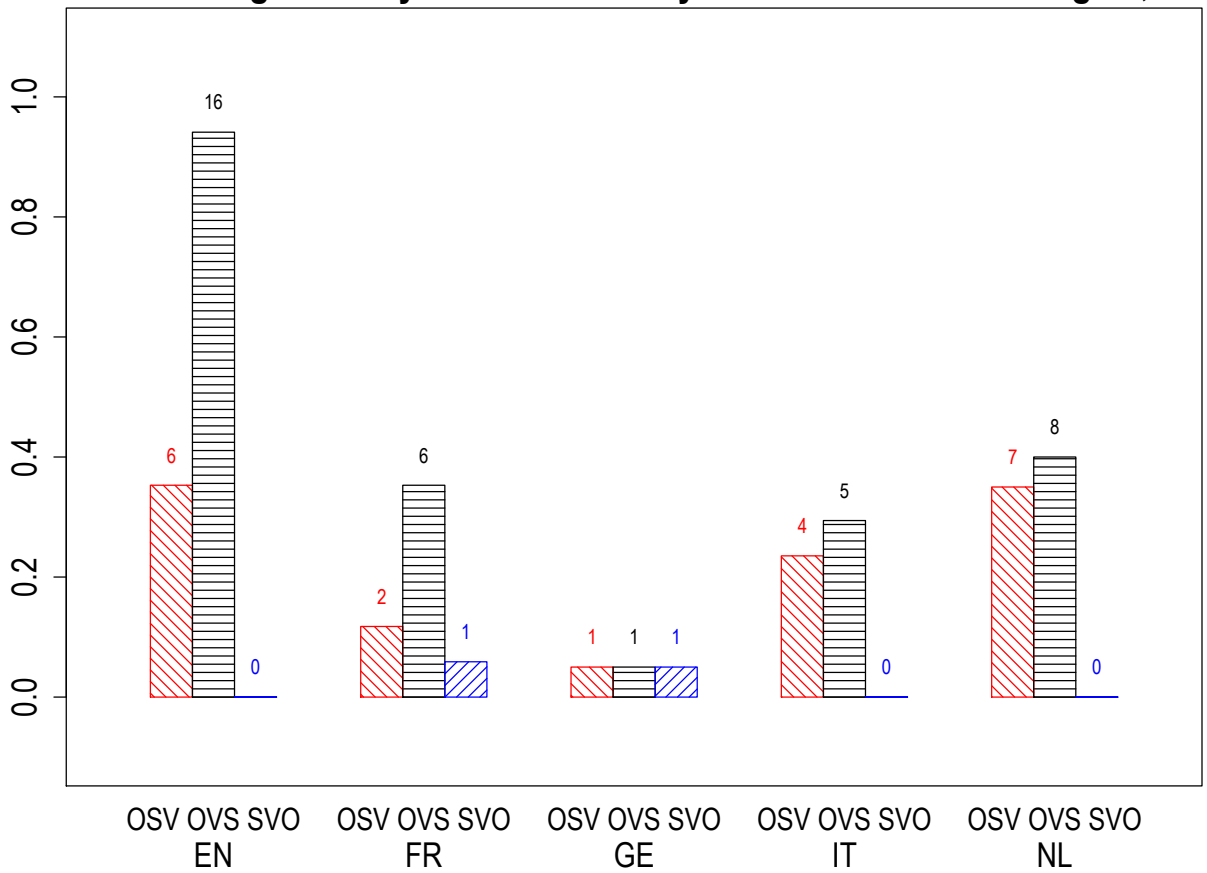
Graph 9

Learners significantly below chance by L1: **OSV**-**OVS**-**SVO** targets, T1



Graph 10

Learners significantly below chance by L1: **OSV**-**OVS**-**SVO** targets, T2



Not unexpectedly, no learner performs below chance on SO targets, at least at T1. At T2, in contrast, two learners, one French and one German, appear to do so. This result is most unexpected and highly unlikely, and witnesses to some confusion in the learner's mind, or, put otherwise, to a very stormy restructuring of the learner variety.

Secondly, all L1 groups with the partial exception of the Germans seem to be more consistent on OVS targets than on OSV ones, although to a different extent depending on the L1. This would suggest that OVS targets are more likely to be interpreted as agent-first than OSV targets are, resembling in this their SVO equivalents. This question will be detailed in section 3.4.1, p. 115.

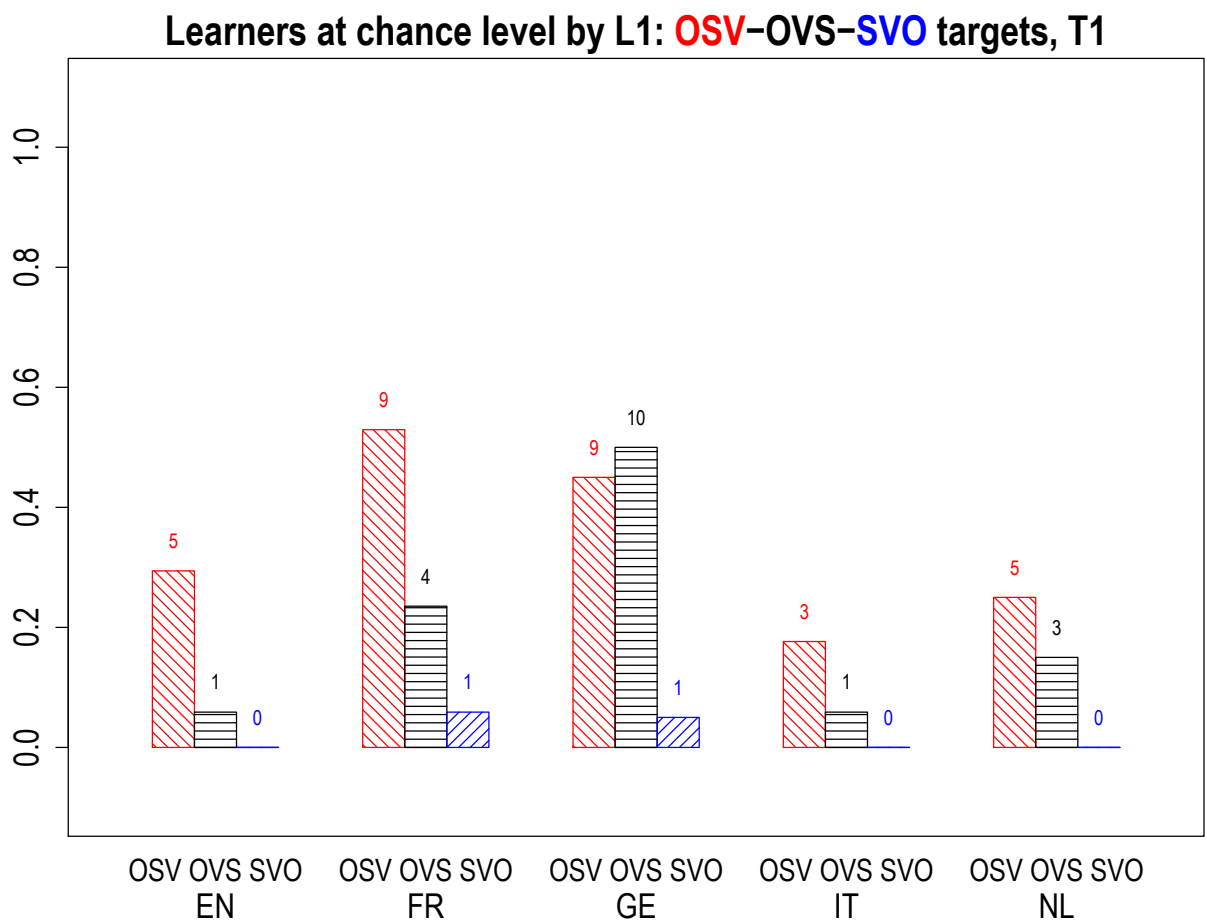
Regarding the effect of input exposure, finally, between T1 and T2 proportions on OS targets generally decrease from about 60% to 40%, indicating that at

least some learners abandon the positional principle and move on to chance level response, or possibly to a morphosyntactic principle. The English group is an exception in this respect, since the proportion of learners performing below chance on OVS targets remains close to 100% and even increases from T1 to T2.

3.3.2. Comprehension test: learners at chance level

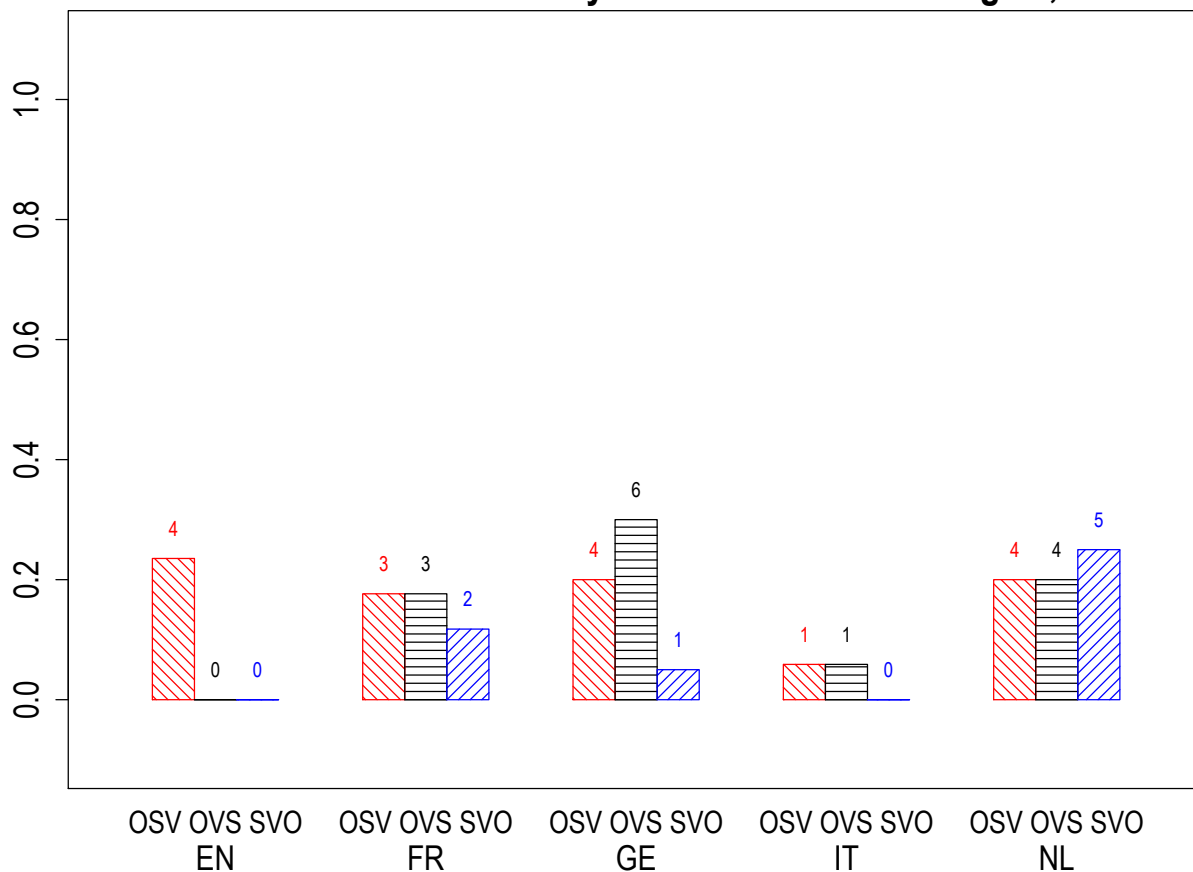
We now turn to learners truly performing at chance level, who can be thought of as performing the test by pure guessing. Graph 11 and Graph 12 present the relevant data at T1 and T2 respectively. The height of the bars indicate the proportion of learners on the L1 group, whereas the digits above them specify the actual number.

Graph 11



Graph 12

Learners at chance level by L1: OSV-OVS-SVO targets, T2



Again, chance level is rarely encountered on SO targets, especially at T1. The proportion increases at T2, though, which suggests that learners may be restructuring their interlanguage by formulating new hypotheses. Certainly they now have doubts about the principles of utterance organisation operating in the target language, and come to question the positional principle on which they probably relied initially. In parallel with the results presented above, OSV targets tend to elicit more chance responses than OVS, especially at T1, again suggesting that the two OS structures may be processed rather differently.

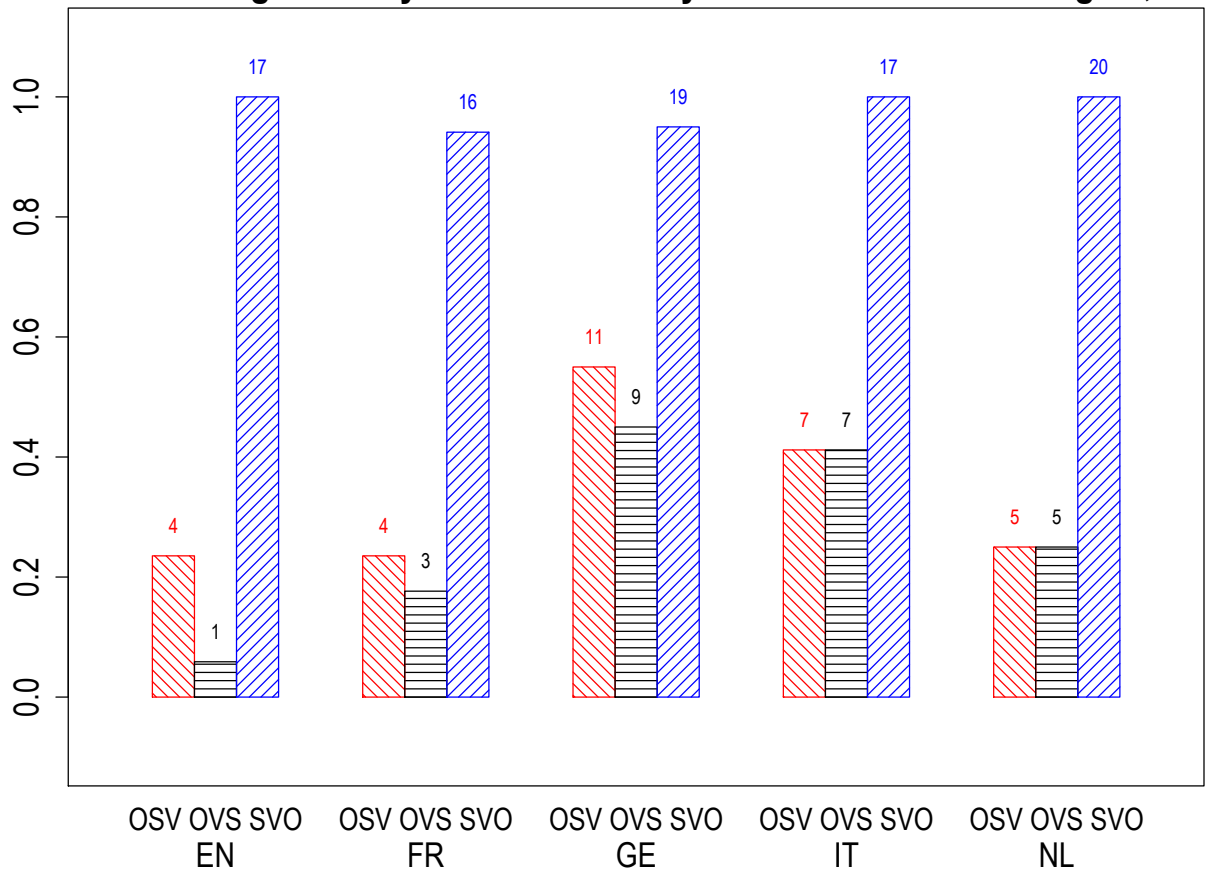
Regarding the effect of time, finally, proportions generally tend to decrease from T1 to T2, probably indicating a shift towards target-like processing, that is, a morphosyntactic principle.

3.3.3. Comprehension test: learners above chance

We can now turn to discussing the linguistic behaviour of those learners who appear to have performed the comprehension test on the basis of a morphosyntactic principle. The analysis focusses on targets which can only be correctly processed with a morphosyntactic principle, that is, OS target sentences. In the case of SO targets, arguably, both a positional and a morphosyntactic principle will allow the learner to correctly retrieve the meaning of the utterance, so success in this condition is not particularly informative. On OS targets, in contrast, applying a positional principle would lead to systematic errors, the only way to correctly deduce meaning being applying a morphosyntactic principle. Because of its format, the Comprehension test is particularly apt to guessing strategies, so this statistical approach appears to be appropriate to the situation. Graph 13 and Graph 14 present the relevant data. Bar height indicates the proportion of learners reaching significance a in group, whereas the digits above them specify their actual number.

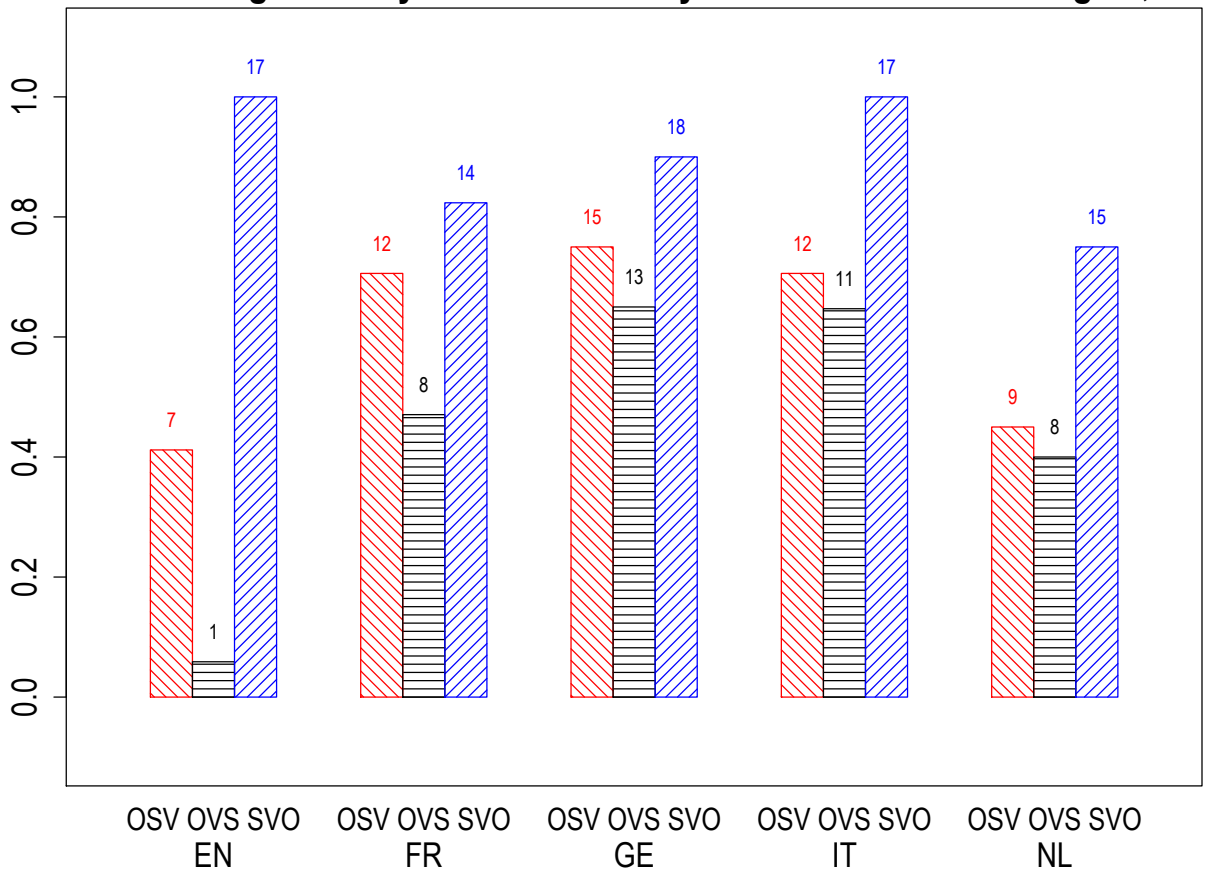
Graph 13

Learners significantly above chance by L1: **OSV**-**OVS**-**SVO** targets, T1



Graph 14

Learners significantly above chance by L1: **OSV**-**OVS**-**SVO** targets, T2



The first noteworthy observation concerns the obvious difference between the processing of SO targets, on the one hand, and of OS ones, on the other hand. Although this tendency is general, it appears that the gap between the two structure types may vary depending on the L1. This fact is hardly surprising, given that SO targets may be interpreted correctly based on both a morphosyntactic and a positional principle, which effectively means independently of morphology. For the same reason, though, it is curious that at least some learners may *not* perform above chance on this type of targets, as shown in the two preceding sections. The phenomenon only concerns some of the L1s, specifically: French and German at T1, both with a single learner, and again French, German and Dutch at T2, with 3, 2 and 5 learners respectively. The number of subjects exhibiting this behaviour, thus, appears to grow with

further exposure to the input. We will return to this topic in section 3.4.2, p. 123.

In parallel, and contrary to our expectations, there are quite a lot of learners who perform above chance on syntactically marked OS word orders. On OSV, specifically, we find 31 learners at T1 and 55 at T2, on a total of 91 participants. All L1s are represented, although at times in dramatically different proportions. The same is true for OVS targets, on which 25 learners performed above chance at T1, and 42 at T2.

In both data sets, the number increases from T1 to T2, which witnesses to a positive effect of further exposure to the input.

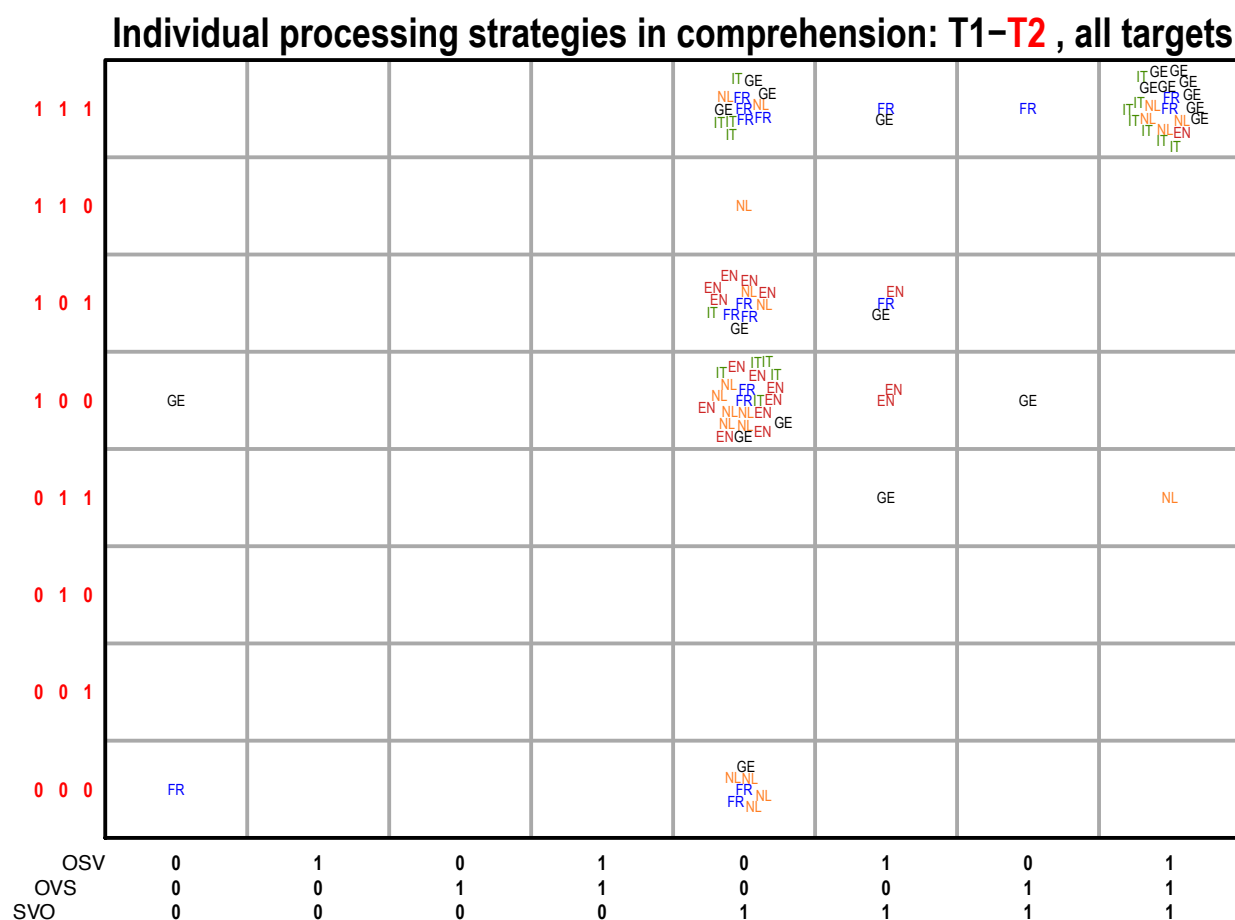
3.3.4. Comprehension test: a comprehensive picture

To conclude our description of the data, we present a general picture which allows us to take into account all the factors which are treated in our analysis: the learners' L1, word order, and time. The objective is to perform a rudimentary cluster analysis to verify whether learners can be grouped based on these factors. Graph 15 is the tool which we will use to this purpose.

Scores at T1 are presented on the horizontal axis, scores at T2, on the vertical axis. On both axes, scores, or, more properly, scenarios, are defined by the combination of learner performance on the three target word orders: SVO, OVS and OSV. A score of 1 indicates that the learner performs above chance on the corresponding target structure, whereas 0 indicates response at chance level.

Each learner is thus identified by a combination of scores at T1 and T2, that is, by his position in one of the 64 squares in which the graph area is divided.

Graph 15⁵³



As a first observation, we could point out that of the 64 theoretical possibilities, only a few are realised in practice, and fewer still - four, in fact - reunite the bulk of the subjects. As these correspond to varying degrees of success in the test, we could interpret this information as a hint to the existence of a hierarchy in the development of morphosyntactic competence in comprehension, identifiable both synchronically and diachronically. In the first case, we consider the column sum, for T1, or the row sum, for T2. In this way, we compute the number of subjects performing in a specific manner at either test time, without taking into account their performance at the other test time.

⁵³ This graph was created using the statistical software R (R Core team 2014) and the packages *wordcloud* (Fellows 2014) and *extrafont* (Chang 2014).

It is easy to see that at T1, most learners (56) show the following score: SVO 1, OVS 0, OSV 0, which clearly corresponds to a positional strategy. 23 learners, in contrast, already exhibit a well-developed morphosyntactic strategy (SVO 1, OVS 1, OSV 1). In between these two groups, 8 learners correctly process SVO and OSV targets, but not OVS, and only 2 do the opposite, which suggests that OSV structures should be more accessible compared to their OVS equivalents.

At T2, the number of learners applying a positional principle is greatly reduced to 26, whereas those using a morphosyntactic strategy now number 36. 15 subjects, finally, perform better on OSV than OVS targets. The picture at T2, therefore, confirms the situation at T1, with a tendency for results to become more target-like.

We can now study the evolution of learners' processing strategies over time, which is the true objective of Graph 15. To do so, we will describe the evolution of those learners who at T1 were found to apply a positional principle (SVO 1, OVS 0, OSV 0), which represent the vast majority of the subjects.

Of these, 23 learners do not change their processing strategy at all, consistently adhering to a positional principle at T2 as well. In contrast, 13 learners move all the way towards a morphosyntactic strategy, and at T2 can consistently derive meaning from both SO and OS targets. 12 learn to process OSV structures, but still not OVS; only 1 learner is found in the reverse situation. This further information is in accordance with the synchronic data, which showed us that, both at T1 and T2, OSV targets are correctly processed by a greater number of learners than their OVS equivalents. Diachronically, it appears that case-marking in the two OS word orders can develop either at same time, or separately. In the latter case, OSV develops first. A more detailed discussion of the apparently differential processing of OSV and OVS targets can be found in section 3.4.1, p. 115.

Finally, 7 subjects move to a stage in which no word order is processed above chance. There may be various reasons for this: the trivial explanation is that these learners by T2 got tired with the VILLA project and simply performed the

comprehension test at random, with no effort, that is, by pure guessing. If we assume them to have performed the test to the best of their abilities, in contrast, then we should conclude that they are not yet ready to consistently process OS targets, but have acknowledged their existence to the point that sometimes they interpret even SO targets as having a OS structure too. For a more detailed discussion of this possibility, see section 3.4.2, p. 123.

The last consistent group of learners reunites those who already reasoned morphosyntactically at T1. With only one exception, they not surprisingly keep doing so at T2 as well.

This concludes our presentation of the data elicited with the comprehension test, as well as our description of the main tendencies which can be observed.

3.3.5. Comprehension test: inferential statistics

Having provided this overall description of the dataset, we now turn to verifying what factors are most influent in determining these results. To this purpose, we fit a generalised linear model with binomial error structure and logit link function. The dependent variable of this model is the proportion of learners who performed above chance on the total number of learners in each group. The fixed effects again include the L1, word order and time, as well as their interaction. The results of an analysis of variance performed on this model are shown in Table 31.

Table 31: Comprehension test linear model: ANOVA

	Df	Deviance	Resid. Df	Resid. Dev	Pr(> Chi)
NULL	NA	NA	29	245.43	NA
L1	4	22.09	25	223.34	< 0.01
Word Order	2	168.31	23	55.03	< 0.01
Time	1	11.20	22	43.83	< 0.01
L1:Word Order	8	22.45	14	21.38	< 0.01
L1:Time	4	3.12	10	18.25	0.54
Word Order:Time	2	15.84	8	2.41	< 0.01

All linear predictors are significant, as well as the interaction between the L1 and Word Order, on the one hand, and Time and Word Order, on the other hand. No interaction was found between the L1 and time.

3.4. COMPREHENSION TEST: PROBLEMATIC SCENARIOS

The next two sections are devoted to the minoritarian but troubling behaviour exhibited by some learners, namely the differential processing of OSV and OVS targets, on the one hand, and errors in the processing of syntactically unmarked SVO structures, on the other hand.

3.4.1. Comprehension test: differential processing of OS word orders

We first turn to quantifying and trying to account for the differences in the processing of OS targets which were identified in the previous sections. Obvious gaps in scores were found between SO targets, on the one hand, and OS targets, on the other hand. These differences are not problematic in that they can easily be explained by the processing principle required to extract meaning from them: positional or morphosyntactic in the former case, necessarily morphosyntactic in the latter. When it comes to OS targets,

however, there should be no differences in processing accuracy, as both OSV and OVS targets share the same relative order of subject and object. Nevertheless, it does seem that OVS targets prove consistently harder to process than OSV ones. We identified evidence for such claim in various contexts: first, we observed marked differences in the mean scores (section 3.2, p. 100); then, we found that the processing of OVS targets is very often significantly *below chance* (section 3.3.1, p. 103); further, scenarios in which OVS structures are processed more accurately than OSV are rare synchronically, and diachronically, OSV almost always develops first (section 3.3.4, p. 111); in addition, the shift towards a morphosyntactic strategy between T1 and T2 appears to be more radical in the case of OSV than OVS targets. In this section we will first describe the phenomenon in detail and then perform a statistical test to verify whether the observed differences are statistically significant and require a specific explanation.

Graph 16 presents an overall picture of each learner's processing strategy of OS targets at both T1 and T2. The processing scores of OSV targets are represented on the horizontal axis, with learners behaving at chance level on the left (scenarios 2 and 4), and learners above chance on the right (scenarios 1 and 3). Conversely, the processing scores of OVS targets are represented on the vertical axis, with learners behaving at chance level at the bottom (scenarios 2 and 3) and learners above chance at the top (scenarios 1 and 4). Taken together, the two scores provide an overall picture of learners' behaviour on both OVS and OSV targets at the same time. This information was lacking in the descriptive statistics presented in section 3.2, p. 100, but we deem it essential to try and identify the principles according to which each learner performs the comprehension task. The main area is divided into four main squares, each representing a processing scenario at T1 according to the conventions summarised in Table 32.

Table 32: scenarios, rationale

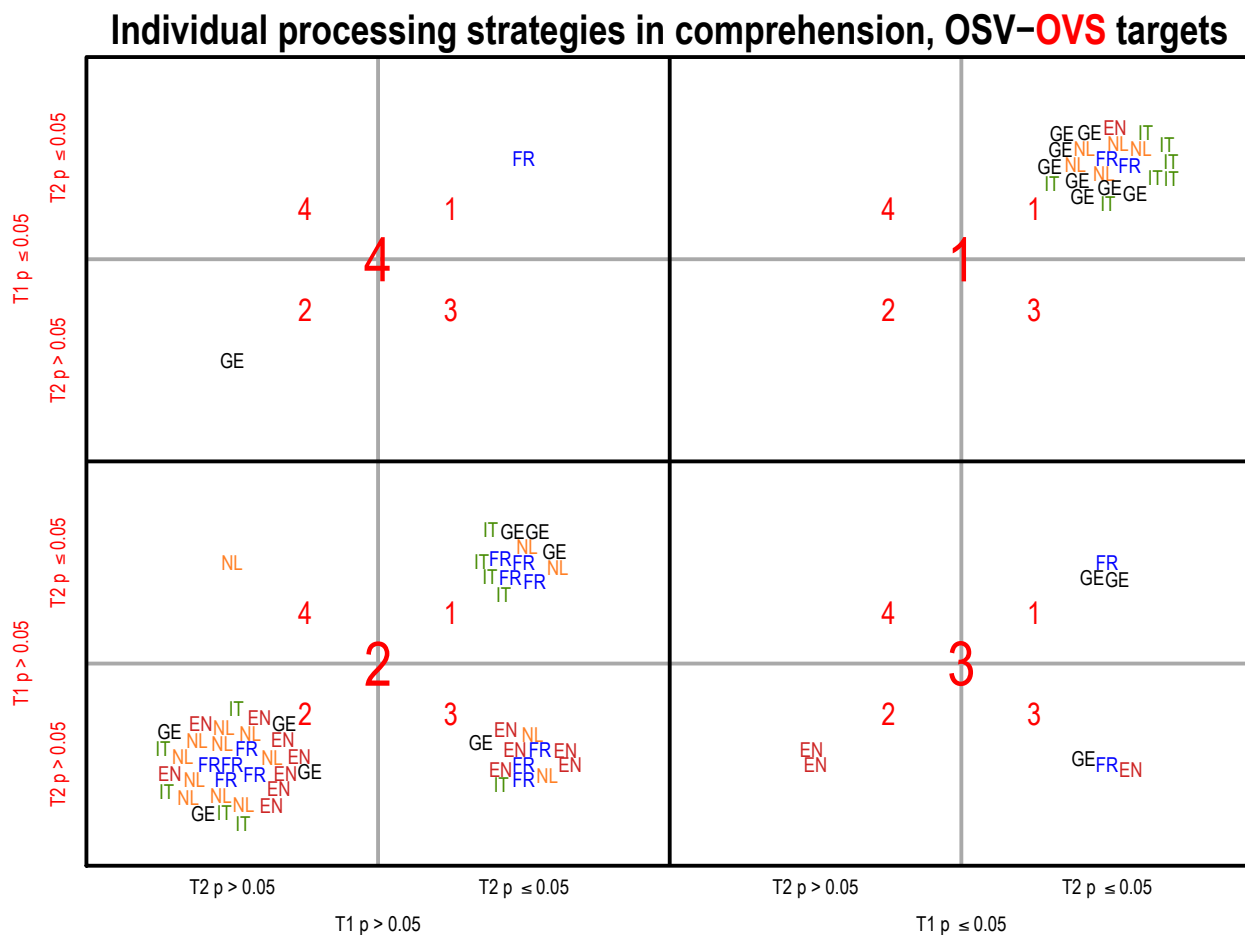
		OSV	
		$p > 0.05$	$p \leq 0.05$
OVS	$p \leq 0.05$	4	1
	$p > 0.05$	2	3

Scenario 1 indicates that the learner processes both OSV and OVS targets based on a morphosyntactic principle. Scenario 2 is the reverse, that is, both types of target are processed positionally, which naturally leads to an incorrect interpretation. In scenario 3, OSV targets are processed in a target-like manner, whereas OVS targets are not; the opposite happens in scenario 4. We are particularly interested in the last two scenarios, which suggest a discrepancy in the processing of the two types of OS target.

The main squares are further divided into four smaller squares each, which represent the same processing scenarios, in the same order, but relative to T2. In this way, we also have an indication of the evolution in time of learner processing strategies. Learners are identified by the two digits corresponding to their scenarios at T1 and T2, in that order, e.g. 2;1.

Since we do not have any *a priori* hypothesis as to whether any scenario should be more common, we proceed with no further delay to describing the data.

Graph 16⁵⁴



The two main clusters which can be identified concentrate in scenarios 2;2 and 1;1, both representing an extreme picture. In scenario 2;2, learners consistently process both types of OS targets based on a positional principle, that is, incorrectly. Their situation is stable between T1 and T2.

Conversely, learners in scenario 1;1 apply a target-like morphosyntactic strategy from the very beginning, and maintain it all the way to T2.

The two other clusters both start from scenario 2 at T1, which indicates a positional principle on both types of targets. However, these learners evolve differently with time: those in scenario 2;1 move on to scenario 1 at T2, which means that in the short time intercourses they learnt to generalise a

⁵⁴ This graph was created using the statistical software R (R Core team 2014) and the packages *wordcloud* (Fellows 2014) and *extrafont* (Chang 2014).

morphosyntactic strategy to all OS targets. Those in scenario 2;3 managed to do so only on OSV targets, and not on OVS ones. The reverse situation, with learners processing correctly OVS targets, but not OSV ones, is only instantiated by a single learner. This suggests that in diachrony, OSV targets are acquired first.

Synchronically, more learners appear in scenario 3 than in scenario 4 at both T1 and T2, as shown in Table 33 and Table 34 respectively.

Table 33: Comprehension test, OS targets, learner distribution across scenarios, T1

		OSV	
		$p \leq 0.05$	$p > 0.05$
OVS	$p \leq 0.05$	23	2
	$p > 0.05$	8	58

Table 34: Comprehension test, OS targets, learner distribution across scenarios, T2

		OSV	
		$p \leq 0.05$	$p > 0.05$
OVS	$p \leq 0.05$	40	1
	$p > 0.05$	15	35

To summarise, it would seem that OSV targets are indeed easier to process than OVS, for reasons which we will now try to identify.

In SVO targets (25), agent and patient appear in utterance-initial and utterance-final position respectively, whereas the verb is in utterance-medial position. As SVO is the dominant order in both the target language and the learners' L1s, this structure may be considered to be the prototype of transitive utterances.

- (25) a. *siostr-a* *woł**a* *brat-a*
 sister-NOM calls brother.ACC
 "(the) sister calls (her) brother"
- b. *brat* *woł**a* *siostr-ę*
 brother.NOM calls sister-ACC
 "(the) brother calls (his) sister"

The interpretation of this sort of utterances is quite natural when a positional principle is used: morphosyntax, on the contrary, may be a hard cue to use to extract meaning. This is because of the diffused syncretism involving the same ending across different paradigms, as in (25a). Here, both nouns are marked by the same ending *-a*, despite the different syntactic functions they are assigned. In fact, this is simply an unfortunate coincidence resulting from the fact that the nominative case of feminine nouns and the accusative case of animate masculine nouns both end in *-a*. In order to be applied successfully, the morphosyntactic principle thus requires that the learner knows the grammatical gender of the noun to process, as well as its inflectional paradigm. OVS targets also present the two nouns in utterance-initial and utterance-final position and the verb in utterance-medial position: this time, however, the patient comes first. This structure is therefore identical to SVO as far as the relative order of phrases is concerned. The only way to correctly process this type of targets, or to distinguish them from their SVO equivalents, is to process inflectional morphology.

- (26) a. *brat-a* *woł**a* *siostr-a*
 brother-ACC calls sister-NOM
 "(the) sister calls (her) brother"
- b. *siostr-ę* *woł**a* *brat*
 sister-ACC calls brother.NOM
 "(the) brother calls (his) sister"

Here too morphology may be tricky: in (26a), again, both nouns are marked by an identical ending, whose interpretation requires knowledge of the grammatical gender of the two nouns and of their paradigms, in addition to fairly good automatised processing to perform the task in real time. This, combined with the modest prominence of case endings, may easily confuse learners, leading them to mistake these targets for instances of SVO utterances. The point we are making here is that learners may simply mistake OVS targets for SVO ones because of their structural similarity: the pressure of online processing would cause items which look very similar to the prototype to be interpreted as such. In other words, whenever a learner encounters a sentence with the structure NP - V - NP, they interpret it as SVO. It may not be a chance that this tendency is particularly strong with English and French learners, that is, speakers of languages whose word order is particularly rigid, which in turn leads to a very stringent association between the linear order of phrases and meaning.

The picture changes entirely with OSV targets (27), in which the structure of the utterance is quite different: the two noun phrases come first, followed by the verb. This is an awkward constituent order which hardly ever appears in the input, and is therefore unfamiliar to the learners. This seems to be enough for them to notice the difference from the prototype, rather marked in fact, and pay attention to inflectional morphology, or perhaps interpret the utterance as OS simply because it appears so different from the prototypical SVO structure.

- (27) a. *brat-a* *siostr-a* *wola*
 brother-ACC sister-NOM calls
 "(the) sister calls (her) brother"
- b. *siostr-ę* *brat* *wola*
 sister-ACC brother.NOM calls
 "(the) brother calls (his) sister"

To conclude, we will now verify statistically the tendencies we have discovered so far in the data. Specifically, we will ask whether or not for any specific L1, the number of learners who can correctly process the two OS word orders is significantly different (Table 35). For each combination of L1 (column 1) and time (column 2), the table presents the number of learners performing at chance level and above chance on SVO targets (column 3 and 4), OVS (5 and 6) and OSV (7 and 8). The column "p OVS vs. OSV" indicates the probability that the two OS types of targets are processed similarly⁵⁵. A value less than 0.05 in this column would indicate a significant difference.

Table 35: Comprehension test, pairwise comparison: L1/WO interaction, OS targets

		SVO		OVS		OSV		
L1	Time	chance	morph	chance	morph	chance	morph	p OVS vs. OSV
EN	1	0	17	16	1	13	4	1.00
EN	2	0	17	16	1	10	7	0.39
FR	1	1	16	14	3	13	4	1.00
FR	2	3	14	9	8	5	12	1.00
GE	1	1	19	11	9	9	11	1.00
GE	2	2	18	7	13	5	15	1.00
IT	1	0	17	10	7	10	7	1.00
IT	2	0	17	6	11	5	12	1.00
NL	1	0	20	15	5	15	5	1.00
NL	2	5	15	12	8	11	9	1.00

As can be seen, all values but one are exactly 1, which indicates that speakers of each L1 process both OSV and OVS targets in a similar way, and that they have equal probability of success. The differences identified above, therefore, are not strong enough to indicate the existence of a regularity. Still, the data do show a tendency towards easier processing of OSV targets, which, although not

⁵⁵ This value was computed via Fisher tests with Holm correction for multiple comparisons.

obvious enough to reach statistical significance, may be attributed to the factors discussed in the qualitative analysis.

3.4.2. *Comprehension test: errors in the processing of SO targets*

In this paragraph we explore the possibility that some learners may have devised a morphosyntactic principle in which the relation between case endings and syntactic functions is inverted. In other words, these subjects would have associated case endings to syntactic functions, but according to a scheme which is the opposite of that required by the target language: in this scenario, *-/a/ NOM* corresponds to the object function, and *-/e/ ACC* corresponds to the subject.

To prove that a learner is indeed behaving according to this bizarre principle, we need the number of correct responses on all types of targets (SVO, OSV, OVS) to be not at chance level, but significantly below it⁵⁶, which corresponds to only 1 or no correct answers. As low scores on OS targets may also result from a simple positional principle, Table 36 presents the details of only those learners who failed to score above chance on SVO targets.

⁵⁶ Technically, that means that the probability of that result in the *lower* tail of the distribution is below 5%.

Table 36: Comprehension test, learners not above chance on SVO targets

Subject	Time	L1	p SVO	p OVS	p OSV
4120	2	GE	0.99	0.63	0.96
1119	2	FR	0.96	0.99	0.99
2115	2	NL	0.93	0.99	0.99
1118	2	FR	0.85	0.96	0.99
2110	2	NL	0.77	0.99	0.99
2102	2	NL	0.50	0.50	0.22
1101	2	FR	0.36	0.85	0.85
4112	1	GE	0.36	0.96	0.63
1119	1	FR	0.14	0.63	0.63
4106	2	GE	0.14	0.00	< 0.01
2109	2	NL	0.06	0.00	< 0.01
2114	2	NL	0.06	0.93	0.77

Only two of them scored significantly *below* chance on SVO targets; indeed, one of them (1119) did so on all types of targets, the other one (4120) on SVO and OSV, while performing at chance level on OVS targets. These two are the only learners for whom one can reasonably hypothesise an inverted morphosyntactic principle. If we exclude the usual suspects such as fatigue, lack of attention, boredom and so on, this rather unexpected result points to a restructuring of the system which did not go in the direction of the target language, but somewhere else. In other words, these learners might have formulated hypotheses which are not consistent with the structure of the target stimuli, and led to errors. This situation is not as uncommon as it might seem: some models of language acquisition, in fact (e.g. Carroll 2001), predict that acquisition results precisely from a failure of the present principles of utterance organisation, detected through a mismatch between the resulting interpretation of the input and their knowledge of the present situation. In other words, they know that what they have understood must be wrong, and this leads to a further restructuring of the system, possibly in the direction of the target

language. Of course, our learners did not have a realistic communicative context against which to check the result of their processing, so that they receive no feedback as to its efficiency.

All other learners scored at chance level on SVO targets, sometimes only marginally so, and have various outcomes on the other word order targets, ranging from a clear positional principle, to random behaviour, all the way to a morphosyntactic principle. In the latter case, the failure to reach significance on SVO targets should probably be interpreted as due to contingent reasons of little linguistic relevance. 2109 and 4106 have top scores on OS targets, and fail to reach significance on SO targets only by a slight margin. In fact, this means that these learners only got one or two responses too many wrong on SO targets, and we can imagine that they were, in fact, applying a rule, rather successfully at that. Others also have high scores on SVO targets, although their performance on OS is not as good (1119 at T1, 2114). The rest have low scores on all targets, which suggest a lack of any strategy and perhaps of engagement with the test altogether.

3.5. COMPREHENSION TEST: AN INTERIM SUMMARY

This chapter described the results of the Comprehension test, in which learners were asked to listen to a short transitive utterance and to identify its syntactic structure by selecting the appropriate picture. As we attempted to identify the principle of utterance interpretation adopted by learners in the processing of target sentences, we came across two options: a target-like morphosyntactic principle, and the default positional principle. The former requires that the learner associates case endings to grammatical meaning, independently of their position in the utterance; the latter, in contrast, systematically assigns the subject function to the first noun phrase, and that of object to the second one. In order to verify what principle learners refer to, we observed their treatment of OS targets, which, due to their syntactically marked word order, can only be correctly processed by adopting the morphosyntactic principle.

To this purpose, we tried to verify whether learners, taken on an individual basis rather than as a whole group, performed the test according to a morphosyntactic strategy as opposed to a positional strategy. The picture which emerged confirms the first impression obtained from the descriptive statistics. SVO targets are processed with far greater accuracy than their OSV and OVS equivalents. Second, within OS targets, OSV seem to be more accurately processed than OVS, or better, the latter seems to be more consistently interpreted as a subject-first structure. Regarding the effect of time, learner processing strategies overall evolve in the direction of the target language, although unexpected errors in the processing of syntactically unmarked SVO targets were also found.

Finally, the source language seems to exert a relevant influence on the learners' processing strategy. Speakers of L1s whose syntax is rather rigid, like French and English, tend to perform more poorly than those whose L1 admit OS structures too. The English learners stand out particularly in this respect as they consistently adhere to a positional principle when processing OVS targets, showing no sign of evolution over time.

To verify these claims statistically, we ran a generalised linear model which found significant effects for word order, time and L1, as well significant interactions between word order and L1, on the one hand, and word order and time, on the other hand.

The former is mainly due to the English L1 group, and can be described as such: while at T1 all groups show a clear difference in their performance on SO targets, on the one hand, on OS ones, on the other hand, at T2 this difference is no longer significant for most L1 groups. In fact, this trend is common to all learners, with the sole exception of the English, who still show a very clear preference for the agent-first interpretation of all targets. While the other four languages are by no means homogeneous with respect to traits such as the presence of case on noun phrases and the structure and rigidity of word order, for the time being one could safely identify the cause of the English learners' difficulty in the very rigid SVO syntax of their L1. We hypothesise that this

trait may represent an obstacle to acquiring a new system based on the category of case in association with potentially free word order.

The interaction between word order and time concerns the evolution over time of the strategies employed by learners to process targets of different word order. The number of learners correctly processing both OVS and OSV targets increases between T1 and T2, which indicates that over time, more and more subjects learn to correctly extract meaning from these structures by applying a morphosyntactic principle. The proportion of learner correctly processing SVO targets, in contrast, decreases between the two test times. This result is quite unexpected, as SVO targets should not pose any particular difficulty. A possible explanation is that at least the learners in question have become so aware of the presence of OS targets in the L2 and in the test, as to over-generalise this pattern to unmarked targets as well.

Finally, we turned to a pair of scenarios which do not seem to fit completely in the picture presented so far. The first concerns the alleged differential processing of OSV and OVS targets, in which the former seems to be favoured, and which was not envisioned by our initial hypotheses. Qualitative analysis of the two structures suggests that the reason for such discrepancy might lie in the strong resemblance of SVO and OVS structures, which in fact can only be distinguished by the relative position of case endings in the utterance. Now, case endings are not particularly prominent, may or may not be attended to by initial learners, and because of widespread syncretism across paradigms, they are not necessarily unambiguous. In sum, it may be that any structure constructed according to the sequence NP - V - NP is interpreted by some learners as SVO, whereas OSV targets, which clearly deviate from this pattern, are more easily interpreted as marked in terms of structure and meaning. In spite of the tendencies observed, however, statistical analysis did not find significant differences between the processing of the two OS word orders.

Finally, we turned again to those learners who did not seem to consistently process SVO targets correctly. We verified whether it could be maintained that they developed an inverted morphosyntactic system, that is, one in which the

target association between case endings and syntactic functions is reversed. In order to fall within this scenario, learners had to have improbably low scores on all types of targets: SVO, OSV, OVS. In fact, it turns out that only two learners potentially fit this picture.

To conclude, we could select the following facts as the most relevant outcome of our analysis. First, as hypothesised, SO structures are more easily interpreted than their OS equivalent. Second, the learners' familiarity with OS targets increases with time, so that by T2, a good half of the subjects can consistently process all target structures. Finally, all L1 groups behave in a comparatively similar manner with the only exception of the English learners, who appear to be much more biased towards a agent-first interpretation of any target structure.

4. AN OVERALL PICTURE OF MORPHOSYNTACTIC COMPETENCE

The purpose of this chapter is to verify whether or not learners acquired a morphosyntactic principle of utterance organisation, according to which syntactic functions are encoded by case endings independently of word order. We will attempt to do so by correlating the results of the Repetition and the Correlation tests discussed earlier. As Pallotti (2006:373) states,

"Validation studies are fundamentally based on the 'triangulation' of various methods. The fact that a structure has emerged could thus be demonstrated on the basis of several elicitation procedures, which can be similar (e.g. a range of communicative tasks) or different."

Indeed, the methodology we are now going to apply makes it possible to verify whether or not the target morphosyntactic principle is already used by the learner, and if so, if it developed at the same time in both abilities which are founded on it, namely comprehension and production, or if on the contrary it is more prone to developing in either of them first.

4.1. THE COMPREHENSION TEST AS A DISAMBIGUATOR TO THE REPETITION TEST

Before we start, it is useful to show one of the most immediate ways in which the two tests in combination can provide us with a clearer picture of the learners' processing. Specifically, the comprehension test can help us to answer some of the questions which emerged from the analysis of the repetition test. Given an utterance like (28), it is impossible to establish *a priori* whether or not the learner truly attempted to encode some specific meaning, and, if so, what this may be.

(28) [artistk-a pozdravja twumatʃk-a]
 artist-NOM cheers interpreter-NOM

The question is particularly important in the case of syntactically marked OS targets, which, as shown in chapter 3, are themselves difficult for learners, so that one may argue that learners can fail not only in the repetition of the marked ACC ending, but also in the interpretation of the target sentence, which could be processed as a default subject-initial utterance. Given a target utterance like (29):

(29) /artistk-e pozdravja twumatʃk-a/
 artist-ACC cheers interpreter-NOM
 "the interpreter cheers the artist"

the output observed in (28) may represent the realisation of at least three underlying structures (30):

(30) a. /artistk-a pozdravja twumatʃk-a/
 artist-OBJ cheers interpreter-SUBJ
 b. /artistk-a pozdravja twumatʃk-a/
 artist-SUBJ cheers interpreter-OBJ
 c. /artistka/ /pozdravja/ /twumatʃka/
 ARTIST TO CHEER INTERPRETER

(30.a) corresponds to target-like comprehension of the OS target. In (30.b), in contrast, the utterance is interpreted as subject initial. In (30.c), finally, no grammatical meaning is encoded, and the learner only retrieves and produces isolated lexical items with no syntactic link among them. In the latter case, the utterance has no overall meaning and it is doubtful whether it should be considered as an instance of communication, or rather as the mere execution of a purposeless exercise.

Unfortunately, in the absence of a translation test it is impossible to shed light on these doubts directly. The comprehension test, however, represents an indirect way to at least reduce our degree of uncertainty as to the learners' behaviour. More precisely, if the results of the comprehension test show that a particular learner is incapable of processing OS targets, then it is highly unlikely that the same learner should process them correctly in the repetition test. In this case, then, the learner might have attempted to encode an SO utterance, at most, or no meaning at all, in the worst scenario: in other words, we could at least exclude option (30.a). In this scenario, the source of incorrect output lies in comprehension, most probably, and in repetition, quite surely. If learners performs above chance on the comprehension test, in contrast, the possibility exists that they might have tried to produce a syntactically marked utterance, while at the same time failing to encode grammatical meaning through case endings. The difficulty, then, should be localised at the level of repetition, rather than comprehension.

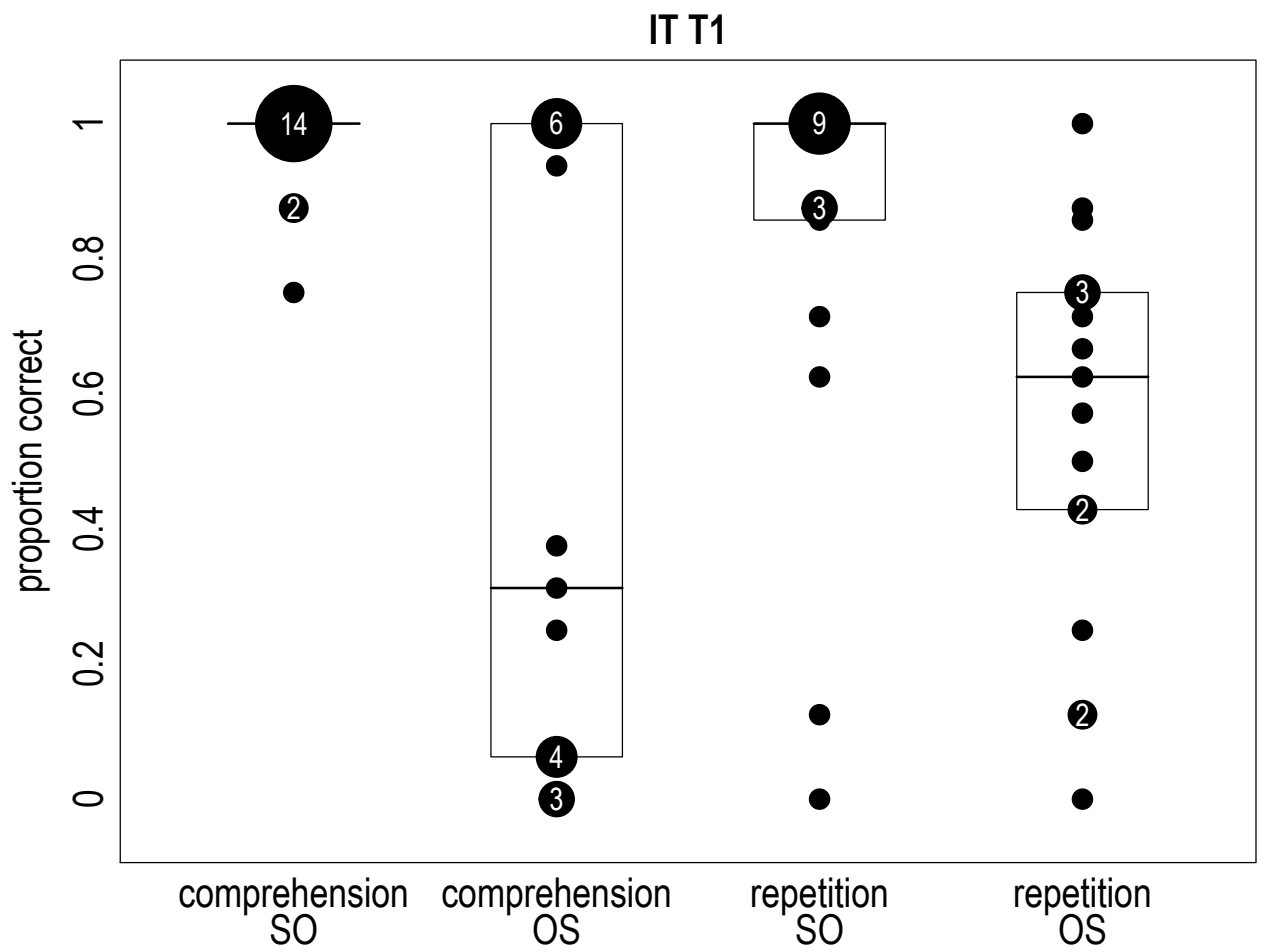
Even so, correlating the two tests does not provide us with definite answers as to the learner's strategies of utterance organisation when producing output. On the contrary, this methodology simply allows us to exclude one possibility out of the various scenarios one could envisage to justify the output observed (see 2.7, p. 88). To summarise, the comprehension test does tell us whether or not it is likely that a learner could have correctly processed OS targets in the repetition test, but it does not tell us that that necessarily happened. In view of future studies, we learn from this experience that a translation test would seem a more appropriate tool to disambiguate the results of a repetition test.

For the time being, we will attempt to correlate the two tests in order to obtain a complete picture of the degree to which learners are able to manipulate case endings. Rather than on the interpretation of ambiguous results, as was the case with utterance (28) above, this analysis is based on the visible effects of case processing in comprehension and repetition, namely correct interpretation of target utterances and correct repetition of target endings.

4.2. CORRELATING THE REPETITION AND COMPREHENSION TESTS: METHODOLOGY

When correlating the results of the repetition and of the comprehension test, we might be tempted at first to plot learners' scores in the various conditions they were tested on (task, word order, time) side by side, as in Graph 17 below, which present the scores of the Italian group at T1.

Graph 17



This sort of representation, however, has several drawbacks. First of all, as already pointed out in section 2.5 (p. 70), differences in mean scores, however significant statistically, do not necessarily have a linguistic interpretation. Second, the behaviour of individual learners cannot be traced in the various conditions, as data points do not identify specific learners across the various

conditions. Finally, each graph only describes the performance of a single L1 group at either test time. It would take 10 such graphs to fully describe tendencies in the data-set, with all the drawbacks just stated.

To remedy the first problem, Table 37 presents the proportion of learners scoring statistically above chance for each test, word order value, and test time, based on the methodology described in section 2.5, p. 70.

Table 37: learners scoring above chance

	OS								SO							
	T1				T2				T1				T2			
	CS		SI		CS		SI		CS		SI		CS		SI	
	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
EN	1	16	0	17	1	16	0	17	17	0	0	16	16	0	0	16
FR	4	13	0	17	9	8	2	15	16	1	1	16	14	3	5	12
GE	10	7	7	13	15	5	11	9	19	1	8	12	18	2	9	11
IT	7	10	6	11	11	6	8	9	17	0	13	4	17	0	14	3
NL	5	15	2	18	8	12	5	15	20	0	6	14	15	5	7	13

This time the data are arranged and interpreted in such a way that they have an immediate linguistic interpretation, namely whether or not the learners (most probably) applied a morphosyntactic strategy.

Again, however, we are left with no information as to the score of individual learners in various condition, which is essential to our understanding of the co-occurrent manifestations of morphosyntactic competence.

4.3. SCENARIOS

To this end, we introduce the methodological tool of 'scenarios'. Scenarios represent a global score for learners' processing skills in both comprehension and repetition. For each test, scores are considered to be 'good' if they are significantly above 50% according to a binomial distribution identified by the

number of correct responses and the total number of responses, as described in greater detail in section 2.5, p. 70. The resulting combinations are presented in Table 38.

Table 38: scenarios, rationale

		comprehension	
		+	-
repetition	+	1	4
	-	3	2

Scenarios 1 and 2 correspond to a correlation between the two tests, that is, both tests are performed above chance or at chance level respectively. Scenarios 2 and 4, in contrast, point to a dissociation between the two abilities, i.e. a situation in which learners perform well in one and poorly in the other one. In general, this makes it possible to pursue the question of whether the two skills are correlated in the learners' competence. The alternative hypothesis is that either skill might develop earlier in time.

We exemplify the use of this tool by plotting the results obtained by all learners at T1 on OS targets. The area of Graph 18 below is divided into four squares. Each corresponds to a scenarios, which is indicated by the large number in red. Learners are identified by a coloured dygraph according to their L1 (Table 39).

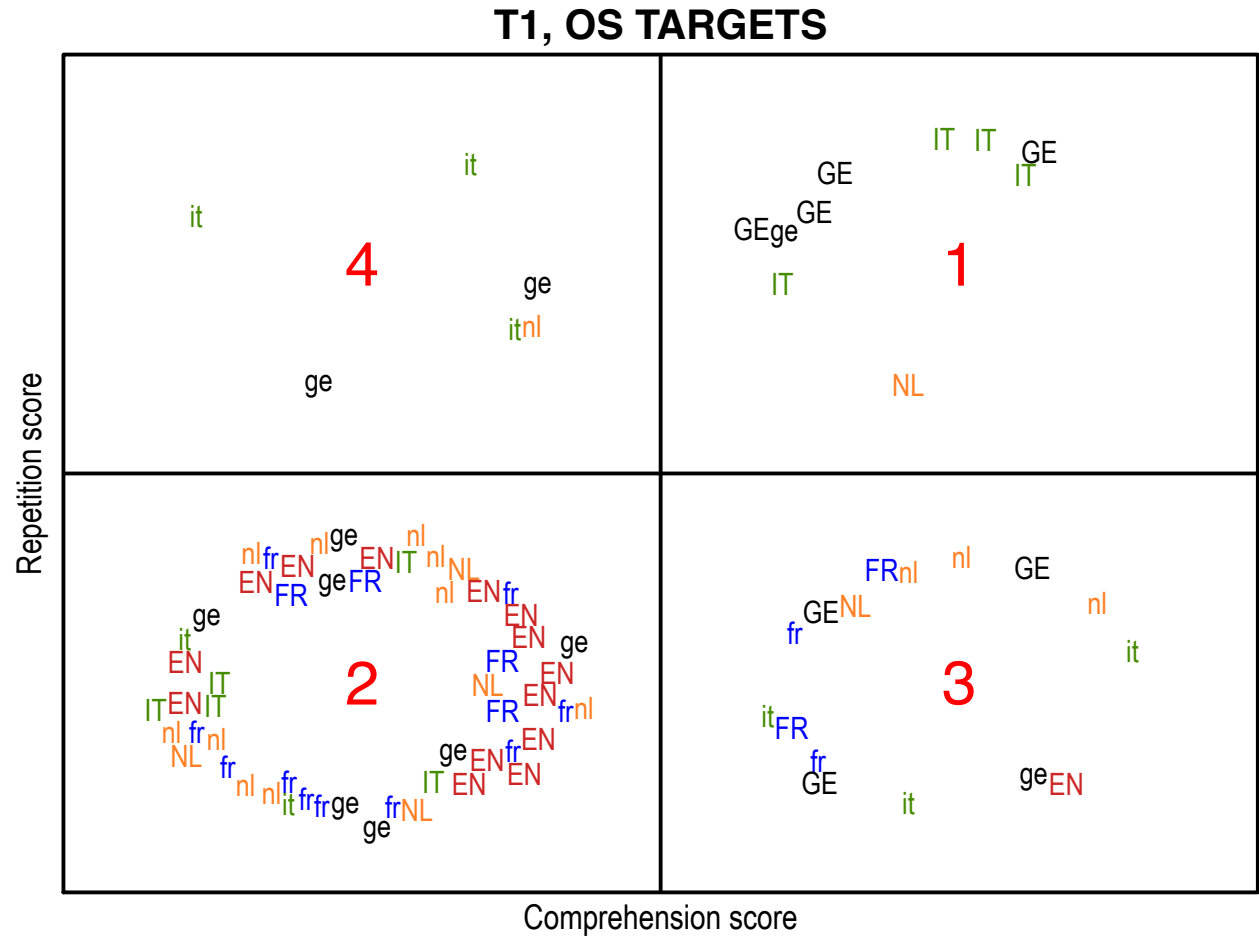
Table 39: Graph 18, identifiers

L1	English	French	German	Italian	Dutch
Identifier	EN	FR	GE	IT	NL
Colour	red	blue	Black	green	orange

Further, depending on whether or not the performance of the learner in question varies from T1 to T2, the dygraph is printed in lowercase or uppercase letters, respectively. Said otherwise, learners identified by capital letters will no

longer be in the same position in the graph depicting the situation at T2 (Graph 21). See section 4.3.1, p. 143 for a detailed discussion of the effect of test time. Thus, learners are not identified by their univocal VILLA ID code, but only by their L1. As will become clear in the following few pages, this is not a problem as participants are identified univocally by their position in a single graph, so that no information is lost, differently from the graphical representation presented in the beginning of this chapter (Graph 17, p. 132).

Graph 18⁵⁷



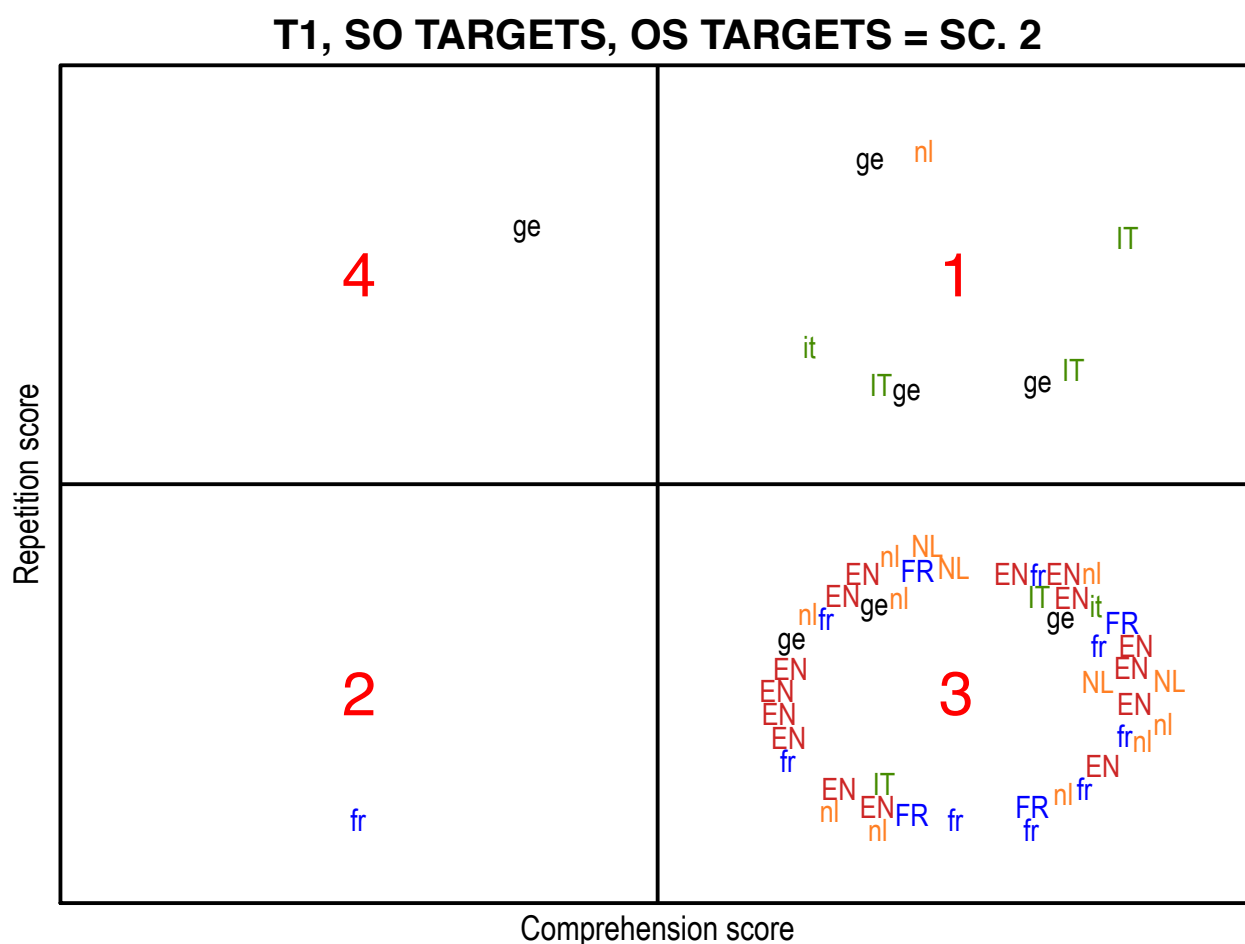
⁵⁷ This graph was created using the statistical software R (R Core team 2014) and the packages *wordcloud* (Fellows 2014) and *extrafont* (Chang 2014).

By far the most obvious cluster, comprising more than a half of the dataset, is located in scenario 2, indicating that neither test was performed with above chance accuracy. The second largest cluster corresponds to scenario 3, indicating that the comprehension test is performed above chance, but not the repetition test. Finally, an even smaller group of learners can be found in scenario 1, indicating that already at the first test time, 10 learners managed to process OS targets morphosyntactically in both comprehension and repetition. Coherently with the assumptions of the repetition test, finally (see section 2.2, p. 56), very few learners are found in scenario 4, which corresponds to above chance performance in repetition, but not comprehension.

The picture presented so far is still incomplete, as it only shows learner performance on a single type of targets, namely OS structures. In contrast, it would be desirable, if we are indeed to provide a comprehensive picture of learners' processing, to have a synoptic representation of their performance on SO targets as well.

We exemplify our claim by focussing on the learners located in scenario 2 in the preceding graph. Graph 19 depicts their performance on SO targets following the same rationale, so that the graph reads in the same manner described above, the only difference being in the type of target structures considered.

Graph 19⁵⁸



Again, an obvious cluster can be identified, this time in scenario 3. This is hardly surprising as SO targets can be easily processed using either a positional or a morphosyntactic principle: no wonder then that most learners perform well in the comprehension test. The exiguity of data points in scenario 1, in contrast, witnesses to the much greater difficulty of the repetition test, although 8 learners do exhibit above chance accuracy. Scenario 2 and 4, finally, run contrary to the assumptions of both test rationale and word order manipulation, and are coherently empty.

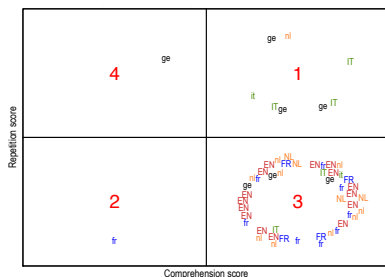
The next step consists in merging the information presented by Graph 18 and Graph 19 into a single, comprehensive representation. This will be achieved as

⁵⁸ This graph was created using the statistical software R (R Core team 2014) and the packages *wordcloud* (Fellows 2014) and *extrafont* (Chang 2014).

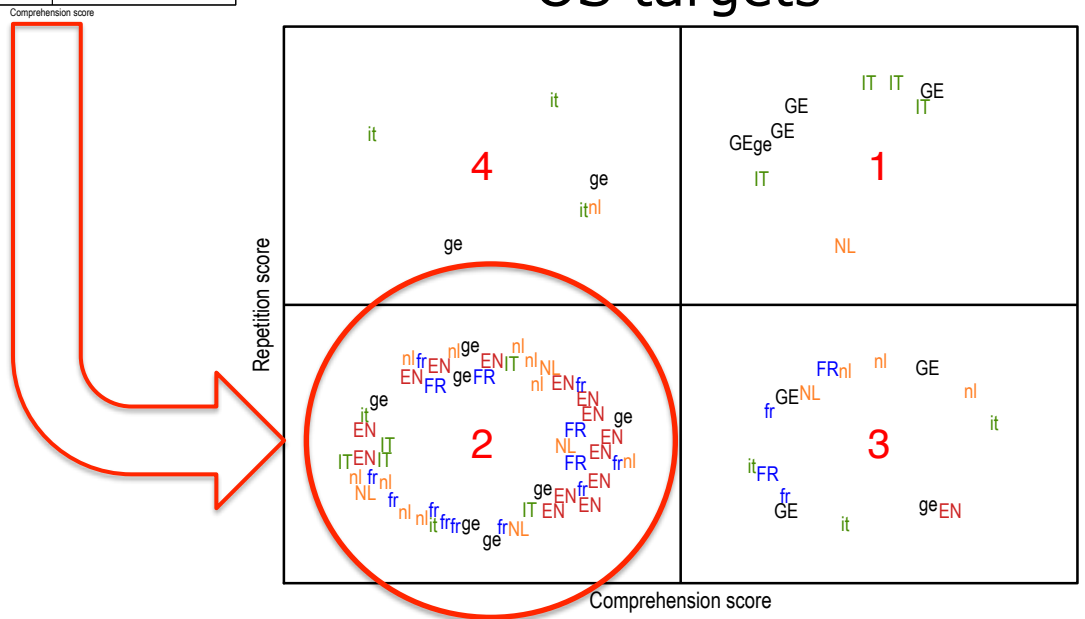
shown in Figure 4. The main squares of the graph representing the processing of OS targets will be further divided into 4 smaller squares, representing the processing of SO targets. Both representations make use of scenarios, arranged in the same order (clockwise: 4,1,3,2) regardless of the type of targets.

Figure 4

SO targets, OS = 2

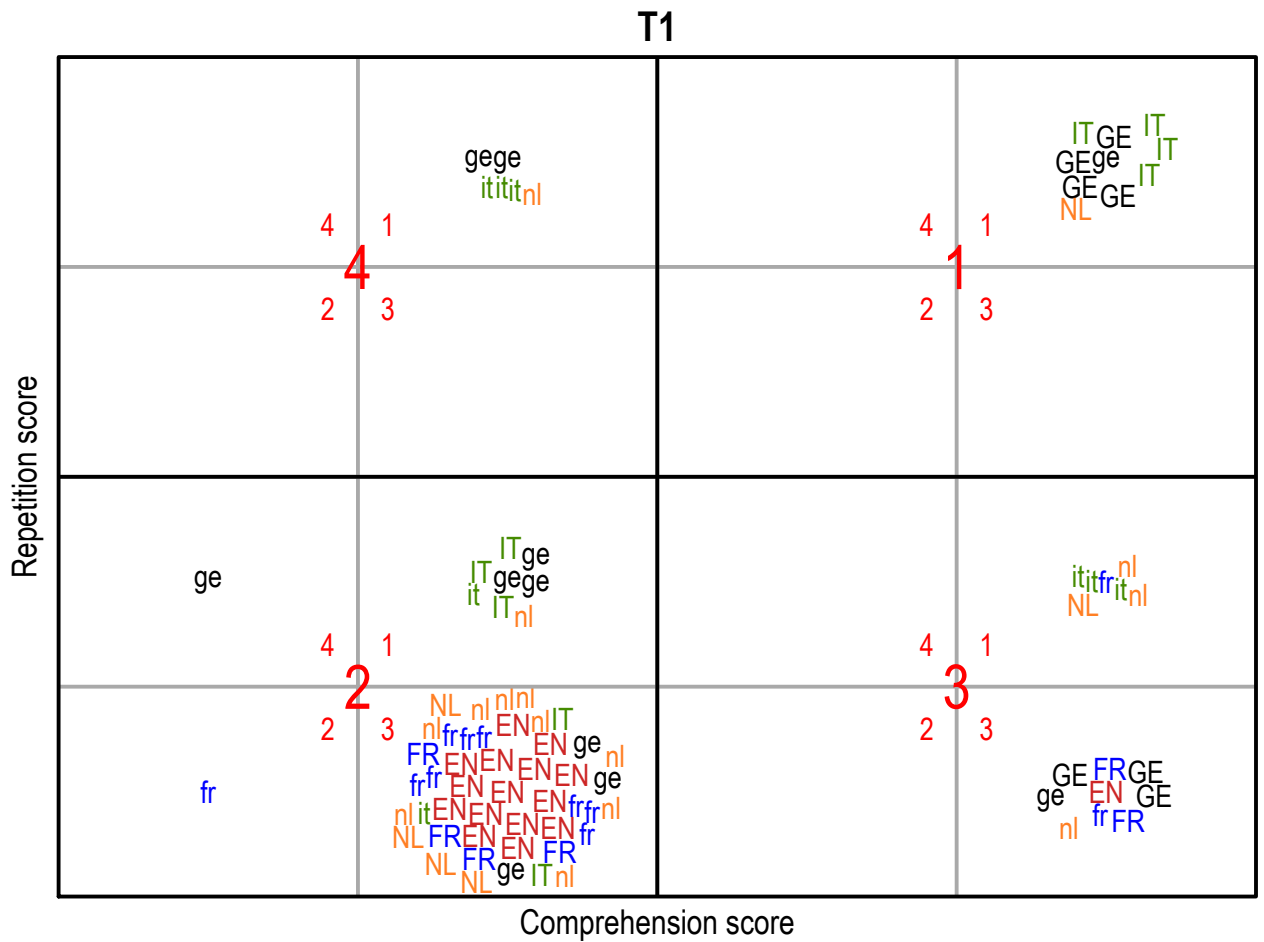


OS targets



The final result is shown in Graph 20.

Graph 20⁵⁹



The graph area is divided into a total of 16 squares, which correspond to unique combinations of OS and SO processing scenarios. We will identify them by providing two coordinates, namely the OS scenario (large number) followed by the SO scenario (smaller numbers). Scenario 2;3, for instance, corresponds to the largest cluster observed (second from left, bottom row).

Again, some of the resulting scenarios are theoretically unmotivated, and are accordingly empty.

Below we provide a rationale of the scenarios which are theoretically motivated and in which data points are effectively encountered.

⁵⁹ This graph was created using the statistical software R (R Core team 2014) and the packages *wordcloud* (Fellows 2014) and *extrafont* (Chang 2014).

- 1;1 both tests are performed with above chance accuracy on both types of targets. This scenario corresponds to a full morphosyntactic principle;
- 3;1 with SO targets, both tests are performed above chance; with OS targets, only the comprehension test is. This scenario shows that if only one ability is missing, most probably that is going to be the repetition of OS targets, which can be thus identified as the hardest task for the learners;
- 2;1 both tests are performed with above chance accuracy on SO targets; on OS targets, neither is. This scenario suggests that independently of the test, OS targets are harder than SO ones;
- 2;3 only the comprehension test on SO targets is performed with above chance accuracy. This scenario corresponds to the typical positional principle.

The scenarios identified in Graph 20 are represented analytically in Table 40. Each is broken down into its task and word order components, with the intention of identifying the combination that affects performance the most. To this purpose, we simply compute the number of learners who perform above chance in each context (last row).

Table 40: implicational hierarchy at T1

OS		SO			
repetition	comprehension	repetition	comprehension	scenario	n. (tot = 88)
+	+	+	+	1;1	10
-	+	+	+	3;1	7
-	-	+	+	2;1	8
-	-	-	+	2;3	47
+	-	+	+	4;1	5
-	+	-	+	3;3	9
-	-	+	-	2;4	1
-	-	-	-	2;2	1
15	26	31	86		

A hierarchy emerges. In the scenarios comprised in it (in black), all the cells to the right of the leftmost + sign are also occupied by + signs. The linguistic interpretation is that if the leftmost task is performed with above chance accuracy, then all tasks to its right must be performed above chance as well, along the following implicational path:

OS repetition \supset OS comprehension \supset SO repetition \supset SO comprehension

This regularity holds for the vast majority of the learners considered (72 on 88). A few observations are possible. To start, either task performed on OS targets is harder than either performed on SO ones. Within a given value of constituent order, further, the repetition test is harder than the comprehension test. It is also useful to signal that all scenarios which are coherent with our initial hypotheses concerning constituent order and the repetition test are indeed part of the hierarchy. There is an exception to this rule, however: scenario 3;3, comprising 9 learners, is not in agreement with the hierarchy and yet does not violate any assumption:

3;3 the comprehension test is performed above chance with both SO and OS targets, whereas the repetition test is always at chance level.

This scenario suggests that independently of the target structure, the repetition test is harder for the learners than the comprehension test. Truly, it was not predicted that if a learner is able to process OS targets in comprehension, he or she should be able to process SO ones in repetition. However, the vast majority of the learners appear to follow this pattern. In fact, scenario 3;3 comprises 9 learners, whereas the closest scenario compatible with the hierarchy, namely 3;1 (success in comprehension on OS targets; success in both tests on SO targets) comprises 7 learners, so that the two possibilities seem to be equally likely. One could conclude that whether a learner at a given developmental stage should be comprised in scenario 3;1 or 3;3 is a matter of learning style: specifically, some learners may be more prone to start speaking - at all, or in a grammatically correct fashion - earlier than others, who may prefer to focus on comprehension for a longer time. Indeed, the so-called "silent period", common to both L1 and L2 acquisition, may vary dramatically in length from person to person (Krashen 1985, Granger 2004).

Another 5 learners are found in scenario 4;1:

4;1 with SO targets, both tests are performed above chance; with OS ones, however, the repetition test is above chance, and the comprehension test is at chance level.

This contradicts the assumptions of the repetition test, according to which repetition requires that the target sentence should be first understood. It is quite clear that the test did not work properly and learners repeated the target sentences on the basis of their phonological memory alone. The distractor (see 2.2, p. 56) proved insufficient to engage the phonological loop, so that the participant effectively repeated a string of sounds in its phonetic details

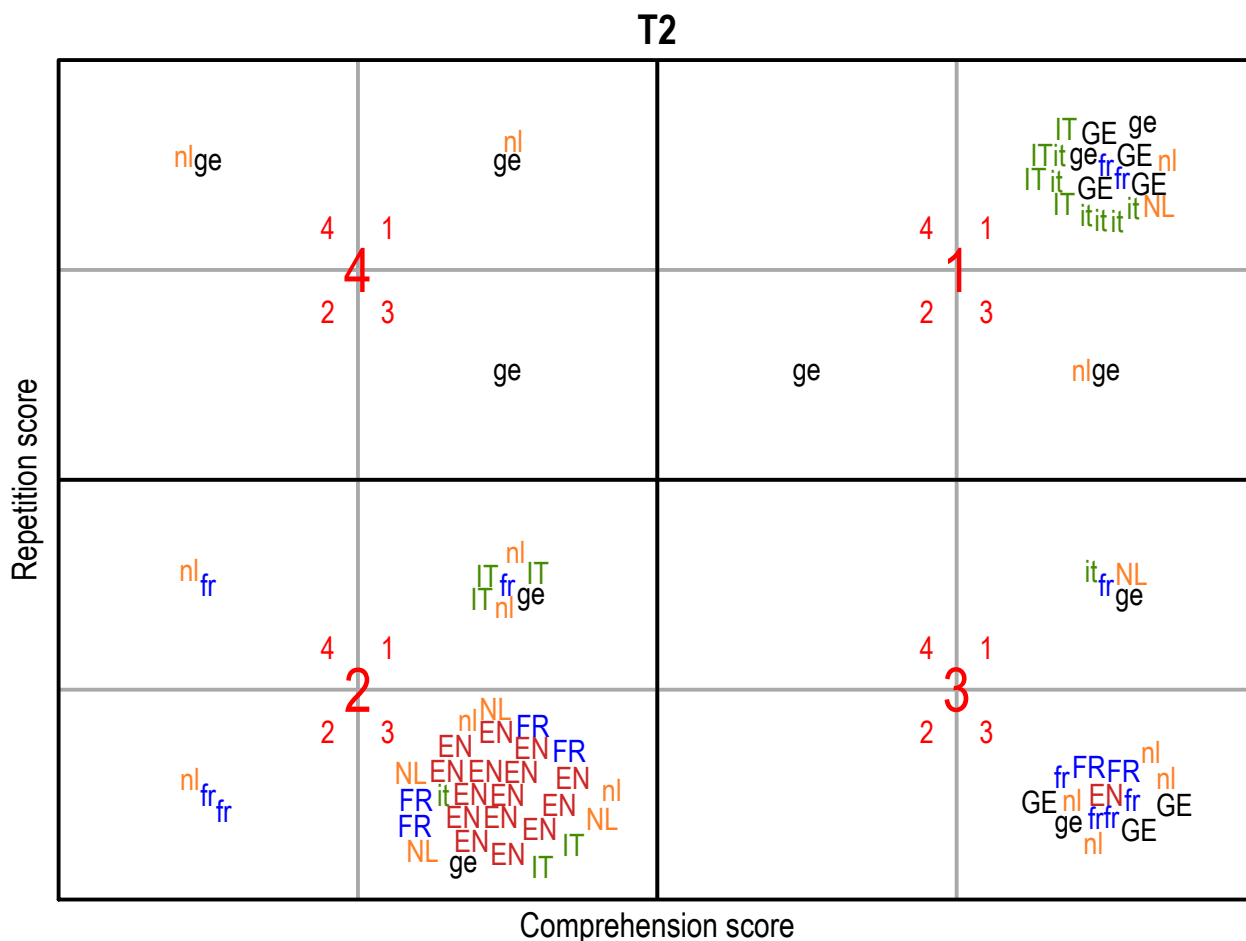
without processing the target sentence for meaning. More demanding distractors typically require test takers to perform verbal or intellectual tasks, such as counting from 20 to 29, doing simple calculations, answering questions, or pronouncing a series of words. In contrast, the distractor used here only consisted in drawing a simple geometrical figure. Great individual variability can be observed in the approach that learners adopted to this task. Since no explanation was provided as to the role of the distractor phase, some learners indeed took it very seriously and drew the geometrical figure with great care, as though that was an essential part of the test. It often took them so long that phonological memory would decay spontaneously, so that the distractor could be said to be effective. Others, in contrast, tried to be as quick as they could with the drawing and focussed instead on repetition. The latter approach may allow the test taker to rely on short-term memory, which, combined with some phonological sensitivity, could explain scenario 4;1. This possibility is explored in greater detail in section 4.2, p. 149.

The two remaining scenarios comprising a single learner each (2;4, 2;2) make little sense linguistically, and are probably due to a lack of commitment on the side of the participants.

4.3.1. Effects of additional exposure to the input

The picture presented above describes the situation at T1, that is, after 9 hours of exposure to the input. As the test was administered twice, it is worth applying the same procedure to the scores obtained at T2, that is, after an additional 4h30 of instruction. Graph 21 below presents the data with the aid of the same structure employed in the previous graphs.

Graph 21⁶⁰



The main patterns observed at T1 seem to hold at T2 as well. Specifically, the largest cluster is still located in scenario 2;3, although the proportion of learners comprised in it has decreased. The cluster corresponding to a full morphosyntactic principle (1;1), in contrast, nearly doubled in size, suggesting that additional exposure indeed steers the developing linguistic system towards the target variety. Finally, greater dispersion is observed than at T1, which could be taken as evidence of various autonomous strategies of input processing being developed by learners as they test hypotheses as to the structure of the target language.

The same tendencies are highlighted in Table 41 below.

⁶⁰ This graph was created using the statistical software R (R Core team 2014) and the packages *wordcloud* (Fellows 2014) and *extrafont* (Chang 2014).

Table 41: implicational hierarchy at T2

OS		SO			
repetition	comprehension	repetition	comprehension	scenario	n. (tot = 88)
+	+	+	+	1;1	20
-	+	+	+	3;1	4
-	-	+	+	2;1	7
-	-	-	+	2;3	29
+	+	-	+	1;3	2
+	+	-	-	1;2	1
+	-	+	+	4;1	2
+	-	-	+	4;3	1
+	-	+	-	4;4	2
-	+	-	+	3;3	15
-	-	+	-	2;4	2
-	-	-	-	2;2	3
28	42	37	80		88

While cluster analysis may not prove particularly insightful, it may be worthwhile to try and detect any regularities in the evolution of learner processing strategies over time. We are now going to verify whether there exist in the data any preferential paths of evolution which learners may take when provided with additional input. Such paths will have the following structure:

T1 coordinates - T2 coordinates

In other words, paths are simply a combination of the scenarios in which a learners is found at T1 and T2. Two research questions will guide our analysis:

1. Is there an implicational sequence of development? It may be the case that in order to improve in one particular ability (e.g. the repetition of

OS targets), learners must be able to master others first (e.g. the comprehension of OS targets and the repetition of SO ones). This requires that the implicated abilities should be already developed at T1, or develop between the two test times. This hypothesis is based on the synchronic implicational scale described in the preceding section.

2. Can cross-linguistic differences be observed? This question really regards possible interactions between the implicational scale of development mentioned in the previous research question, if it should be found, and the learners' L1. It may be the case, for instance, that speakers of a certain L1 improve on particular tasks rather than others, perhaps because their native language facilitates or hinders them.

Table 42 is the methodological tool which we will use to explore these questions. The first column ("pattern") lists all the combinations of scenarios at T1 and T2 observed at T2. Note that this set, composed of 30 items, is only a small part of the possible set of combinations, which amounts to 256 patterns. The second columns shows the proportion of learners who follow each pattern, along a decreasing order. In the following columns, proportions are provided for each native language.

Table 42: patterns of morphosyntactic processing over time

	tot	EN	FR	GE	IT	NL
pattern	(n = 88)	(n = 16)	(n = 17)	(n = 18)	(n = 17)	(n = 20)
2_3 2_3	0.28	0.94	0.24	0.00	0.12	0.20
1_1 1_1	0.10	0.00	0.00	0.22	0.24	0.05
2_3 3_3	0.08	0.00	0.24	0.06	0.00	0.10
3_3 3_3	0.07	0.06	0.12	0.17	0.00	0.00
3_1 1_1	0.05	0.00	0.06	0.00	0.18	0.00
2_1 2_1	0.03	0.00	0.00	0.00	0.18	0.00
2_3 1_1	0.03	0.00	0.00	0.06	0.06	0.05
2_3 2_1	0.03	0.00	0.06	0.00	0.00	0.10
2_3 2_2	0.03	0.00	0.12	0.00	0.00	0.05
2_1 2_3	0.02	0.00	0.00	0.00	0.06	0.05
2_3 4_4	0.02	0.00	0.00	0.06	0.00	0.05
3_1 3_3	0.02	0.00	0.00	0.00	0.00	0.10
4_1 1_1	0.02	0.00	0.00	0.00	0.12	0.00
1_1 1_2	0.01	0.00	0.00	0.06	0.00	0.00
2_1 1_1	0.01	0.00	0.00	0.06	0.00	0.00
2_1 1_3	0.01	0.00	0.00	0.06	0.00	0.00
2_1 4_1	0.01	0.00	0.00	0.06	0.00	0.00
2_2 2_4	0.01	0.00	0.06	0.00	0.00	0.00
2_3 2_4	0.01	0.00	0.00	0.00	0.00	0.05
2_3 3_1	0.01	0.00	0.06	0.00	0.00	0.00
2_3 4_1	0.01	0.00	0.00	0.00	0.00	0.05
2_4 2_3	0.01	0.00	0.00	0.06	0.00	0.00
3_1 3_1	0.01	0.00	0.00	0.00	0.00	0.05
3_3 1_1	0.01	0.00	0.06	0.00	0.00	0.00
3_3 1_3	0.01	0.00	0.00	0.00	0.00	0.05
3_3 3_1	0.01	0.00	0.00	0.06	0.00	0.00
4_1 2_1	0.01	0.00	0.00	0.06	0.00	0.00

4_1 2_3	0.01	0.00	0.00	0.00	0.00	0.05
4_1 3_1	0.01	0.00	0.00	0.00	0.06	0.00
4_1 4_3	0.01	0.00	0.00	0.06	0.00	0.00

The first striking observation regards precisely the lack of obvious patterns in the data. Second, among the five most common patterns, comprising 28% and 10% of the data respectively, are patterns in which no change takes place. If we want to draw a generalisation, then, we could say that the bulk of learners tends not to evolve over a 4h30 period.

Next, the English L1 group is the most consistent in its evolution patterns. No improvements at all takes place, all learners but one being found in scenario 2;3, corresponding to a clear positional principle. All other language groups show much greater dispersion, with most clusters comprising as little as a single learner.

The data suggest the following two conclusions. First, the data set for each L1 group is too limited to identify any significant tendency. Second, the time elapsed between the two test times is probably too short, so that what we observe is in fact a picture taken from a collection of individual linguistic systems in fluid development, whose not necessarily straight-forward evolution in theory should lead from the principles of utterance organisation typical of the earliest, spontaneous learner variety all the way to the target language.

4.1. INFERENCE STATISTICS

To statistically verify the tendencies identified so far, we are going to fit a generalised linear model with binomial error structure and logit link function. The dependent variable is given by the proportion of learners who perform above chance under each combination of predictors. Independent variables include test type, word order, L1 and test time; we also probe a four-way

interaction between test, word order, time and L1, as well as all implicated three-way and two-way interactions.

Table 43: Overall morphosyntactic competence, ANOVA

	Df	Deviance	Resid. Df	Resid. Dev	Pr(> Chi)
NULL	NA	NA	39	417.82	NA
test	1	101.14	38	316.68	< 0.01
WO	1	98.23	37	218.45	< 0.01
L1	4	93.13	33	125.32	< 0.01
Time	1	5.54	32	119.77	0.02
test:WO	1	49.49	31	70.29	< 0.01
test:L1	4	19.28	27	51.01	< 0.01
WO:L1	4	27.72	23	23.29	< 0.01
test:Time	1	0.31	22	22.98	0.58
WO:Time	1	7.23	21	15.75	0.01
L1:Time	4	2.69	17	13.07	0.61
test:WO:L1	4	2.29	13	10.78	0.68
test:WO:Time	1	6.23	12	4.55	0.01
test:L1:Time	4	2.26	8	2.29	0.69
WO:L1:Time	4	0.81	4	1.49	0.94
test:WO:L1:Time	4	1.49	0	0.00	0.83

All linear predictors are statistically significant, while at the same time engaging in significant two- and three-way interactions.

4.2. REPETITION IN THE ABSENCE OF COMPREHENSION

The rationale of the analysis presented so far is that learner comprehension and repetition scores combined should provide us with a comprehensive picture of the principles of utterance organisation observable in the learner variety.

However, the validity of our conclusions strongly depends on the assumptions of the repetition test, namely that it is not possible to repeat a sentence without having processed it for meaning. Phonological memory, that is, that device which allows human beings to recall meaningless strings of sounds for a short time after hearing them (Baddeley 2003), should not play any significant role in this test (Okura & Lonsdale 2012). Now, some learners in the data set appear to violate precisely this assumption. Their number is not great: scenario 4;1 comprises 5 learners at T1 and another 2 at T2. Three learners appear in scenario 2;4, indicating that the only task in which they score above chance is the repetition, but not the comprehension, of SO targets. This scenario might be a clear hint to the fact that the learners in question only used phonological memory to perform the test, while the difference between the repetition of SO and OS targets should be interpreted in terms of perceptual prominence, as argued in section 0, p. 88. However, this scenario is really too rare to draw any sensible generalisations. For each relevant combination of test time and word order, Table 44 provides the comprehension and repetition scores, as well as the probability of observing such distribution in the absence of a rational morpho-syntactic principle. Information as to the learners' performance in terms of scenarios at T1 and T2 is also provided in the last two columns. The table shows that the repetition and comprehension test scores are consistently high very high and very low, respectively, which excludes that the subject were assigned to scenario 4 only because of a rigid interpretation of border-line scores. On the contrary, the picture seems rather clear.

Table 44: learners in scenario 4

	time	WO	subj.	L1	rep. score	rep. p	comp. score	comp. p	sc. T1	sc. T2
a.	1	OS	2108	NL	0.88	< 0.01	0.00	1.00	4_1	2_3
b.	1	OS	4105	GE	0.80	0.03	0.44	0.60	4_1	4_3
c.	1	OS	4108	GE	0.83	0.02	0.31	0.89	4_1	2_1
d.	1	OS	5104	IT	0.71	0.06	0.38	0.77	4_1	1_1
e.	1	OS	5106	IT	0.75	0.04	0.00	1.00	4_1	3_1
f.	1	OS	5109	IT	0.86	0.01	0.00	1.00	4_1	1_1
g.	2	OS	2118	NL	0.88	< 0.01	0.00	1.00	2_3	4_1
h.	2	OS	4105	GE	1.00	< 0.01	0.31	0.89	4_1	4_3
i.	2	OS	4110	GE	1.00	< 0.01	0.31	0.89	2_1	4_1
l.	1	SO	4112	GE	1.00	< 0.01	0.50	0.36	2_4	2_3
m.	2	SO	1119	FR	0.75	0.04	0.12	0.96	2_2	2_4
n.	2	SO	2115	NL	0.88	< 0.01	0.14	0.94	2_3	2_4

In addition to this, others also exhibit higher scores in repetition than comprehension, although their behaviour is hardly rational in that they fail to score above chance in the comprehension of SO targets, which, as we know, can be indifferently processed based on word order or inflectional morphology. These learners (5 in total, comprising both T1 and T2) can be found in scenarios 2;4 and 4;4. A final learner is located in scenario 4;3, which surprisingly means that he or she scores above chance in the repetition test on OS, but not SO targets, and the other way round in the comprehension test. Such behaviour is evidently a little erratic and does not seem to require a specific explanation.

However, the very fact that some learners may show better performance at the repetition test than at the comprehension test should make us a little more cautious as to whether or not the assumptions of the repetition test really hold. As stated above, this question is of capital importance for our analysis, because if it were not the case, then the two tests would no longer measure the same skill, namely morphosyntax, from two different points of view. Instead, they

would measure two entirely different phenomena, specifically comprehension skills and working memory capacity.

In this section, therefore, we will try to verify whether or not it is really possible to perform the repetition test in the absence of comprehension. To do so, we administered the VILLA repetition test to new groups of Italian, French and German students selected according to the VILLA guidelines (see section 1.3.2, p. 22)⁶¹. However, these new test-takers were not exposed to any Polish input, which guarantees that they cannot process target sentences for meaning. Consequently, one can safely state that the only skills they bring to the task is their phonological memory: the way in which they perform the test, therefore, should be comparable to that of a VILLA learner who did not process targets for meaning, but only repeated them as a string of sounds. The concrete question we ask at this stage is whether these participants with no comprehension skills will be able to repeat ACC case endings, especially when they appear in the poorly prominent utterance-medial position. If so, we should conclude that the VILLA repetition test does not fulfill the assumptions of this kind of task.

A couple of examples of repetitions of the sentence /dʒɛvtʃɪnkɛ tʃɔŋgnjɛ portuɡalka/⁶² are presented in (31).

- (31) a. [tʃɛfnɛ nɛ'tswɔ nɑ portuɡalɑ]⁶³
b. [dʒɪkɪntʃɪ kɔntʃɛ portuɡartʃɛ]⁶⁴
c. [tsɪptɪrnɛ tʃɔ portuɡal ktɑ]⁶⁵

Several observations are possible. First, the first two non transparent words are hardly recognisable, whereas the transparent word in final position sounds

⁶¹ Italian participants were recruited by the authour with the help of prof. Bernini and tested at the University of Bergamo; French and German participants were recruited and tested by Marzena Wątopek at the CNRS SFL, Paris, and by Christine Dimroth and Johanna Hinz at Münster University, respectively. We sincerely thank all of them for their helpful effort.

⁶² little girl-ACC pulls portuguese woman-NOM

⁶³ German group, subject 2

⁶⁴ French group, subject 1

⁶⁵ Italian team, subject 3

decidedly closer to the target word. Even here, though, accurate repetitions only concern the part of the word which is common to the target language and the subject's L1, namely the stem. Suffixes and inflectional endings are mostly omitted or substituted with random linguistic material. At the same time, we notice that in some cases the segments corresponding to case endings are correctly repeated, in spite of being attached to a more or less random sequence of sounds. What should we conclude from this? Since we can surely exclude any sort of processing for meaning, we have to admit that the repetition of those segments can only be due to phonological memory. This too, however, is by no means a rule: working memory also seems prone to errors and inaccuracies, as witnessed by example 31.b.

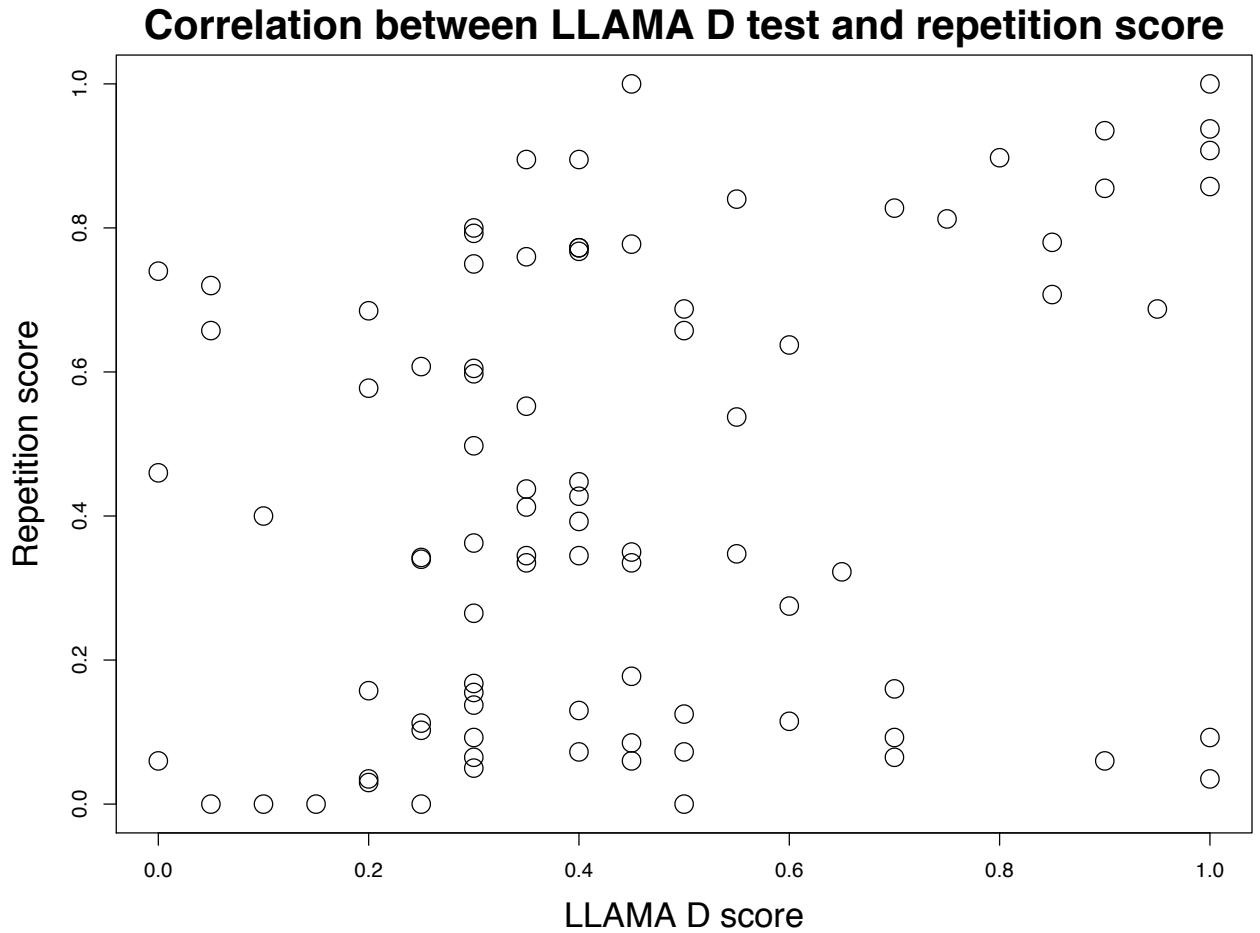
But above all, the very fact that these subjects managed to produce output which at least partly resembles target sentences witnesses to the insufficiency of the distractor. Without the possibility of processing targets for meaning, test-takers should not be able to remember anything of the sentence heard if the distractor were truly effective. Instead, they still maintain an auditory impression of it, however feeble and distorted. Based on these considerations, we should conclude that the VILLA repetition test does not guarantee that learners should necessarily engage in processing for meaning, and for this reason is not a reliable measure of implicit competence.

This does not necessarily mean that the VILLA repetition test is not effective, though. Indeed, two observations should be taken as strong evidence that in most cases learners do try to encode meaning according to the means at their disposal, be it morphosyntax or word order. The first is that the scores of most learners are consistently lower in repetition than in comprehension; the second is that the marked accusative ending is substituted by the basic word form based on the nominative case, which is consistent with what is known about early learner varieties. However, we must accept that some learners do repeat targets with very consistent accuracy and yet fail to interpret them with equal consistency, which almost inevitably leads to the conclusion that the repetition

test was performed with the aid of working memory and not through processing for meaning.

A further argument can be added. Repetition scores seem to be positively correlated with a working memory scores as measured by the LLama D test (Meara 2005), which works as follows. Learners hear a target sentence and are then presented with a shorter string of sounds: their task is to decide whether the shorter string was comprised in the target sentence. While there may be superficial similarities with the structure of the repetition test, there is a crucial difference: the sound material of the LLama test is based on the names of flowers and natural objects in a British Columbia indian language. In other words, it is highly unlikely that any VILLA learner should be able to process them for meaning, which brings the LLama test closer to the repetition test in the absence of input. Indeed, Graph 22 shows that there exists a 0.28 significant correlation ($t = 2.75$, $df = 86$, $p\text{-value} = 0.007$) between the LLama test and the score for the repetition of $-/e/$, averaged for time and word order.

Graph 22



After all, it appears that the ability to retain strings of sounds in working memory may be of help in the repetition test, at least as long as the repetition of grammatical morphemes is concerned.

Even in that case, though, we should not be too hurried in drawing our conclusions. It seems unlikely that the output of the VILLA learners comprised in scenario 4 could be the result of processing strategies comparable to those observable on first exposure. In (32) we present the repetition of the same target sentence discussed above (*/dʒɛvtʃɪnke tɔŋgnje portuɡalka/*⁶⁶) as performed by the learners comprised in Table 44, that is, learners who perform above chance in repetition, but not in comprehension.

⁶⁶ little girl-ACC pulls portuguese woman-NOM

- (32) a. dżewtʃinke tʃɔngnie portugalka.
b. dżewtʃenkə portugalska
c. dżewtʃinknɛ portu porta bazu port portugalka
d. dżjeftʃinka dʒ tʃɔɲɲe portugalka.
e. dżjeftʃinka kn eh tʃɔɲe portugalkon.
f. portugalka.
g. dżewtʃinke tʃɔngnie portugalka.
h. dżewtʃinə tʃorgo portugalka
i. dsziewtʃinkɛ tʃɔngnie portugalka
l. BLANK
m. portugalka tʃɔɲe dʒɛfkinkje.
n. portugalke tʃɔngnie dżewtʃinke.

Leaving inflectional endings aside for the moment, one immediately notices that the output produced by these learners is quite different from that of the informants in example 31. Lexical items are clearly recognisable and produced with considerable accuracy. It seems unlikely that such results could be the results of working memory alone. Instead, as anticipated informally earlier in this work, we would like to propose an explanation which lies somewhat in between the two extreme of morpho-syntactic processing and repetition based on working memory alone. We could argue that the learner variety of the learners in question is sufficiently mature to recognise lexical items, so that the corresponding lexical meaning can be processed and words⁶⁷ do not have to be stored in working memory as a bare sequence of sounds. At the same time, though, form-function associations between case endings and syntactic functions have not developed yet. The learners may well know that Polish words can appear in several word-forms: because of the implicit approach

⁶⁷ Because of the structure of the test, we have no way of distinguishing lexical words from functional ones. However, we should expect some differences in the processing of the two, as establishing form-meaning association should be considerably easier in the case of items with lexical rather than grammatical meaning.

adopted in the design of the course, however, they have not yet identified the rule which governs the use of different forms. It may be the case that these learners did not even attempt to produce a meaningful utterance, but only repeated the lexical items they recognised, in the word-form in which they identified them. Lexical items, in other words, are not *memorised*, but *recognised*, which could be considered as a sort of first step on the way to processing. Most probably, word-forms are not recognised on the basis of their function, that is, as an "accusative form" or an "object form", but of their form, as in "the form ending in *-/e/*".

This is only a working hypothesis, though, which will have to be verified by further work on the data from the VILLA repetition test and from repetition of first exposure. For now, we have demonstrated that the VILLA repetition test is by no means not flawless, but at the same time produces results which are generally in accordance with its assumptions and provide us with a fairly accurate picture of learners' morphosyntactic competence in production.

4.3. SUMMARY

This chapter aimed at providing a comprehensive picture of morphosyntactic competence, defined as the ability to derive and encode meaning through inflectional morphology only, without the help of context or phonology. This picture is deemed comprehensive because the same ability is probed using two different tasks, which in turn recreate the skills which together make up morphosyntactic competence as a whole, namely comprehension and production, or, otherwise said, the decoding and encoding of meaning. The results show that the two skills indeed differ in complexity and, consequently, accessibility by initial learners.

In order to exclude that learners could rely on strategies other than morphological or syntactic, we correlated two highly structured tests, described and analysed in detail in the previous chapters, which systematically exclude any contextual, semantic or phonological hint. The comprehension test

(Chapter 3, p. 97) requires test takers to listen to a target sentence and then select the picture which best represents it. A choice is offered between two options, differing only in the syntactic functions performed by the referents involved. The repetition test (Chapter 0, p. 3) was used to probe morphosyntactic competence in repetition. Participants are asked to listen to a target sentence, then draw a simple geometrical figure as a distractor, and finally repeat the target sentence as accurately as possible. The rationale of this procedure is that the test does not simply consist in repeating a string of sounds: in fact, this is made impossible by the distractor, which engages short-time working memory. Instead, test-takers have to decode meaning and then re-encode it, both operations being performed according to the rules of the learner variety. Therefore it can be considered as a speaking test, with the advantage of full control onto the target structures, which may be rare or absent from even long samples of spontaneous speech. Although the test may not be completely valid from an ecological point of view, it certainly represents a good compromise to study the implicit processing of marked structures in production.

In both tests, the learners' ability to decode and encode meaning using case endings was probed by manipulating word order. The rationale for this choice is that while SO structures can be successfully processed by relying on default word order (SO, agent-patient, controller-theme, topic-comment etc), OS targets can only be processed on the basis of inflectional morphology.

We first addressed the research question through cluster analysis, designed to group learners together on the basis of their linguistic behaviour. Several scenarios could be identified, providing a synoptical picture of learner processing strategies in both comprehension and repetition, and with both SO and OS target types. Four such scenarios, comprising more than 80% of the data set, have a direct, meaningful linguistic interpretation, which provides partial answers to our research questions.

- 1;1 both tests are performed above chance with both types of targets. This scenario corresponds to a full morphosyntactic principle;
- 3;1 on SO targets, both tests are performed above chance; on OS targets, only the comprehension test is. This scenario shows that if only one ability is missing, that is going to be the repetition of OS targets;
- 2;3 only the comprehension test with SO targets is performed above chance. This scenario corresponds to the typical positional principle.
- 2;1 both tests are performed above chance with SO targets, whereas with OS ones neither is. This scenario suggests that independently of the test, OS targets are harder than SO ones;

Subsequently, scenarios were broken down into their test and word order component, so that they could be described in terms of the set of skills which the learners had acquired by the test time. For the bulk of the data set, skills could be ordered along the following implicational hierarchy:

OS repetition \supset OS comprehension \supset SO repetition \supset SO comprehension

The interpretation of this scale is that if a learner is able to perform a given task with above chance accuracy, then the same must be true for all tasks to its right.

Indeed, most of the learners behave in accordance with this scale. A few do exhibit linguistic behaviour which is both incompatible with the scale and hard to interpret linguistically, so that it is best attributed to methodological faults or lack of commitment on the side of the test taker.

Cluster analysis was performed both on the results collected at T1 and T2. Although there were slight changes in the size of the clusters, the implicational scale was confirmed. The positive effect of additional exposure to the input was made evident by the growing number of learners located in scenario 1;1, indicating full morphosyntactic competence. Scenario 2;3, in contrast, indicating a pure positional principle, dramatically reduced its size.

Regularities were searched for in the evolution patterns displayed by learners, understood as the change from one scenario at T1 to another at T2. This question was addressed on the basis of both the whole data set and of the individual L1 groups.

Although only a minority (30) of the possible combinations of scenarios at T1 and T2 (256) was instantiated in the data, no persuasive tendencies could be identified. The most common pattern, in fact, involves no change at all, so that learners are found at T2 in the same scenario that they were at T1. The paths taken by learners who do change their processing strategies from T1 to T2, in contrast, are too diverse to be arranged in a defined picture. Most probably, the time interval between T1 and T2, namely 4h30, was too short to produce clear common changes. The very variability of the data, however, witnesses to the development of individual strategies of input processing.

Finally, to verify the tendencies highlighted at the descriptive level, a generalised linear model was fitted. This showed the clear significance of all the hypothesised predictors, as well as the existence of complex interactions between time, task, word order, and the learners' L1.

This concludes our analysis of the learners' morphosyntactic competence in the structured tests. Clear tendencies emerge, pointing to the greater difficulty of the repetition test, on the one hand, and of OS targets, on the other hand. Even though the majority of learners consistently apply a positional principle of utterance organisation, the most impressive result is that at least a small proportion of them seems to be able to apply a morphosyntactic principle at T1 already. Their number, moreover, greatly increases with additional, albeit limited input exposure, suggesting that even complex target structures may be acquired spontaneously with no explicit instruction and within only hours from the first contact with the target language.

5. LEARNER SEMI-SPONTANEOUS OUTPUT IN INTERACTION

5.1. INTRODUCTION

Our work so far has been focussing on data elicited experimentally in a laboratory setting. The structured tests considered in the previous chapters, while undoubtedly providing the researcher with a fully controlled environment to test linguistic hypotheses, still seriously lack ecological validity as well as realism. One could arguably imagine real-life situations resembling the comprehension task, in which meaning can only be retrieved through morphosyntactic analysis; even then, though, one could expect to perform such analysis once or twice, in contrast to the twenty-four target sentences of the test. But consider the repetition test. While it can be argued that the ability to repeat native input may be helpful in the acquisition task (to "achieve understanding", in the terms of Bremer et al. 1996), this particular task requires learners to process targets lacking any plausible meaning and uttered with a flat prosodic contour, which is usually not the case in real life. The repetitiveness it shares with many psycho-linguistic tests, as well as the lack of a true interlocutor, further make it difficult to even consider it as a communicative situation.

From this we should conclude that collecting data through structured tests, and then generalising insights to the learner's competence in general, may not be an unproblematic operation (Ellis 1985:289–290). Truly, implicit and explicit critiques to this approach are not new. Among many, Krashen (1981) proposed his well-known distinction between acquisition and learning, which may well parallel the different types of situation in which a learner may use his L2; Klein and Perdue (1997), along with all research conducted within the Learner Variety approach, based their work on the qualitative analysis of mostly free production data; and Pienemann (1998) founded the claim of his Processability

Theory on the quantitative analysis of data extracted from spontaneous speech. Within this last framework, studying the most advanced stages of language acquisition posits a methodological problem, as the corresponding marked structures, such as OS word order, may be rare in spontaneous speech, or absent altogether. Most studies have therefore employed so-called "tasks", that is, games or simulations in which participants are given the opportunity to produce semi-spontaneous speech along the lines of a more or less strict structure, which should facilitate the elicitation of the target structure of interest. To make an example, alongside free conversation, picture description and story-telling tasks, Di Biase and Kawaguchi (2002:286) used "an 'object-first' communicative task, devised to elicit structures with object clitics. This task presented first the picture of a food item (the semantic object) and then that of an animal (the semantic agent) who was meant to contribute this item to a communal animals' dinner". The test-taker had to verbalise the pictures in the order in which they occur. Ecological validity and control over the target structures to be elicited are to a certain extent mutually exclusive, so that striking a balance may require a sacrifice on either side. The "animal dinner" task described above makes it quite likely that learners should produce object-first structures, but it hardly represents spontaneous speech. In turn, it is more realistic than the repetition test described in this work, which, however, *guarantees* that the learner will be confronted with an OS structures. Results, as shown in the preceding chapters, vary all the same.

Spontaneous speech, in short, is the communicative situation which offers the greatest ecological validity, to the detriment of control over which target structures will be produced. The question now is whether or not different environments may produce not only different target structures, but different accuracy rates too.

Task effects are not a novelty in SLA research (see Michel (2011) for a comprehensive review). Harley and Swain (1984:305) for example state that "the recognition of a conditional form and of its use in a particular task is not necessarily associated with the ability to use similar forms in other tasks and

functions". More generally, Chaudron (1985:11) writes that "given the different characteristics of processing tasks [...] and the interactive nature of more complex processing, one should expect to find different "grammars" emerging with different elicitation measures".

In the same year, Tarone noted that the accuracy with which English L2 articles are produced depends on the communicative situation in which they are elicited. A little later, Tarone and Parrish (1988) observed that depending on the task, learners tend to produce a variety of noun phrases, which in turn require different types of determiners. The same type of noun phrase, moreover, is produced with varying accuracy depending on the communicative context.

Task effects are also reported by Salaberry & Lopez-Ortega (1998), who observed the accuracy with which subject personal pronouns, articles and verbal aspect are produced in the different contexts: narration, multiple-answer test and cloze test. Their analysis shows that both test type and target structure are significant predictors of accuracy.

Tarone (1979; 1985) suggested that the differences in learner output observable across different tasks stem from a number of "styles", characterised, as a result of peculiar features, by various degrees of attention to form. In addition, there may be differences in communicative pressure, that is, the difficulty of reaching a given real-life objective using language, in view of the importance of the objective itself. Styles may be placed on a continuum, whose extremes are represented by careful speech, on the one hand, and casual, "vernacular" speech, on the other hand. The latter is considered to be the most natural, basic style, in which speakers pay the least attention to form. As far as SLA studies are concerned, the author argues that the vernacular style should be accorded primacy on two grounds. First, it is more stable and consistent (Tarone 1983:154). Second, it is supposed to better reflect the actual competence of learners, as it will be the first to include structures which are typical of the learner variety, albeit grammatically incorrect. Only later on will the learner be able to adjust his output to the rules of the target language, a

tendency which is supposedly stronger and more widespread in more controlled styles.

Nonetheless, task effects are by no means obvious or easy to identify. Duff (1993), for example, studied a single learner of English L2 across three types of tasks which we could collectively label as "communicative", namely discussion/interview, picture description, and narration. Although some clear effects were found for task type, such as a greater number of words per turn in narration than in discussion, or a more widespread use of the structure "has NP" as a topic introducer in the discussion context, the author reports considerable differences from one interview to the next, which could not be explained on the basis of the parameters employed to characterise the tasks. One should conclude, then, that either different, perhaps more fine-grained parameters are needed to describe such variation, or this is beyond experimental control altogether. Perhaps the analysis of this question may benefit from Hulstijn's (1989) useful distinction between task and task requirements, where the former term broadly defines the objective and structure of the activity, while the latter term specifies such incidental variables as time pressure or informational versus grammatical focus of attention.

After this brief review, we should point out that most of the studies investigating task effects tend to compare like with like - typically, different types of broadly defined communicative tasks. In this chapter, we will attempt to correlate two radically different contexts of data elicitation: structured tests, on the one hand, and semi-spontaneous interaction, on the other hand. While these two situations share the target structure on which we will focus our attention, namely the case-marking of nouns performing the object function, they radically differ in their structure and in the kind of output required of the learner. This difference will allow us to pursue two different research questions. First, our analysis may usefully contribute to the debate on task effects by exploiting the fully controlled methodology granted by the VILLA project, as well as the fact that identical target structures are being compared. Second,

and more importantly for our purposes, we will be able to verify if the extent to which learners are able to process morphosyntax varies depending on the task used to elicit output, implying varying degrees of communicative pressure in the sense of Tarone and Parrish' (1988).

The choice of the two contexts is not random: interaction has often been seen as a privileged environment not just for data elicitation, but for acquisition too. With his Interaction Hypothesis, Long (1990) as well as others (Gass and Mackey 2006, Gass et al. 1998, Mackey 1999, Pica 1994) argue that this is the context in which learners's L2 competence finds a meaningful application, as they need to rely on their interlanguage to understand their partners and make themselves understood. In addition, feedback from the interactional partners may help L2 speakers to notice any differences between their output and the target language, thus providing precious occasions for focus on form and hypothesis testing. In turn, modifying one's own output upon negative feedback is thought to be a particularly precious exercise to enhance intake (Schmidt 1990; Pica 1994). In this respect, Swain and Lapkin (1995) further argue with their Output hypothesis that providing learners with the opportunity to produce output may drive them towards a syntactic mode of processing (Givón 1985), as opposed to the pragmatic or semantic mode.

Other models foresee a different scenario, though. Skehan and Foster, in their so-called Limited Attentional Capacity Model (Skehan 1996, 2009, Skehan and Foster 2001), suggest that humans are equipped with only limited attentional capacity, so that, when producing output in an L2, the different ongoing processes compete with each other for attentional resources. It follows that only some of them can receive attention and produce accurate output. The model predicts that when competition arises, learners will concentrate on their communicative goal, focussing on meaning to the detriment of form (VanPatten 1990). In this case, we should expect lower overall morphosyntactic accuracy, as learners will be focussing on meaning, here represented by lexical access and integration within the interactional flow. Only those who have automatised the relevant bits of the target

morphosyntactic system will be able to correctly inflect nouns depending on their syntactic function. All others will focus on more urgent levels of language, and rely on other means to express grammatical meaning. In line with the findings of research within the Learner Variety approach (Klein & Perdue 1997), we may foresee semantics (degree of control) and pragmatics (information structure) as two candidates for that role.

5.2. METHODOLOGY

Among its many resources, the VILLA project offers us the possibility to look into this topic too. In addition to the many structured tests and free production tasks devoted to specific target structures, learner output was also observed during classes as part of routine classroom interaction. Participants sat in front of a directional microphone which digitally recorded everything they said during the whole course, as illustrated in Figure 5.

Figure 5: VILLA classroom set-up



Thanks to this methodology, it is possible to study learner output when prompted by the teacher, which, among other things, should show us the very first items produced by a learner straight after the first exposure to a given target.

More importantly for our purposes, though, the VILLA course comprises several dialogic episodes during which participants could interact with each other. They were typically given a simple task to perform in pairs, usually involving a particular grammatical structure or vocabulary set which had been previously practised with the teacher.

The data discussed in this chapter⁶⁸ were collected precisely during one such occasion, which took place during lesson 7.2, i.e. after roughly 10 hours of exposure to the input. In order to contextualise these figures, it is helpful to remind the reader that T1 and T2 in the structured tests correspond to lessons 6.2 and 9.2, respectively, which in turn indicate 9 hours and 13:30 hours. Accordingly, the competence portrayed in the output discussed below should not be too dissimilar from what we observed when discussing the structured tests at T1, provided that we assume a gradual, continuous model of language acquisition⁶⁹.

Without having to move from their seats, participants were divided into 7 groups of 2 and a group of 3. They were then given a set of paired cards, which contained information about several referents that learners were asked to describe to each other. Each card in a learner's set only contained part of this information: the remaining details could be found on the corresponding card in the partner's set, so that, in order to obtain a full description of the referent, information had to be exchanged between the two.

While the first participant described the referent based on the information on his or her card, the partner would try to identify the relevant character, asking questions to complete the information set. In doing so, learners were encouraged to use all structures presented during the course, including the transitive constructions which constitute the object of this work, as in [ɔn ma duze zute samoxut], "he has a big, yellow car". An example of the dialogue between learners 5101 and 5202 is provided in Table 45.

⁶⁸ In this chapter we will focus on the Italian Meaning-based input group, which attended the VILLA course at the university of Pavia in September 2013.

⁶⁹ See Rastelli (2014) for an alternative view on this topic.

Table 45: interaction between 5101 and 5102

5102	[ɔn ma na im ɔn ma na imje pjer].	"His name is Pierre."
	[ɔn naziva je pjer zo'la].	"His name is Pierre Zola."
	[ɔn ɔn jest strazakjem].	"He is a fireman."
	[i ɔn pratsuje v straze pozarnje].	"and he works at the fire department."
	[ɔn jest frantsuzem].	"he is a Frenchman."
	[mje]ka vɔ frantsi v parizu].	"(he) lives in Paris."
	[ɔn ma duze zute samoxut].	"he has a big car."
	[i ɔn ma zɔne eve].	"And he has a wife, Ewa."
	[ɔn ma dzjet]i anna i karlo].	"He has children, Anna and Carlo."
	[i ɔn lubi matematike literature].	"And he likes math and literature."
	[i lu lubi jexat] na roveze].	"And he likes riding (his) bike."
	[i jest dinamit] jest dinamit]ne].	"And he is active."
5101:	[dze pratsuje pjer]?	"where does Pierre work?"
5102:	[pjer pratsuje v straze pozarne].	"Pierre works at the fire department."
5101:	[i kto kocha pjer]?	"and who does Pierre love?"
5202:	[pjer xoxa sin brata i]ɔstra].	"Pierre loves (his) son, (his) brother and (his) sister."

For this study, the audio fragments relative to the episode considered were extracted from the main track and merged according to the small groups in which learners were divided. Following this operation, each resulting track contained only the speech of the two or three participants who were part of the same group. Following Saturno (2015), these tracks were finally transcribed in broad IPA using ELAN (Brugman & Russell 2004) and CHAT (MacWhinney 2000). Synchronised video recordings proved of great help to identify participants during the transcription stage, providing the transcriber with an additional clue, in addition to the sound of learners' voices. For analytical

purposes and ease of reading, transcripts were further divided into one-verb utterances. In this chapter we only consider utterances in which at least one feminine noun occurs, so as to ensure the full compatibility of the analysis with the results of the structured tests. Altogether, our corpus comprises a total of 60 utterances, containing one or more feminine nouns.

5.3. AN OVERVIEW OF LEARNER OUTPUT

We will start our analysis by presenting a brief survey of the types of utterance which we may encounter in the data, considered from the point of view of correct or incorrect case marking of the subject and object functions.

The subject is commonly expressed by a person name, a pattern found in 15 utterances. New characters are typically introduced by a copular construction with presentative function, in which the topic is expressed by a personal pronoun (*on*, "he" or *ona*, "she") and the complement is represented by the name (33).

- (33) [ɔna jɛst dʒovann-a].
she is Giovanna-NOM
"she is Giovanna"
[dʒovann-a ɛst nautʃitʃɛlk-on].
Giovanna-NOM is teacher-INS
"Giovanna is a teacher"

When a referent has been thus introduced, it is typically referred to using personal pronouns (34), although names may be repeated in consecutive utterances, too (35):

- (34) a. [to jest marj-a].
 this is Maria-NOM
 "this is Maria"
- b. [ɔna lubi herbat-a].
 she likes tea-NOM
 "she likes tea"
- (35) a. [to jest marj-a].
 this is Maria-NOM
 "this is Maria"
- b. [marj-a jest njemk-on i tumat{k-on}].
 Maria-NOM is German-INS and interpreter-INS
 "Maria is German and an interpreter"

The last type of utterances found in the input presents zero anaphora (36b), following a pause (indicated as <(...)> in the transcript) and another utterances in which the subject is expressed by either a name or a pronoun (36b).

- (36) a. [ɔna jest portugalk-on (...)].
 she.NOM is Portuguese-INS
 "she is Portuguese"
- b. [i ɔna zna jeski angjelsk-i].
 and she knows language.NOM/ACC(?) English-NOM/ACC
 "and she speaks English"

In no examples can we find common nouns expressing the subject function, as that would be pragmatically inappropriate. Since names and pronouns lie beyond the scope of our work, we will not pursue the analysis of the nominative further. Additionally, the nominative case by hypothesis represents the unmarked entry in the developing nominal paradigm, and perhaps the only

one, in case nouns in the interlanguage grammar are still limited to an invariable word-form. For all these reasons, the rest of this chapter will only consider the case-marking of nouns in the object function, to which we now turn.

In the first example all nouns in the object function are correctly marked as accusative, both in a sequence of feminine nouns only (37 a) and in sequences of feminine and masculine nouns alike (37 b):

- (37) a. [ɔn lubi tʃokolad-e kav-e i erbat-e].⁷⁰
 He likes chocolate-ACC coffee-ACC and tea-ACC
 "he likes chocolate, coffee and tea"
- b. [kɔxa ʒɔn-e ev-e kɔt-a i ps-a].⁷¹
 loves wife-ACC Ewa-ACC cat-ACC and dog-ACC
 "(he) loves his wife Ewa and (his) cat and (his) dog"

At the other end of the spectrum we find occurrences in which all feminine nouns in the object function are marked as nominative (38):

- (38) a. [ɔna lubi herbat-a].⁷²
 she likes tea-NOM
 "she likes tea"
- b. [ɔna lubi tʃekolad-a kav-a i erbat-a].⁷³
 she likes chocolate-NOM coffee-NOM and tea-NOM
 "she likes chocolate coffee and tea"

In other cases still, nouns belonging to the same paradigm and with the same function of object are marked as either nominative or accusative, with no apparent regularity (39):

⁷⁰ 5106

⁷¹ 5102

⁷² 5111

⁷³ 5119

(39) *ona lubi erbat-e i kav-a.*⁷⁴
 she likes tea-ACC and coffee-NOM

Finally, we add that all utterances produced by learners are structured according to the SO word order. We have no means to tell whether this reflects their imperfect competence, or simply the discourse situation: OS structures typically serve the purpose of topicalising the object in order to create contrastive emphasis, and it can be argued that in the task discussed here there is no such need.

5.3.1. *Lexemes*

We now turn our attention to the lexemes produced by the learners during the interactional episode. For each word, Table 46 presents the following information: the first column ("lexeme") indicates lexemes along with their English translation (second column, "translation"). The third column ("mean") indicates the overall accuracy with which the word received accusative marking, computed as the ratio between the number of correct occurrences of a lexeme and the total number of occurrences as produced by all learners combined (fourth column, "freq"). The fifth column ("subjects") specifies the number of learners who produced the lexeme. The last four columns provide the frequency with which the word occurred in the input at the time of the test, i.e. after 10:30 hours. The table provides figures for three word-forms, namely the nominative case ("NOM"), the accusative case ("ACC"), and all other cases combined ("other"), along with their sum ("tot").

⁷⁴ 5107

Table 46: lexemes produced by learners in interaction

lexeme	translation	Mean	Freq.	subjects	NOM	ACC	other	Tot
<i>Anna</i>	Anna	0.00	1	1	243	0	2	245
<i>Chorwacja</i>	croatia	1.00	2	1	1	0	2	3
<i>Córka</i>	daughter	0.00	6	5	27	4	0	31
<i>Czekolada</i>	chocolate	0.30	10	6	13	58	0	71
<i>Ewa</i>	Ewa	1.00	2	1	123	4	34	161
<i>Herbata</i>	tea	0.78	9	9	16	52	0	68
<i>Kawa</i>	coffee	0.30	10	6	36	48	0	84
<i>Kola</i>	coke	0.50	2	2	8	30	0	38
<i>Kuchnia</i>	cuisine/kitchen	0.75	4	3	20	22	0	42
<i>Lalka</i>	doll	0.00	1	1	22	7	9	48
<i>Literatura</i>	literature	0.71	17	11	1	53	0	54
<i>Mama</i>	mum	0.50	2	1	54	8	0	62
<i>Matematyka</i>	math	0.57	7	5	0	9	0	9
<i>Muzyka</i>	music	0.00	1	1	1	19	0	20
<i>Piłka</i>	ball	0.33	3	3	30	23	0	53
<i>Pizza</i>	pizza	0.00	1	1	12	31	0	43
<i>Rodzina</i>	family	1.00	1	1	32	0	12	44
<i>Siostra</i>	daughter	0.00	3	2	19	13	3	45
<i>Tata</i>	dad	1.00	1	1	54	19	0	73
<i>Żaba</i>	frog	0.50	2	2	61	48	0	109
<i>Żona</i>	wife	1.00	2	1	44	4	0	48

We will limit the analysis to common nouns, thus excluding three lexemes: the person names *Anna* and *Ewa*, and the country name *Chorwacja*, "Croatia". The rationale for their exclusion lies in their semantic specificity, which may cause them to be processed otherwise than common nouns: specifically, learners might think that being names, they should not follow the inflectional pattern of other nouns, despite their belonging to the same inflectional paradigm. Regarding their frequency we observe that *Chorwacja*, with only 1 occurrence,

is an extremely rare input item; in spite of the fact that one of the main characters of the course, *dziadek Dawid*, "grandpa Dawid", is actually from Croatia, and that nationalities are an important component of the input, the preferred construction to express that meaning is *Dawid jest chorwatem*, "Dawid is a Croatian". The other two names, on the other hand, are extremely common, but mainly appear in the nominative case. This tendency is particularly extreme in the case of Anna, whereas Ewa also exhibits a fair number of other word forms. Virtually all of them belong to the genitive case with the meaning of possession, as in (40).

(40) *to jest pokój Ew-y*
 this is room.NOM Ewa-GEN

Let us then proceed with our analysis. Our first observation inevitably regards the fact that, given the nature of the task, not many occurrences were elicited for each lexeme: moreover, as we are only considering nouns belonging to the paradigm of feminine nouns in *-/a/*, our dataset is further limited. The most common item (*literatura*, "literature") occurred 17 times; 6 words (e.g. *rodzina*, "family") only occurred once.

The number of learners producing each lexeme is rather variable, ranging from a minimum of 1 (e.g. *żona*, "wife"); to a maximum of 11 for *literatura*, "literature", an item which justly deserves this record on several grounds: in addition to being fairly common in the input, it almost always appears in the accusative case, which may have facilitated a privileged association with the object function. Finally, it is a very transparent word and a classic of dialogues on likes and dislikes.

A certain degree of variability can also be found in the overall frequency in the input of the items considered here, the two extremes of the continuum being *matematyka*, "mathematics" with 9 occurrences and *żaba*, "frog", with 109.

We finally turn to the mean accuracy ratio. We start with a brief overview of the non-target-like endings found on nouns in the object function. Out of a

total of 47 examples, 6 end in a bare consonant, and another one in [ən] (Table 47).

Table 47: interaction, endings other than -/a/ or -/e/

learner	utterance	learner form	target
5101	[i ma dzurk].	[dzurk]	/tsurke/
5118	[paolo lubi matema'tik i mu'zik].	[matema'tik]	/mate'matike/
5118	[paolo lubi matema'tik i mu'zik].	[mu'zik]	/'muzike/
5115	[on lubi literatura i matematik].	[matematik]	/mate'matike/
5117	[ɔn lubi matematik i psa].	[matematik]	/mate'matike/
5117	[marta lubi kav i kerbatən].	[kav]	/'kave/
5117	[marta lubi kav i kerbatən].	[kerbatən]	/her'bate/

In most cases, one could hypothesise the influence of other known languages, especially German, as in [matema'tik], compare to ger. *Mathematik* /matema'ti:k/. In other cases, the ending may be modelled on other word forms present in the input, whether extracted from the paradigm of the paradigm of the lexical item they are attached to or not. In the case of [kerbatən], for instance, the ending [ən] may be a trace of the instrumental masculine in *-em*. Moreover, these few examples were produced by only four learners, who often produced more than one instance. It seems, therefore, that this phenomenon should be a matter of individual variability whose causes are beyond our control even under the privileged conditions of the VILLA project. All remaining examples of incorrectly marked nouns are modelled on the nominative case in *-a* (Table 3) confirming that we are dealing with a mostly binary alternative.

Table 48: interaction, non-target-like ending

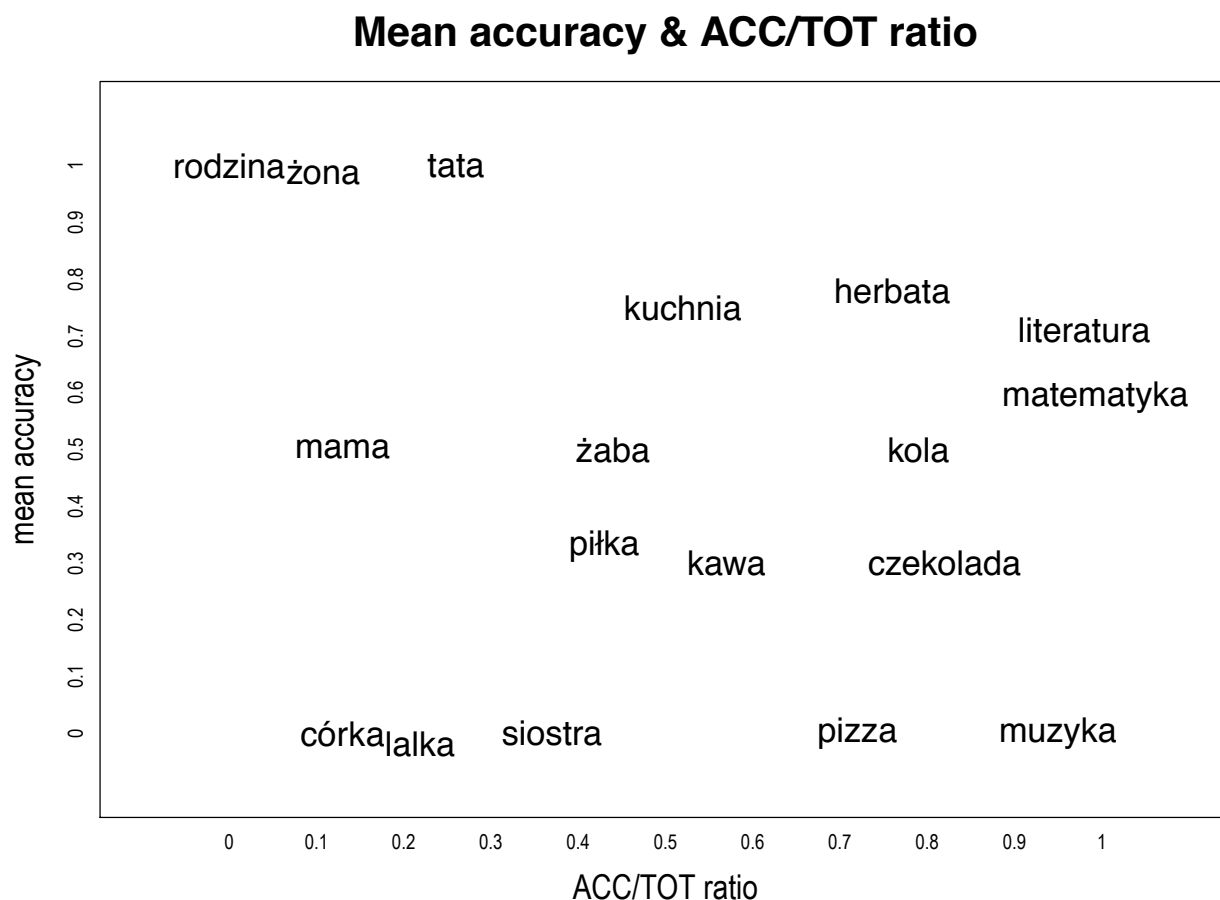
learner	utterance	learner form	target
5117	[i on ma rodzina].	[rodzina]	/ro'dzine/

Turning back to our quantitative analysis, while some lexemes reach 100%, others scored 0% altogether: in fact, the whole continuum is represented. As example (39) above makes it clear, different nouns may receive either marking even within the same utterance.

These observations lead us to formulate the hypothesis that such distribution may result from a biased distribution in the input: if a given lexeme mostly occurs in a specific word-form, then learners may associate it with the corresponding syntactic function or, at least, with the corresponding case marking. To make an example, it may be the case that a word which only occurs in the accusative case, like *matematyka*, "mathematics", may be noted and stored by learners in its accusative form. If this is so, we should expect two phenomena to happen. First, accuracy for accusative case marking should be very high, in principle 100%. Second, and conversely, this word-form should overextend to all others, including the nominative case. While the first prediction should hold in any case, provided of course that form-function associations are established on the basis of the distribution of word-forms in the input, the second makes stronger assumptions and requires greater theoretical commitment. A scenario in which both predictions are realised simply indicates that the learner elected the accusative case as the basic word-form of the basic variety. For this to happen, in the context of an input in which nominal paradigms are dominated by the nominative case, at least in terms of frequency, there must be no room for any sort of abstract generalisations, or rules. It follows that such predictions can only stem from a rather extreme form of usage-based approach, which we will not endorse in this work. Nonetheless, we will verify whether or not the accuracy of accusative marking is influenced by the proportion of instances in which a given lexeme appears in that word-form. We already know that this is not necessarily the case for all learners, as two of them never mark any noun correctly in their output, while others reach accuracy levels close to 100%. It may be that we witness different levels of development of the learner variety; but we cannot exclude, because of the lack of evidence regarding subject

marking, that different learners simply selected different word forms as the basic form of a lexical item, as indeed is the case in spontaneous acquisition (Broeder, Extra & Van Hout 1989; 1993). For those learners who mark nouns correctly only from time to time, a provisional hypothesis may be that apparently random discrepancies might in fact depend on the lexeme, more specifically on the relative frequency of its accusative form, computed as the number of accusative word-forms on the total number of occurrences for that particular lexical item. In short, the rationale is that learners should be facilitated to produce a specific lexical item in the accusative word-form if this occurs more often than the other forms of the paradigm. If we plot the relative frequency of accusative forms against case-marking accuracy, then, a straight line with a positive slope would indicate that the hypothesis is confirmed. Graph 23, however, shows no apparent pattern, indicating that the expectations are not borne out in the data. Specifically, a few words which in the input are rare or absent altogether show a mean accuracy of 100%, while other, whose ACC word form is much more common, exhibit much lower accuracy scores.

Graph 23⁷⁵



We now turn to consider overall frequency, i.e. the frequency of all word-forms lumped together. The rationale is now that if a word is very frequent in the input, then it should be more available to the learners, and therefore more easily retrievable. If the learner is quicker and more comfortable in retrieving the lexeme, then perhaps he could devote more resources to less crucial levels of language, such as inflectional morphology. A word of caution is needed here: as not all learners produced all lexemes (Table 49), and even the most frequent lexeme in learner output, *literatura* "literature", only occurred 17 times across 11 participants, statistical tools are not particularly reliable here.

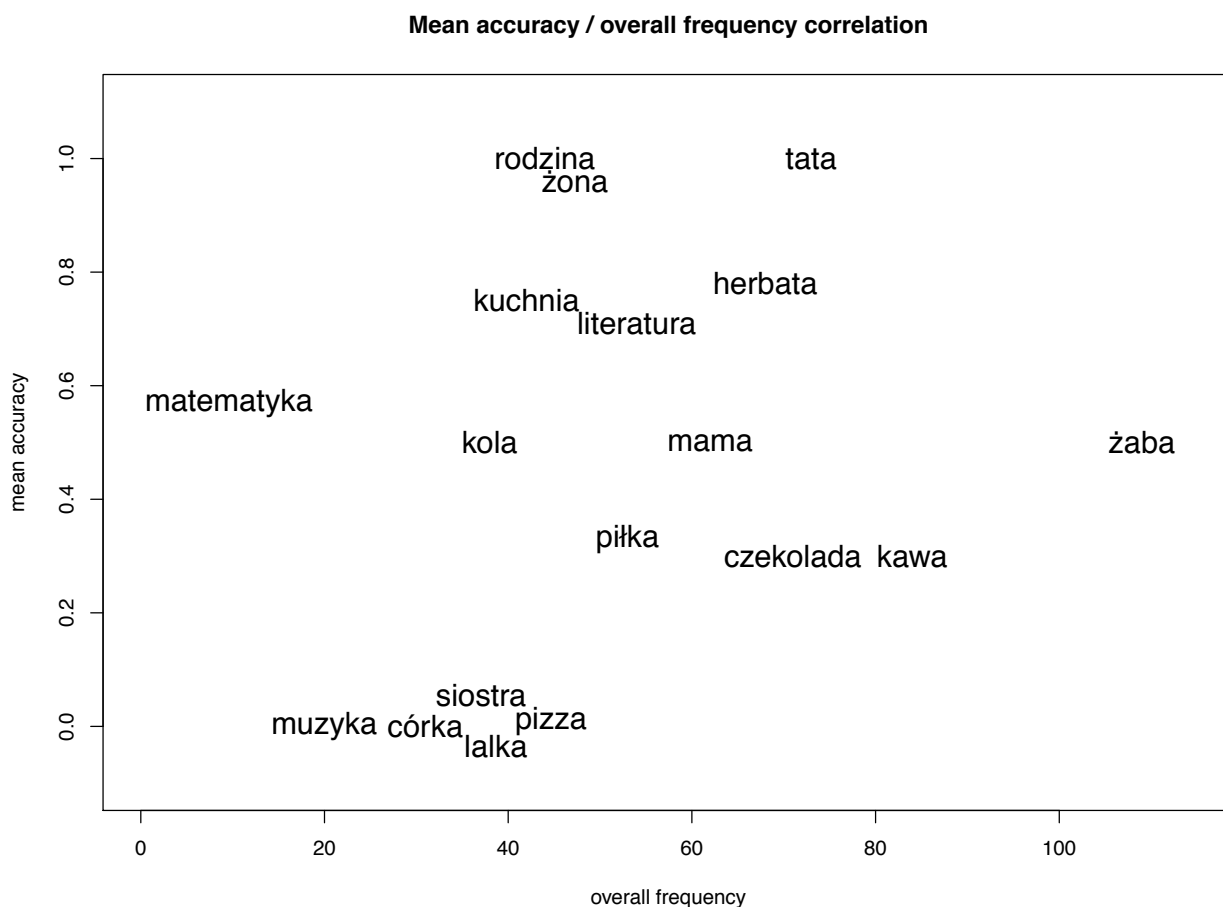
⁷⁵ This graph was created using the statistical software R (R Core team 2014) and the packages *wordcloud* (Fellows 2014) and *extrafont* (Chang 2014).

Table 49: lexemes produced by learners

subject	lexemes
5101	<i>czekolada, kuchnia, literatura, siostra</i>
5102	<i>córka, Ewa, literatura, matematyka, piłka, siostra, żaba, żona</i>
5104	<i>córka, lalka, literatura</i>
5105	<i>Herbata</i>
5106	<i>czekolada, herbata, kawa, kola, literatura</i>
5107	<i>córka, herbata, kawa</i>
5109	<i>chorwacja, czekolada, herbata, kawa, kuchnia, literatura, matematyka</i>
5111	<i>herbata, literatura</i>
5112	<i>kuchnia, literatura, pizza</i>
5113	<i>czekolada, herbata, kawa, literatura, żaba</i>
5114	<i>Anna, czekolada, herbata, kola, piłka</i>
5115	<i>córka, literatura, mama, matematyka, tata</i>
5116	<i>literatura, rodzina</i>
5117	<i>herbata, kawa, literatura, matematyka</i>
5118	<i>córka, matematyka, muzyka</i>
5119	<i>czekolada, herbata, kawa, piłka</i>

When these limitations are acknowledged, the rationale of the analysis becomes as follows. We start from the assumption that the effects of frequency might be the same for all learners, and that regularities found in the output should be generalisable to the potential output of other learners. If there were a correlation between overall frequency and accuracy, lexemes in Graph 24 should distribute along a positive slope, with words being processed more and more accurately as their frequency in the input increases. Quite clearly, this is not the case.

Graph 24⁷⁶



The information just shown should reassure us as to the fact that differences in learner processing of morphosyntax are not due to the biased distribution of word-forms in the input.

5.4. SAME-WORD UTTERANCES

Among the implications of the findings just discussed is that the most common word-form of a lexical item is not necessarily selected as its basic word-form. The analysis was carried out applying statistical tools on the basis of the whole learner set. We will now try to establish if the same holds for individual

⁷⁶ This graph was created using the statistical software R (R Core team 2014) and the packages *wordcloud* (Fellows 2014) and *extrafont* (Chang 2014).

participants by investigating the number of word-forms in which each learner produces specific lexemes.

This question may be pursued by looking at the output of learners with a mean accuracy different from 0 or 1, and in which the same lexical item occurs more than once. Table 50 presents the relevant utterances. The rationale is as follows: if the rule governing case marking is simply unstable, then repeated lexical items should appear sometimes in their nominative, sometimes in their accusative form. If, on the other hand, lexical items only occur in their invariable, basic word-form, whatever its model in the input, then each word should always appear in the same word-form.

Table 50: same-word utterances

subject	utterance	Lexeme
5102	[i ɔna lubi literature].	[literature]
5102	[i ɔn lubi matematike literature].	[literature; matematike]
5102	[i lubi matematike].	[matematike]
5102	[i kɔxa ʒɔne eve i kɔta i kɔta i psa].	[ʒɔne; eve]
5102	[i ɔn ma ʒɔne eve].	[ʒɔne; eve]
5104	[ɔn ma kɔta i sɔn i tsurka].	[tsurka]
5104	[ɔn ma sɔn i tsurka i fortɔpjan].	[tsurka]
5109	[ɔna lubi kava ɚbate leteratura].	[kava; ɚbate; leteratura]
5109	[ɔn lubi literatura i matematike].	[literatura; matematike]
5109	[ɔn lu ɔn lubi literature matematike].	[literatura; matematike]
5109	[ɔna lubi kava i korvate tʃekolada].	[kava; korvate; tʃekolada]
5109	[ɔn lubi lɔde tʃekolada i korvatʃe].	[tʃekolada; korvatʃe]
5113	[ɔna lubi literatura kave i i kɔt].	[literatura; kave]
5113	[i lubi literatura i kino i kerbate i kava i psa i kot].	[literatura; kerbate; kava]
5115	[kristina xoxa mama mame].	[mama mame]

We start with evidence of variability. The only relevant utterance by 5115, presented in (41), suggests that rule-based, explicit processes may be at play in producing lexical items in a specific word-form.

- (41) [kristin-a xoxa mam-a mam-e].
Cristina-NOM loves mum-NOM mum-ACC
"Cristina loves (her) mum"

The learner tries out different forms of the same word, both desumed from the input: *mam-a* "mum-NOM" occurs 54 times, *mam-e* "mum-ACC" 8 times. This example witnesses to instability in the system, which depending on the further input received might lead to evolution towards target-like output, fossilisation, or the development of a new system.

Two other learners also exhibit variable case-marking with the same lexical items. 5109 produces three instances of *literatur-a*, "literature-NOM", and one of *literatur-e*, "literature-ACC"; 5113 produces one instance of *kaw-a*, "coffee-NOM", and one of *kaw-e*, "coffee-ACC".

With only these three exceptions, all other lexical items always occur in the same word-form, which can be indifferently *-e/ ACC* (in the speech of learner 5102) or, more commonly, *-a/ NOM*. Based on these very limited data set, there is little we can add to what we already knew, and, more importantly, did not know. We have shown that the same lexical item tends to occur in the same inflected word-form: when that is the nominative case, we could simply think of the basic word-form of the learner variety. When the learner consistently produces the *-e/ ACC* form, we still cannot tell if he is applying a morphosyntactic rule, or if that lexical item, for that learner, simply presents a basic word-form in *-e/*. The reasons why a specific word-form should be selected are probably complex and hard, if not impossible, to control. Frequency would appear as an intuitive factor in this choice, but the analysis presented above showed no obvious correlation. More personal factors should

then be taken into consideration, such as attention, previous experience, and individual mental associations.

A methodologically sound way to solve this question would be to collect several examples in which the same word performs different grammatical functions (Pallotti 2006). If word-forms vary on the basis of grammatical function, then the learner has developed a morphosyntactic principle of utterance organisation: if not, that word simply occurs in a word-form which is not modelled on the nominative case, but on another form of the paradigm. Regrettably, the VILLA data do not contain enough information to solve this question. For the time being, then, the most reliable hint as to the developmental stage of a learner's competence remains the statistical set described in section 5.5, p. 183.

5.5. EVALUATING A SYNTACTIC PRINCIPLE OF UTTERANCE ORGANISATION

We finally turn to verifying whether or not learners use inflectional morphology in a target-like and systematic manner, that is, following a morphosyntactic principle. The relevant information is presented in Table 51, which for each participant⁷⁷ provides the mean score, the number of utterances, the number of words produced, and the ratio between the latter two values, in which a value of 1 indicates that each lexeme occurs once in each utterances, while higher values indicate that at least one occurs more than once. The last three values are useful in order to obtain a more complete picture of the learner variety: specifically, while very high mean scores might suggest that the learner has productively acquired the rule of case marking, a reduced number of utterances might shed some doubt on this claim. This is the case of subject 5105, for example, about whom very little can be said from the only relevant utterance he produced. Further still, a reduced number of lexemes might suggest that the learner is not applying a rule, but only

⁷⁷ Only 16 participants out of 17 are included in the table as one of them (5120) did not produce any utterance comprising feminine nouns.

replicating chunks deduced from the input and not necessarily analysed in terms of morphosyntax. The last column provides the probability that learners did *not* consistently use case-marking to express syntactic functions⁷⁸.

Table 51: interaction, morphosyntactic principle probability by learner

subject	mean	correct	contexts	lexemes	utterances/lexemes	p
5101	0.67	4	6	4	1.50	0.11
5102	0.83	10	12	8	1.50	< 0.01
5104	0.25	1	4	3	1.33	0.69
5105	1.00	1	1	1	1.00	< 0.01
5106	1.00	8	8	5	1.60	< 0.01
5107	0.33	1	3	3	1.00	0.50
5109	0.46	6	13	7	1.86	0.50
5111	0.50	1	2	2	1.00	0.25
5112	0.67	2	3	3	1.00	0.13
5113	0.43	3	7	5	1.40	0.50
5114	0.20	1	5	5	1.00	0.81
5115	0.33	2	6	5	1.20	0.66
5116	1.00	2	2	2	1.00	< 0.01
5117	0.50	2	4	4	1.00	0.31
5118	0.00	0	3	3	1.00	0.88
5119	0.00	0	8	4	2.00	1.00

A few learners have a p value close to 0, which should witness to their systematic use of the target morphosyntactic principle. On closer inspection, though, we note that learners 5105 and 5116 only produced one and two relevant utterances, respectively, which makes it difficult to generalise the results. As for the rest of this group, learner 5106 made no errors at all;

⁷⁸ For each learner, the p-value was calculated based on the upper tail of a binomial distribution described by the number of correct responses produced and the total number of responses. Chance is modelled as a set probability of 50%. The same methodologies is employed and discussed in more detail in the chapters devoted to the structured tests.

regarding 5102, the passage in which her two errors occur is reported in (42):

- (42) *STU: [ɔna ɔn ma brat-a].
 she he has brother-ACC
 "she has (a) brother"
 *STU: [ma siostr-a].
 has sister-NOM
 "(she) has (a) sister"
 *STU: [i ma dzurk].
 and has daughter-?
 "and (she) has a daughter"

On a purely speculative level, one might hypothesise that the non-target-like marking of the feminine noun may bear the influence of the preceding masculine noun, correctly inflected. Since *brat*, "brother" is a masculine animate noun, its accusative ends in *-a/*, just like the nominative case of feminine nouns. Regarding the ending in a bare consonant of the word [dzurk], "daughter", we can only suppose that the learner picked up this rather infrequent lexical item in that specific form for some reason beyond our control⁷⁹. In any case, the otherwise high accuracy rate of this learner makes it more probable that the examples presented above may be due to simple mistakes, rather than to an imperfect competence.

At the other end of the spectrum, two learners (5118 and 5119) did not produce any accusative marking in their three and eight respective obligatory contexts. For these learners, all nouns probably occur in a single invariable word-form in *-a/* and morphological variation does not take place.

All other cases present a somewhat mixed picture. Thresholds are intrinsically arbitrary, which is why we present the approximate *p* value instead of presenting the reader with a classification, as was done, on practical grounds,

⁷⁹ In all honesty, we cannot exclude that this awkward example could be due to the poor quality of the recording.

in the chapters devoted to the structured tests. For reasons of consistency with the previous analysis, though, we could take 10% as a working threshold. P values below this figure indicate that we have a 90% probability of being right in saying that the arguably good results observed were achieved not because of luck, but thanks to the systematic application of a syntactic principle. Conversely, we suffer a reasonable 10% risk that the learner in question has been particularly lucky in his guessing.

Regarding the participants whose p values are above 10%, it can be excluded that they systematically mark all feminine nouns with the object function as accusative, as required by the target language. Nonetheless, they sometimes do, which witnesses to the fact that they managed to notice that there is some morphological variation in the input, first, and to identify what word forms lexical items can assume, second. What is still missing is the ability to use the correct word-form in the appropriate syntactic context, that is, a form-function association between syntactic function and word-form.

5.6. PRINCIPLES OF UTTERANCE FORMATION AND INTERPRETATION

Our analysis has just shown that only a minority of subjects consistently uses morphology to express meaning in their spontaneous output, yet all managed to successfully complete their task. On what principles did they rely then to express and decode meaning? To answer this question, we will perform a qualitative analysis of learner output.

Most often, the referents involved in the utterance differ in their animacy, so that the animate referent has by far the greatest probabilities of being the subject of the sentence, or, on a more general, semantic level, the "controller" of the event (43a). When referents do not differ in their animacy, default SO word order is still available to derive meaning (43b). In fact, the entire corpus of learner output does not contain a single OS utterance, although it could be argued that such structures were simply not required pragmatically.

- (43) a. [dʒulj-a lubi herbat-e kol-a i tʃokolad-a]⁸⁰
 Giulia-NOM likes tea-ACC coke -NOM and chocolate-
 NOM
 "Giulia likes tea, coke and chocolate"
- b. [luk-a xoxa hann-a].⁸¹⁸²
 Luca-NOM loves Anna-NOM
 "Luca loves Anna"

Both animacy contrast (44a) and default word order (44b) are relied upon even by those learners who were shown to use inflectional morphology productively.

- (44) a. [ɔna lubi literatur-e].⁸³
 she likes literature-ACC
 "she likes literature"
- b. [i ɔn ma ʒɔn-e ev-e]⁸⁴
 and he has wife-ACC Ewa-ACC
 "and he has a wife, Ewa"

Quite strikingly, the same principles appear to operate in native speech as well, as witnessed by the quantitative analysis of teacher input. The vast majority of transitive utterances (n = 136) involves both a contrast in animacy and default SO word order (45a); if the latter is missing (45b), animacy still ensures that meaning can be easily decoded (n = 35). In case more than one referent is animate (45c), decoding relies on SO word order (n = 28). Only in a minority of utterances is morpho-syntactic analysis indispensable to decode meaning

⁸⁰ 5114

⁸¹ 5114

⁸² Note that Hanna, being a person name, is not included in our computations of morphosyntactic accuracy. See section **Errore. L'origine riferimento non è stata trovata.**, p. 8.

⁸³ 5102

⁸⁴ 5102

correctly, as both referents share the same value of animacy in the presence of marked word order (45d).

- (45) a. *Filip pcha wózek*
Filip.NOM pushes cart.ACC
"Filip pushes the cart"
- b. *muzyk-ę lubi Leon*
music-ACC likes Leon.NOM
"Leon likes music"
- c. *mąż kocha żon-ę*
husband.NOM loves wife-ACC
"the husband loves (his) wife"
- d. *brat-a kocha Juli-a*
brother-ACC loves Julia-NOM
"Julia loves (her) brother"

In sum, it appears that even if a learner simplifies his output by omitting case morphology, this does not impede the expression of meaning, since such semantic and syntactic means as animacy contrast and word order are sufficient to achieve the learners' communicative goals.

This system may appear efficient at the present stage of interlanguage development; however, we can foresee at least two sources of pressure towards the target language. The first is functional: as clearly shown by Klein & Perdue (1997), phrasal, semantic and pragmatic principles will conflict, sooner or later, and the learner will find the system inadequate to express more complex meaning. As a result, new means will be developed based on the input received by native speakers, which presumably should bring the learner variety closer to the target language.

In such a context as the VILLA course, moreover, it is implicitly understood that the learners' task is to try and learn the target language, including its grammar. Combined with the modified VILLA input, this should push learners

to identify regularities in the input and to try and develop rules accounting for them, for the sake of grammatical correctness and broadly understood scholastic success. Therefore, even if case morphology is not necessary to express what learners mean, they may be pushed to acquire it by the very context in which acquisition takes place.

5.7. CORRELATING THE STRUCTURED TESTS WITH SEMI-SPONTANEOUS OUTPUT

We now turn to the main research question of this chapter, namely whether or not morphosyntactic competence differs depending on the context in which it is elicited. For each subject, Table 52, adapted from Table 51 above, presents the number of correctly marked nouns, the total number of obligatory contexts, and the probability of observing this proportion of correct case marking if the learner is not applying a morphosyntactic principle. The lower the p value, the surer we can be that the learner systematically and correctly inflects nouns on the basis of their syntactic function. To make sure that they are not just repeating unanalysed chunks, we also report the number of lexemes on which the calculations are based. The last column, finally, allows us to correlate this quantitative analysis of learner spontaneous output with their performance in the structured tests, here expressed in terms of the relevant scenario (see chapter 4, p. 129).

Table 52: correlation between spontaneous interaction and structured tests

Subject	correct	Total	Types	p	Scenario
5102	10	12	8	< 0.01	1;1
5116	2	2	2	< 0.01	1;1
5101	4	6	4	0.11	1;1
5112	2	3	3	0.13	1;1
5113	3	7	5	0.50	3;1
5118	0	3	3	0.88	3;1
5117	2	4	4	0.31	2;1
5115	2	6	5	0.66	2;1
5119	0	8	4	1.00	2;1
5111	1	2	2	0.25	2;3
5107	1	3	3	0.50	2;3
5114	1	5	5	0.81	2;3
5106	8	8	5	< 0.01	4;1
5109	6	13	7	0.50	4;1
5104	1	4	3	0.69	4;1

With the only exception of 5106, all the subjects with a p value below 10% belong to scenario "a". This is equivalent to saying that for a learner to systematically produce case inflection in spontaneous production, he must be able to process both SO and OS targets in both the EI and the comprehension test. We find quite striking that even learners who in the context of the EI test proved to be able to repeat the ACC ending in SO targets, failed to do so in their spontaneous output. This ability was only exhibited by those subjects who were able to repeat the ACC ending in OS targets, which by hypothesis is a more demanding task. We remind the reader that no learner produced OS structures in the interaction task, although it could be argued that no appropriate context was supplied to do so. In any case, it is quite striking that the ability to manipulate case in syntactically marked structures in the structured tests should be a requisite to produce case marking correctly in

syntactically *unmarked* structures, but in a more realistic setting. This reminds us of Tarone's argument regarding the primacy of spontaneous over careful speech. The latter style, in contrast to the former, offers the learner the possibility to monitor his speech, thus avoiding typical interlanguage structures and focussing instead on target-language ones, perhaps with the help of explicit competence. It follows that only spontaneous speech presents the researcher with a picture of actual, that is, automatised (DeKeyser 2001) learner competence. The same could be said about our comparison. The structured tests undoubtedly make it easier to plan and monitor one's own output, whereas the strict rhythm of real-time interaction arguably produces a more spontaneous version of the learner variety, inclusive of all its creative means to express meaning at odds with the rules of the target language.

5.7.1. Discussion

The results of our analysis fit quite well one of the two scenarios we had hypothesised in the introduction, where we had claimed that morphological accuracy should be lower in interaction than it is in the structured tests because the former task is cognitively more demanding: in addition to retrieving lexical items and producing them in the appropriate word-form, learners have to keep track of the discourse situation and adapt their output to the interlocutor. In such a complex situation, learners can devote fewer attentional resources to such an accessory level as inflectional morphology, and focus on other more urgent levels of language. This is in line with the Limited Attentional Capacity Model proposed by Skehan and Foster (Skehan 1996, 2009, Skehan and Foster 2001), predicting that the greater the complexity of the task, the more learners will have to choose how to allocate their limited attentional resources while always privileging the processing of meaning over form. Truly, the structures produced in spontaneous speech are less complex than those encountered in the structured tests, as no utterance with OS word order was produced. At the same time, though, the effort required of the

learner is much greater in the new communicative context than it was in the tests, which in fact only dealt with mock meaning, or no meaning altogether. In contrast, interaction with a partner requires participants to encode and decode meaning in real time. Moreover, the expression of meaning is functional to achieving a concrete objective, which, although not necessarily important to the learners, still represents an accepted part of foreign language classes. In sum, Skehan's model seems to describe the situation well, and its predictions are indeed borne out in the data.

A suggestive complementary hypothesis which, regrettably, we will not be able to test adequately with the data at hand is that some learners may perform more poorly in interaction than in the structured tests as a consequence of alignment to the speech style of their partner - regrettably, as witnessed by the quantitative data, the reverse does not seem to take place. Alignment consists in imitating to a certain extent the way in which the interactional partner speaks, for instance by using the same words or grammatical structures. Among native speakers, it is a powerful tool to facilitate both comprehension and production, allowing the speaker to focus on more content-related activities such as planning (Pickering & Garrod 2004; Pickering & Garrod 2006). In the case of L2 learners, alignment may not be as profitable because the learner variety is not yet fully automatised, an important requirement, given the mostly unconscious nature of this process. On the other hand, the gains in attentional resources should be much more precious to the L2 learner than to the proficient or native speaker (Costa, Pickering & Sorace 2008).

The theme of alignment intersects with that of (implicit) feedback. Recent work on the same data discussed here (Saturno in prep) has shown that corrective feedback, in addition to often being non-target-like, in general tends to be quite rare in the interaction task, so that it is unlikely that it should prove a major source of morphosyntactic processing. In fact, we suggest here that a much greater role could be played by peer feedback in a broader sense, understood not as corrective interventions, but rather as negative example. Let us suppose that a learner has provisionally identified a form-function association between

case endings and syntactic function, and is structuring his output accordingly. His partner, on the contrary, may still be at a less advanced stage of interlanguage competence, one, let us say, in which all nouns occur in their invariable basic word-form. As the first learner is still in the process of testing the hypothesis regarding the morphosyntactic behaviour of Polish nouns, the ungrammatical output of his partner may unfortunately be interpreted as implicit negative feedback, causing his own output to go back to a previous developmental stage. If at all, the hypothesised scenario may take place at any time during the interaction, so that it could account for the otherwise asystematic variability sometimes encountered in the speech of a single learner. While a detailed analysis of this hypothesis is beyond the scope of this work, and would probably prove inconclusive anyway, it remains a suggestive path to pursue in the future.

In sum, our results show the performance in the structured tasks may exhibit superior accuracy compared to the speech produced by the same speaker in an interactional task, keeping exposure to the input and target structure unchanged.

The final question we have to answer, on a theoretical level, is what performance - structured tests or interaction - we should take as representative of the learner's present competence. This is probably a false problem, as both results were produced by the same real learners, and fit into hierarchies of development. As we have seen in chapter 4, p.129, the four contexts created by the structured tests can be ordered along the following hierarchy: OS repetition \supset OS comprehension \supset SO repetition \supset SO comprehension. Accurate use of morphosyntax in interaction represents a further step to the left of the above sequence. To be sure, there are exceptions to these regularities, but in addition to being rather rare, they can be mostly explained through methodological faults intrinsic to the repetition test employed in the VILLA project (see Saturno (forthcoming) for a discussion). Each step in the sequence above represents an ability which learners may or may not have mastered, along with all the steps to its right, which can be interpreted as its prerequisites. Only two details do

not fit in this account. Firstly, learners who proved able to master OS structures in the tests cannot correctly inflect nouns in SO structures when observed in interaction, despite the fact that the latter lie more to the right in terms of the implicational hierarchy. We have no evidence as to whether or not these learners can process OS targets in interaction, although that seems highly unlikely. Secondly, and conversely, some learners can accurately repeat SO targets in the structured context, but cannot produce target-like morphology in the same target structure in interaction. Concerning the latter group, then, we could ask: can they or can they not produce accurate case marking in SO structures *in general*? Or, in other words: given the different results obtained in the two elicitation contexts, which one reflects more faithfully what learners can *really* do with inflectional morphology?

To answer, we should first decide what "really" means in the preceding sentence. We start by pointing out that while interaction involves the *production* of a message, the structured test only requires its repetition. Even if we accept the assumptions of the repetition test, namely that learners do not merely repeat a sequence of sounds, but decode and re-encode its meaning according to the rules of their learner variety, still, the target is provided by the test, so that learners need not concern themselves with its formulation. They do have to decode it, though. This process in turn may seem less crucial in the particular interactional task described here, mostly consisting of a sequence of monologues and strongly anchored in context.

This said, we come back to the question: can they, or can they not produce accurate inflectional morphology in SO structures? Quite evidently, they can, as witnessed by the transcripts of their speech. It can be argued, though, that they can do so under very specific conditions, in terms of test structure - as just mentioned - but also of setting. Although it is true that Gass and colleagues (2011) did find that interactional patterns do not seem to vary between a class and a laboratory setting, we still argue that correctly inflecting nouns in a quiet room in the presence of a researcher, speaking to a computer and saying nothing meaningful is quite a different task from speaking to a fellow learner in

a loud foreign language class, trying to solve a task in a limited time. So, learners *can* inflect nouns in SO structures, but context matters. We could thus reformulate our question as following: can learners produce morphosyntactically accurate SO structures in a realistic setting? Accordingly, instead of wondering what performance best represents learners' "real" competence, we might ask which result represents what they can really do using the L2 when put to the test in a real-life context. Put in these terms, the answer is rather straight-forward.

5.8. CONCLUSION

This chapter aimed at analysing semi-spontaneous speech in interaction, elicited thanks to a task in which learners spontaneously produced a good number of target structures, namely feminine nouns in transitive sentences.

Among the main results of the analysis, we may list the following.

1. The subject is always expressed by a name (e.g. *Anna*) or a pronoun (e.g. *ona*, "she"), never by a common noun (e.g. *aktorka*, "actress") as was the case in the two structured tests. For this reason, the analysis concentrates on the object function, corresponding to the accusative case.
2. The object is most often represented by an inanimate noun (e.g. *herbata*, "tea"). Animate (e.g. *pies* "dog") and human (e.g. *córka*, "daughter") are relatively rare. This partly reflects the input learners were exposed to (see chapter 6, p.198), in which, based on their semantics, specific lexical items are more likely to perform the subject or object syntactic function.
3. In spite of this uneven distribution in the input, the accuracy of morphosyntactic marking does not appear to depend on the relative frequency of accusative word forms on the total occurrences of a lexeme.
4. Still, it may be that learners individually assumed a specific word-form of a given lexical item to be its basic, invariable word form. To verify this claim, we analysed the output of those learners who repeat the same

lexemes more than once. Alongside limited variability, we found that the same word tends to occur in the same word-form when it is repeated, lending partial support to the hypothesis. The limited amount of data does not allow for any generalisations, though.

5. Statistical analysis of case-marking shows that only few learners inflect nouns with above chance accuracy. Since no data are available for the case-marking of nouns in the subject function, though, we cannot exclude that these learners produced all nouns in their accusative word-form.
6. Qualitative analysis of learner output shows that independently on the accuracy with which case-marking is produced, utterances are shaped by animacy contrasts and default SO word order. The combination of these two principles makes morphosyntax superfluous, at least relative to the level of complexity required by the task discussed here.
7. The same patterns can be found in the input, too. It appears that only very few utterances produced by the native speaker require morphosyntactic analysis to retrieve meaning. Against this picture, the communicative need for case marking appears rather doubtful. Instead, learners may be pushed to acquire it because of the implicit rules and objectives of the scholastic context.

We further correlated the results summarised above with the two structured tests discussed in the preceding chapters, reaching the following conclusions:

8. Most of the learners who performed above chance in interaction also succeeded in both tests and with both word orders. It thus seems that being able to manipulate word order and case marking in comprehension and repetition is a prerequisite for correctly inflecting nouns in interaction, albeit with unmarked word order only.

The ability to manage morphosyntax in an interactional context then can be seen as the last step of the hierarchy identified for the structured tests: OS repetition \supset OS comprehension \supset SO repetition \supset SO comprehension. Although the actual grammatical structures required of the learners may not be as complex as in the structured tests, still interaction appears to be a much more demanding task, which can be justified in terms of the much greater number of variables among which learners have to disperse their limited attentional resources. Since the expression of meaning has priority over the processing of form, we could expect learners to accurately case-mark nouns only if that process has become automatized in their learner variety. We conclude that while the structured tests provide researchers with a good opportunity for observing the cutting edge of the developing grammar, i.e. the very best of what the learner can do, perhaps with some assistance of his explicit competence, the interactional context brings to light the real competence of the same learner, that is, what he can really do in a real-life situation. We label the two types of competence, respectively, "potential" and "actual", and argue that research on SLA, especially if aiming to didactic application, should focus on the latter.

6. INPUT

Unlike most other key points of SLA, that input should be necessary for acquisition is recognised in all theoretical approaches, in addition to be rather self-evident. Even the "poverty of input" (Chomsky 1965; 1980) argument put forward in the Generative framework implies that although in possibly minimal quantity, input is nevertheless required to set UG parameters on the value appropriate for the language being acquired, be it the L1 or an additional L2.

Opinions start to differ broadly, however, when the debate moves to the input parameters which influence acquisition success the most, or to the way in which input, "what is available to go in" is processed and transformed into intake, "what goes in" (Corder 1967:165).

In the generative framework (see Rankin & Unsworth (2016) for a recent review), input is a mere activator of an innate mechanism. The "poverty of input" argument states that a person's competence is more complex than the input which supposedly shaped it (Chomsky 1981), so that it is unlikely that such complex competence could develop on the basis of such imperfect input alone; instead, competence as a system is innate, and input only provides the raw material which the language processor will take care to shape and systematise through the acquisition process. Among the many imperfections of input, Pullum and Scholz (2002)⁸⁵ list the following:

- ingratitude: learners are not rewarded for their achievements;
- finiteness: despite being only exposed to a limited number of sentences, people can create a potentially infinite number of new utterances;
- idiosyncrasy: input is not the same for everyone, yet language as a system is largely shared by all people who are competent in a given code;

⁸⁵ While these authors' work is devoted to L1 acquisition, we find that the same arguments are equally applicable to SLA.

- degeneration: numerous sentences are incorrect or incomplete.

In contrast, emergentist approaches maintain that input contains a wealth of information which learners are equipped to analyse statistically in search of form-function associations, i.e. co-occurrence of a given meaning (e.g. "feminine noun") and a form (e.g. "ka": see Saturno (2016) for the cited example, based on VILLA data). Ellis (2006b:1) describes learners as "intuitive statisticians, weighing the likelihoods of interpretations and predicting which constructions are likely in the current context", and the acquisition process as "the gathering of information about the relative frequencies of form-function mappings". While the focus of our work is on L2 acquisition, similar claims and methods are also common in L1 acquisition research (see Behrens (2006) and Tomasello (2005) for an example). This approach will be examined at length and indeed applied to the VILLA data in section 6.2.1, p. 206.

Even if form-function associations can be identified statistically, their strength is likely to be influenced by a variety of factors which can hardly be studied with a quantitative approach. Among these we could list teaching methodology (Norris & Ortega 2000; VanPatten & Borst 2012), linguistic environment (LaBrozzi 2012), the learners' motivation and attitude, as well as their interactions. Flege, for example (see Flege (2009) for a review and a discussion) has shown that what is typically referred to as "age effect", i.e. the supposed acquisitional advantage of younger people, strongly correlates with important input variables, such as the amount and the quality of native input. Specifically, while younger immigrants usually go to school and have extensive contacts with their native peers, older immigrants commonly spend most of their time within their L1 community. Limited education also limits their access to jobs in which native input is predominant, thus exposing them to pidginised varieties, at most. With this we do not intend to discuss whether or not age effects exist, as that is decidedly beyond the scope of our work⁸⁶; our

⁸⁶ See Andorno & Valentini (forthcoming) and Singleton (2012a; 2012b) for a review on this topic.

point input processing is a complex of process in which numerous factors simultaneously play a role and engage in complex interactions with each other (Rescorla & Wagner 1972; Goldschneider & DeKeyser 2002; Collins et al. 2009; Wulff et al. 2009).

Thus, despite its obvious importance in the acquisition of language and the body of research just briefly referred to, the ways in which input is processed still largely mysterious to us. This is mainly because a detailed analysis of input is methodologically prohibitive. The study of input involves all sorts of practical difficulties related to monitoring and recording the variegated input to which each individual learner is exposed. Yet, without this information, the doubt remains that different acquisition outcomes may simply depend on input differing in quality and quantity, hence the importance of analysing and controlling for this variable.

The intuitive thing to do in order to overcome any uncertainty about this variable would be to monitor all input to which a learner is exposed. This was indeed attempted by Roy and colleagues (2009), who constantly video- and audio-recording the learning experience of a child by means of cameras strategically placed in various rooms of their home. This methodology has several limitations, though: first, such constant monitoring is only possible in specific places (although portable microphones were also developed for similar projects), and second, the analysis of the data collected are only possible through automatic, statistical analysis. The amount of information is simply too large to be manually processed by a researcher. Finally, the findings relative to a specific learner can hardly be generalised to other, as the learning experience of an individual is to a large extent unique. Experimental designs concerning the effect of a specific variable therefore are hardly affordable.

Alternatively, the individual work of several researchers can be brought together in a single database thanks to a common transcription and meta-data system (MacWhinney 2000), as is the case with the CHILDES project (MacWhinney & Snow 1985).

Others still, wishing to testing specific hypothesis in a controlled experimental setting, have turned their attention to artificial languages as a means to expose learners to fully controlled input (Reber 1967; Hulstijn & DeKeyser 1997; Mueller 2006; Robinson 2010). Just like in the VILLA project, the main idea is that by focussing on absolute beginner learners of a language which can only be learned during the experiment itself, everyone's learning experience should be identical. Artificial languages invented *ad hoc* for a specific project certainly present the best environment for such studies, as there is simply no way learners could be exposed to the target language beyond the experiment. However, such methodology may raise doubts as to its ecological validity, as artificial languages typically lack the complexity and idiosyncrasies which are characteristic of natural languages.

Other studies, typically conducted with a longitudinal design (e.g. Perdue 1993; Giacalone Ramat 2003), simply do not even attempt to control the input and rely on the assumption that learners will be exposed to 'average' input, especially over sufficiently long periods of time. Incidental evidence has shown that this is not always the case: Spreafico (2005) has shown that learners may know lexical items which are very rare in the average input, but are particularly salient to them as a consequence of their job or place of residence; at the same time, very common words may be utterly absent from their output. This methodology, moreover, makes it problematic if not impossible to correlate learner output with the input received in a rigorous way.

Others still, like Collins et al. (2009), base their input analysis on samples extracted from a relatively homogeneous context (in that case, classroom L2 Instruction), assuming that the rest of the input should be more or less homogenous. While this approach may enjoy good ecological validity in the case of learners of intermediate and upper proficiency, it becomes problematic when attention is turned to another sensitive topic in SLA research, namely the very initial stages of L2 acquisition, which lies at the heart of the VILLA project. This particular environment is particularly important because in a rather circular manner, studying the strategies of input processing applied by learners

confronted with a completely novel language may be illuminating with regard to the general mechanisms of input processing (Rast et al., 2011; Perdue 2002). In order to do this, however, it is essential to have full control over the input, so as to correlate learner output with the relevant input parameters.

To summarise, it would be highly desirable to investigate the development of an L2 from its very onset, with a natural target language and whilst retaining full control over the input. Precisely such is the methodological challenge undertaken by the VILLA project. Input control is taken to a wholly different level, in that it was integrally recorded and transcribed. This makes it possible to study factors regarding the linear sequence of linguistic elements, such as frequency effects, as well as others regarding the properties of the audio signal, such as perceptual prominence, although this latter option is methodologically more complicated and time consuming.

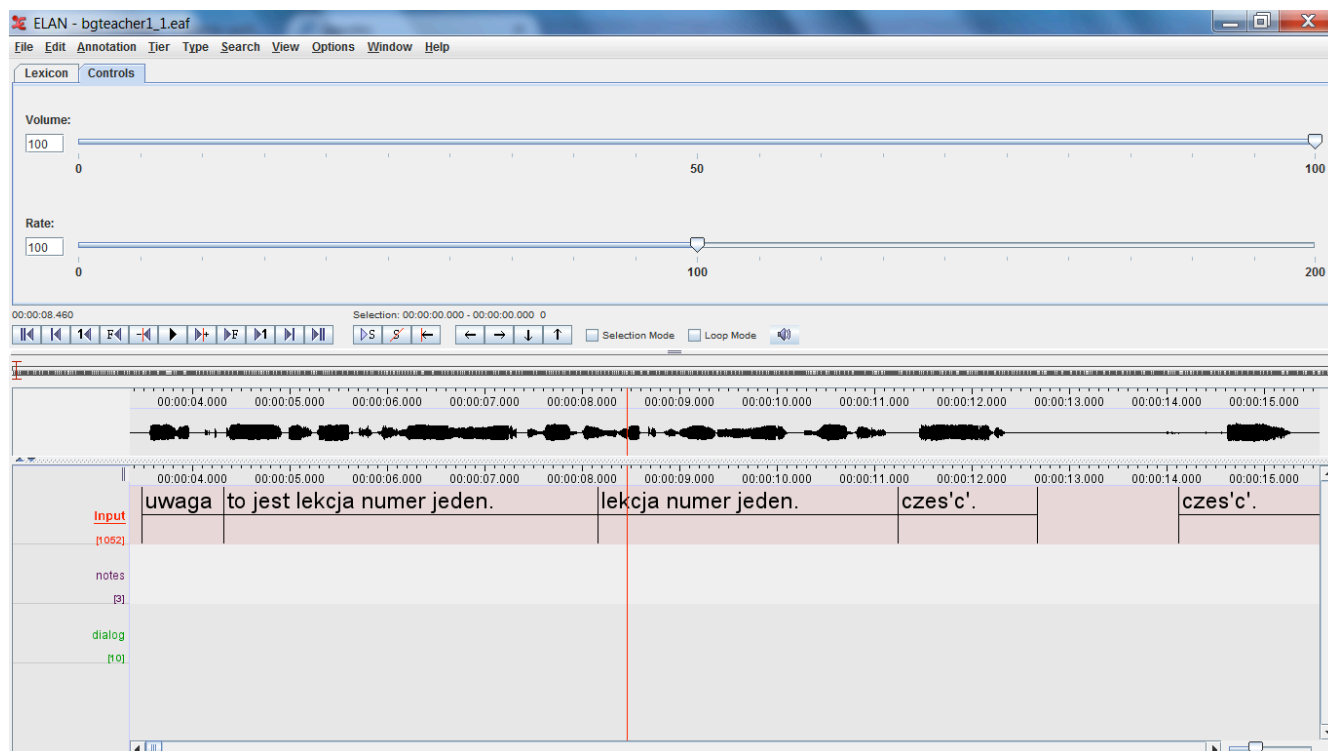
6.1. THE VILLA METHODOLOGY: INPUT CONTROL

Input was carefully planned in advance in terms of topics, vocabulary and frequency of both lexical items and syntactic structures. During the course, the teacher wore a portable wireless microphone which recorded her speech; this in turn, in addition to only a few support powerpoint slides, represent the only input which was provided to the participants.

The resulting tracks were subsequently transcribed⁸⁷ in standard orthography using ELAN (Brugman & Russell 2004). This software makes it possible to time-align transcriptions, i.e. to automatically associate each annotation with the corresponding audio segment (Figure 6). This feature is particularly useful for solving doubts about details in the transcriptions as well as for distinguishing prosodic contours. Transcription is produced from left to right along the horizontal axis; participants are assigned different tiers which are listed from top to bottom.

⁸⁷ The Italian and English input were transcribed by ourselves; the French, German and Dutch editions were internally transcribed by members of the corresponding research teams.

Figure 6: transcription of the VILLA input with ELAN



To further separate input addressed to all learners from comments aimed at individual learners or groups during games, teacher speech was transcribed on two different tiers, labelled *TEA and *TEB respectively. In this work we will only consider the former, which represents the vast majority of utterances anyway.

The resulting ELAN files were exported into the CHAT/CLAN format (Figure 7), thanks to which the transcripts are converted into a text-like top-bottom format.

Figure 7: transcription of the VILLA input in CHAT

6 *TEA: uwaga
7 *TEA: to jest lekcja numer jeden .
8 *TEA: lekcja numer jeden .
9 *TEA: cześć .
10 *TEA: cześć .
11 *TEA: cześć .
12 *TEA: cześć .
13 *TEA: bardzo dobrze .
14 *TEA: cześć .
15 *TEA: to jest kontakt nieoficjalny .
16 *TEA: kontakt nieoficjalny .
17 *TEA: cześć .
18 *TEA: cześć .

A CHAT-CLAN automatic morphological tagging system was developed by Christine Dimroth and Roman Skiba, with a little participation on our side concerning the morphological tagging of Polish entries. (Figure 8).

Figure 8: automatic morphological tagging of the VILLA input in CHAT

*TEA: uwaga
%mor: Noun|uwaga:Nom-fem-Sg=attention
*TEA: to jest lekcja numer jeden .
%mor: Pron|ten:demon-3-Sg-neutr-Nom=this Cop|być:3-Sg=be
Noun|lekcja:Nom-fem-Sg=lesson Noun|numer:Nom-masc-Sg=number
Num|jeden:cardinal=one .
*TEA: lekcja numer jeden .
%mor: Noun|lekcja:Nom-fem-Sg=lesson Noun|numer:Nom-masc-Sg=number
Num|jeden:cardinal=one .
*TEA: cześć .
%mor: Adv|cześć=bye-or-hello .
*TEA: cześć .
%mor: Adv|cześć=bye-or-hello .

On the dependent tier %mor, each word is morphologically tagged with the appropriate values of the relevant grammatical categories, depending on the part of speech considered (e.g. case, gender, number and lexeme for nouns; person, number and lexeme for verbs, and so on). To each item in the original transcript, the algorithm associates the appropriate gloss, retrieving it from a

specially designed lexicon. In case a given form corresponds to more than one label, which is fairly common in Polish due to widespread syncretism, all labels are presented subsequently.

Building on that work, we developed our own system by adapting the same principle to a different tool, namely the software R (R core team 2014) and its package *stringr* (Wickham 2012). An example of the resulting gloss is provided in (46). Appropriate categories and labels were also devised for all grammatical categories, such as verbs, pronouns and adjectives.

(46) *balonik*

sostantivoIN_Acc_Mas_Sg//sostantivoIN_Nom_Mas_Sg:balonik

For the exemplified entry, the gloss provides the information schematised in Table 53.

Table 53: automatic morphological tagging, labels and values

Label 1		Label 2	
grammatical category:	noun	grammatical category:	noun
animacy:	inanimate	animacy:	inanimate
Case:	Accusative	Case:	Nominative
Number:	singular	Number:	singular
lexeme:	balonik	lexeme:	balonik

The input transcript, once glossed, can be searched for appropriate patterns. Searching for the string "SVO sentence with animate masculine nouns as subject and inanimate feminine nouns as object", for example, should retrieve hits like *Leon lubi herbatę*, "Leon likes tea", among others. In addition, it is possible to identify all the instances of a given lexeme or grammatical value such as, for example, "nominative masculine singular". In short, one can retrieve all the values of the categories listed in Table 53, as well as their combinations. This tool will prove useful to address several questions related to

the input. We will consider form-function association between case endings and syntactic function (6.2.1), type frequency distributions for the subject and object function (6.2.2) and the distribution of transitive patterns in terms of gender, animacy and syntactic function (6.2.3).

6.2. STUDYING FREQUENCY EFFECTS IN THE VILLA PROJECT

Frequency is probably the most intuitive characteristic of input one could study, and indeed it has received considerable attention in both L1 (Gülzow & Gagarina 2007; Lieven 2010) and L2 research. In the latter field, this parameter has been extensively treated in a number of studies which it would be impossible to review here: moreover, frequency in itself is such a complex concept, that in the end this body of literature is extremely varied. In what follows we will focus on specific aspects of frequency which are directly relevant to our study: for introductory reviews on this topic, however, we signal among others Goldschneider and DeKeyser (2002), Ellis and Collins (2009) and especially Ellis (2002) and the other studies collected in the same special issue of *Studies in Second Language Acquisition*.

6.2.1. *Form-function association*

In line with the functionalist approach adopted in this work, the first analysis we turn to regards the strength of form-function association, that is, how commonly and unambiguously a given form is used in the input to veiculate a given function, and vice versa.

The idea has been extensively developed in a variety of studies belonging to different theoretical frameworks, concerned with both L1 and L2 development. In a study on the development of verbal morphology by toddlers, Hadley and colleagues (2011) found that the best predictor of age of acquisition was input informativeness, operationalised as the proportion of overt vs. ambiguous verb

forms for tense. The concept behind this measure is certainly reminiscent of other ideas developed in different theoretical frameworks, such as *cue validity* in the Competition model (MacWhinney & Bates 1987), or *uniformity* and *transparency* within the paradigm of Natural Morphology (Dressler 1987). This methodology has been successfully applied to several L1s, leading to the discovery of two important facts. First, the same form-function association may have different relevance in different languages: for instance, the agent function is signalled with the greatest reliability by utterance-initial position in English, but by subject-verb agreement in Italian (MacWhinney, Bates & Kliegl 1984). Second, the strength of form-function association is a better predictor of acquisition success than the degree of complexity of the target structure. In this respect, Kempe & MacWhinney (1998) demonstrated that the Russian case system, although more complex than its German equivalent, is more rapidly acquired because the relationship between case endings and meaning is more easily retrievable by L2 learners. By the same token, L1 learners of a morphologically complex language like Turkish manage to master its case system by age 2, which rarely happens with learners of Indo-European languages (Slobin 1985:275). The reason for that lies in the fact that although Turkish exhibits a far greater number of inflectional morphemes, most of them are univocally associated with a single meaning, whereby in the mainly fusive Indo-European languages, each morph typically expresses more than one meaning. To this one must add widespread syncretism, whereby the same morph may express different and competing combinations of meanings depending on the paradigm of which it is part. Further, it had been noted that in L1 acquisition diminutive nouns are acquired earlier and more accurately than the corresponding primitives. Extensive research has come to the conclusion that they facilitate gender assignment by providing a single, unambiguous ending instead of many, non transparent ones. Evidence is consistent across L1 and L2 learners of different languages (Haman 2003; Dąbrowska 2006; Savickienė & Dressler 2007; Kempe et al. 2009).

In the VILLA project, finally, the same paradigm has proved useful to explaining learner preferences in copular structures (Saturno 2015a) and word-formation strategies (Saturno 2016).

The raw material to quantify the strength of form-function associations is input. While most of the studies conducted within this paradigm relied on input samples, or on the written material used as teaching aid in the case of instructed SLA, the VILLA project offers the possibility to thoroughly compute the frequency of any form and function over the entire experimental course. In what follows, we will attempt to quantify the strength of the association between case endings and the corresponding syntactic functions. The data concerning the relation between the ending *-/e/* and the function "accusative singular of feminine nouns" are presented in Table 54.

Table 54: association index, accusative case

a	B	C	D	e	f	g	h
L1	Corpus	ACC FEM <i>-/e/</i>	ACC FEM	<i>-/e/</i>	c/d	c/e	f*g
EN	41776	415	441	6541	94,10%	6,34%	5,97%
FR	25763	345	365	3789	94,52%	9,11%	8,61%
GE	43219	379	414	6424	91,55%	5,90%	5,40%
IT	53965	598	624	7675	95,83%	7,79%	7,47%
NL	45554	494	519	6570	95,18%	7,52%	7,16%

We first turn to the columns presenting absolute figures (b to e). It is easy to see that the five corpora differ somehow in length. The French corpus, specifically, is significantly shorter than the other four due to several missing transcripts: in other words, some of the 19 sub-lessons are not represented. One may think that such difference is of little importance for our purposes, since we are only concerned with relative figures. However, since sub-lessons differ in the grammatical structures on which they focus, it is possible that the proportion of target structures on the total size of the corpus could vary. Fortunately this is not the case, which witnesses to the fact that most classes in

which transitive structures were introduced are indeed represented in the partial French corpus available.

The instances of accusative case of feminine nouns in *-/e/* (column c) represent on average 1.1% of the total number of words in the corresponding corpus (range: 0.9% - 1.3%). The attentive reader will notice that these figures are much larger than those provided later on in Table 60 and Table 61 (p. 227 and 227), relative to the frequency of full transitive constructions. The examples in (47) may help us to illuminate this apparent incongruity. Thanks to the pro-drop character of Polish, nouns in the accusative case may appear in structures lacking an overt subject (47a) or both subject and verb (47b). Such elliptic structures were not computed as fully-fledged utterances, hence the discrepancy.

- (47) a. *TEA: *lubi kaw-ę tak*
 likes coffee-ACC yes
 "(he) likes coffee yes"
- *TEA: *Daniel lubi kaw-ę*
 Daniel likes coffee-ACC
 "Daniel likes coffee"
- b. *TEA: *co on lubi?*
 what he likes
 "what does he like?"
- *TEA: *Robert?*
 "Robert (learner name)?"
- *TEA: *Fotografi-ę tak.*
 photography-ACC yes
 "photography yes"

Turning back to our form-function association, instances of accusative case of feminine nouns in *-/e/* represent the co-occurrence of the form *-/e/* and the function "accusative case of feminine nouns". To evaluate how strongly the two

are associated, we now need to compute their frequency independently of each other: we will thus search for feminine nouns in the accusative case, but not necessarily ending in *-e/*, and for words ending in *-e/*, but not necessarily representing the accusative case of feminine nouns.

Regarding the former, the table shows that there are hardly any examples of feminine nouns in the accusative case *not* ending in *-e/*. This is not unexpected: as mentioned, feminine nouns were almost exclusively chosen from the *-a* class, which only exhibits one accusative ending independently of semantics, unlike masculine nouns for which animacy is also relevant. The ratio, therefore, should be close to one, a situation which is arguably quite rare in a natural language. It only occurs in the VILLA data because the input was simplified for experimental reasons.

Predictably, the accusative of feminine nouns represents only a very small part of all Polish words ending in *-e/*, graphically represented as *<e>* or *<ę>*⁸⁸. A necessarily inexhaustive list of the most frequent words in *-e/* includes the following:

- the reflexive particle *się*, as in *on boi się*, "he is afraid";
- the negator *nie*, "not";
- 1st person singular of verbs, as in *lubię*, "I like";
- 3rd person singular of verbs, as in *on idzie*, "he goes";
- adverbs, as in *dobrze*, "good";
- the direct case of neuter nouns in *ę*, as *imię*, "name";
- the locative case of most nouns, as in *w Warszawie*, "in Warsaw".

To summarise, almost all feminine nouns in the accusative case end in *-e/*: this represents the form > meaning index, that is, the likelihood that a

⁸⁸ The Polish so-called 'nasal vowel' *<ę>* has various phonetic realisation depending on the context in which it occurs. In the word-final position which is most relevant for our analysis, it is always realised as [e]. Other, more "nasal" realisations like /eŋ/, typical for instance when followed by a fricative, should be considered as hypercorrect and never occur in the VILLA data. See Gussman (2007) for further details.

given function should be expressed by a given form. Conversely, only a little proportion of the words ending in *-e/* represent the accusative case of feminine nouns: this represents the meaning > form index, that is, the likelihood that a given form expresses a given function. The global association index is simply computed as the product of the two indexes just discussed⁸⁹. The information reported in Table 54 is explained schematically in Table 55.

Table 55: form-function association indexes

a	L1;
b	Corpus length in words;
c	occurrences of accusative singular of feminine nouns instantiated by the ending <i>-e/</i> ;
d	occurrences of accusative singular of feminine nouns, independently of their ending;
e	words ending in <i>-e/</i> , independently of their meaning or function;
f	Form > function index: likelihood that the function of interest should be expressed by the particular form investigated;
g	Function > form index: likelihood that the form in question should refer to the function of interest;
h	Association index: product of f and g. Overall strength of the form-function association.

The association index is a relative value comprised between 0 and 1, which indicates how biunivocal is the relation between form and function. Both extreme values are rather unrealistic in a natural language: in our case, for instance, an association value of 1 would indicate that all words in *-e/* are instances of the accusative case of feminine nouns, and conversely, that the

⁸⁹ These measures are strongly indebted to the theoretical and methodological frameworks of Natural Morphology (Crocco Galeas 1998; Wheeler 1993; Wurzel 1989; Dressler 1985; 1987), on the one hand, and the Competition Model (MacWhinney & Bates 1987; Macwhinney, Tomasello & Slobin 2005; Kempe & MacWhinney 1998), on the other hand, of which they represent a simplification and an adaptation to the VILLA data.

accusative case of all feminine nouns ends in *-/e/*. While the latter condition is almost met in the VILLA corpus, in which there are no instances of feminine nouns belonging to other inflectional classes, the first criterion is decidedly less realistic.

As relative figures make little sense on their own, Table 56 below provides the figures relative to the association index between the form *-/a/* and the function "nominative singular of feminine nouns".

Table 56: association index, nominative case

a	B	c	d	e	f	G	H
L1	Corpus	NOM FEM <i>-/a/</i>	NOM FEM	<i>-/a/</i>	c/d	c/e	F*g
EN	41776	2437	2471	7091	98,62%	34,37%	33,89%
FR	25763	1730	1798	4239	96,22%	40,81%	39,27%
GE	43219	2265	2338	7100	96,88%	31,90%	30,91%
IT	53965	3161	3198	9345	98,84%	33,83%	33,43%
NL	45554	2721	2787	7483	97,63%	36,36%	35,50%

The association index for this relation is much larger than in the previous example. The reason is quite simple: on the one hand, the function > form index is practically the same, as again, virtually all feminine nouns in the nominative case end in *-/a/*, as a consequence of the common inflectional paradigm. On the other hand, the form > function index is much higher, mainly because feminine nouns occur much more frequently in the nominative than in the accusative case, while the number of words ending in *-/a/* is not too dissimilar from that of words ending in *-/e/*. Among the former, we may list:

- the accusative case of animate masculine nouns, like *Karola*, "Karol-ACC";
- the adversative conjunction *a*, "and/but";
- 3rd person singular of verbs, as in *on ma*, "he has";
- prepositions, like *na*, "on";

- the pronoun *ona*, "she";
- the nominative case of feminine adjectives, like *sympatyczna*, "friendly".

A comparison of the two association values indicates that it should be easier for the learner to associate *-a* to the nominative case than *-e* to the accusative. This is indeed consistent with the results of the structured tests and of the interactional task, in which feminine nouns in all syntactic functions, if the learner has not yet mastered the inflectional paradigm, typically occurs in one invariable form in *-a*. The strong association index of *-/a/* as an index of the nominative case certainly contributes to explaining why this, and not other forms of the paradigm, should be selected as the basic word form. There may be more than a mere matter of relative frequency, though. The citation form in which lexical items first appear in the course, for instance, is the nominative form. Moreover, the nominative case appears in a much wider variety of contexts than the accusative case. This argument is certainly related, but not identical to a higher relative frequency. Other factors, such as input skewedness based on semantics, may add complexity to the picture. To such questions we turn in the next section.

6.2.2. *Type frequency and input skewedness*

By type frequency we mean the number of lexical items which occur in a linguistic construction. The literature on this topic maintains two positions which may seem mutually exclusive, or perhaps complementary. On the one hand, several researchers argue that learners might benefit from the same construction being instantiated by a greater number of types (Bybee 1985; Plunkett & Marchman 1991; Plunkett & Marchman 1993), especially as far establishing abstract patterns is concerned (McDonough & Kim 2009). The same distribution is encountered in the verb types instantiating particular constructions in the speech directed towards infants by their caretakers, with

facilitatory effects on acquisition of novel constructions (Goldberg, Casenhiser & Sethuraman 2004 for L2, Casenhiser & Goldberg 2005 for L1 learning). Type frequency ensures productivity, as hearing several different lexical items in a certain context makes it less likely that that construction could become associated with any of them. Further, if a construction is instantiated by many items, then it is rather general in meaning and can be easily generalised to others; and finally, high type frequency ensures that that construction is used frequently (Bybee & Thompson 2000). Conversely, Krushke and Blair (2000) argue that learning that a particular stimulus is associated with a particular outcome makes associating the same outcome with another stimulus more difficult. Type frequency in turn might be determined by the semantic scope of a given construction (e.g. past tense with telic verbs), a relation which interests us directly. Other studies confirming the importance of type frequency, or variability, in the generalisation of constructions include Onnis et al. (2008), Ninio (2005) and Bybee (1995).

However, other researchers claim that learners may be equally helped in the acquisition process by a highly skewed distributions, in which a given grammatical constructions is mainly instantiated by a very small number of lexical items. The rationale is that construction learning is a process of categorisation (Goldberg, Casenhiser & White 2007), by which the learner - either child or adult - begins to recognise a similarity of meaning from an identical structure, albeit instantiated by different lexical items. Studies on non-linguistic categorisation have shown that learns are facilitated in the construction of categories by low-variance input (Gentner, Loewenstein & Hung 2007; Casasola 2005). The same is true for language (Casenhiser & Goldberg 2005; Maguire et al. 2008), with the additional difficulty that linguistic constructions are by nature abstract (Gentner & Medina 1998).

A typical example of skewed distribution which is commonly encountered in language is Zipf's law (Zipf 1935), according to which the frequency of a given word is inversely proportional to its rank in a frequency table. As a result, the first few most common words in a text account for a substantial proportion of

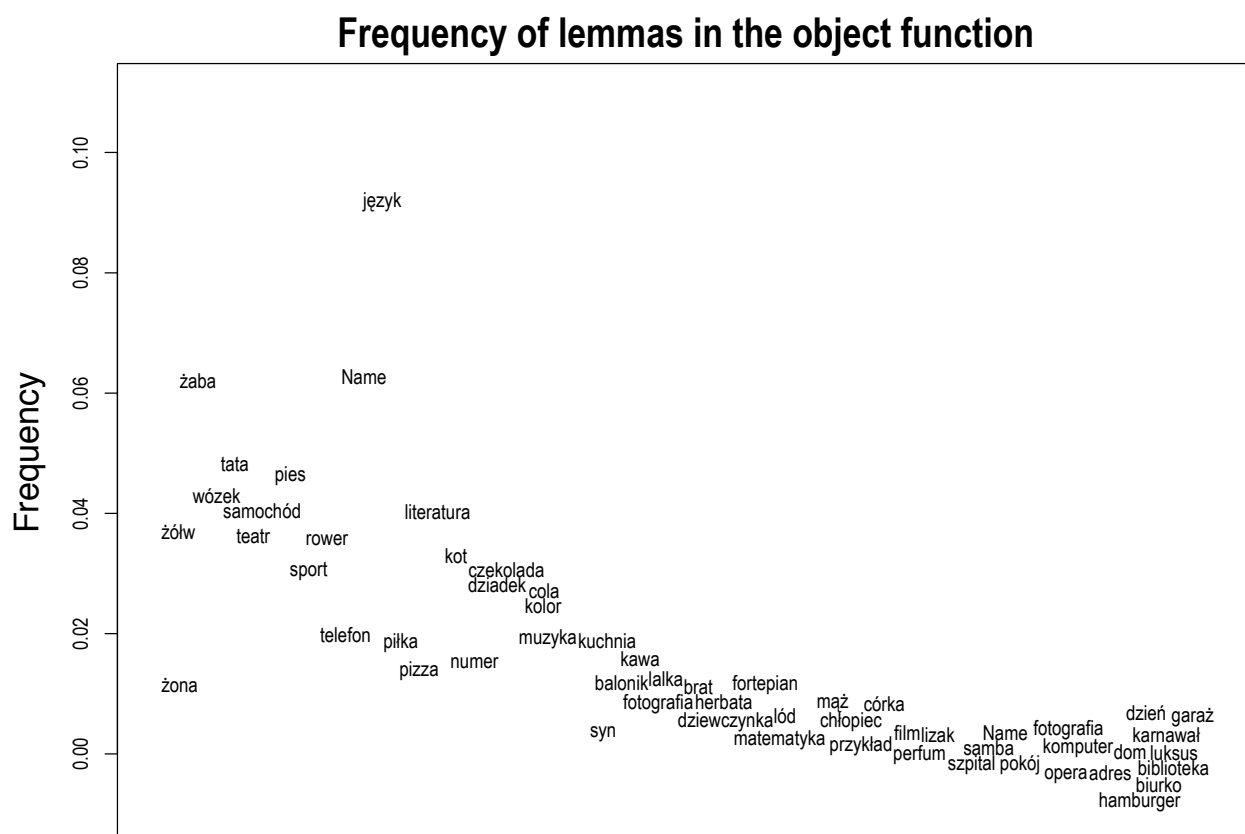
all tokens (Mintz, Newport & Bever 2002). There may be various reasons for this, usually linked with the semantics of the words involved: some verbs, for instance, have far more general meaning than others and therefore can occur in a greater number of constructions: compare for example *go* with more specific manner motion verbs in English, such as *walk*. Specific verbs seem to be closely associated with specific constructions, and *vice versa* (Kidd, Lieven & Tomasello 2006; Thompson 2002), especially when the lexical repertoire is quite limited, as in the case of early learners (Ellis & Ferreira–Junior 2009).

Skewed input has been shown to be common in input, but testing its effect on the acquisition outcome implies comparing learner performance under different input skewedness conditions (Borovsky & Elman 2006; Casenhiser & Goldberg 2005 for L1 research). When this is done, its beneficial effects are not completely clear. Year and Gordon (2009) investigate the effect of a frequent, prototypical ditransitive verb on the acquisition of the English ditransitive construction in Korean children learning English in a classroom setting. The target structure exhibits strong semantic constraints, as it requires an animate beneficiary: it is hypothesised that skewed input might highlight such constraints and facilitate target like acquisition, which accordingly was tested with both a production and a grammaticality judgement task. No statistically significant differences were found between the learners receiving skewed and standard input. Instead, the results suggest that construction learning in this situation is superior with a more balanced set of verbs rather than a focus on a single prototype. Wulff et al. (2009) also examine the role of frequency distributions, reliabilities of form–function mapping, and prototypicality of lexical aspect on the acquisition of tense-aspect morphology. The authors' input analysis was matched against acquisition data by adult learners of English (Bardovi-Harlig 2000). As is often the case, the results did not point to a single, univocal predictor of acquisition success, but rather suggest that all parameters considered jointly drive acquisition.

In sum, the issue of whether high or low variance is beneficial to acquisition is not settled. It may well be that the effects of this factor interact with others, such as the type of construction targeted or the learner's proficiency in the L2. In what follows we will analyse the type composition of the VILLA in order to verify whether or not it is skewed in favour of specific lexical types. Specifically, we ask if the two functions of subject and object tend to be instantiated by only a limited number of lexical items, or if on the contrary the two functions exhibit high type frequency, without forgetting, of course, that the VILLA lexical repertoire is intrinsically quite limited. We do so in order to find possible explanation for the apparently random variability in morphosyntactic accuracy which was observed in the output produced by learners in their semi-spontaneous interaction. As the accusative marking is in the focus of our interest, we will start with the object function and then proceed to the subject function and to verbs. In the analysis which follows, the figures are extracted from full transitive sentences in which subject, verb and objects are realised explicitly, which is not obvious since Polish is a pro-drop language. Words appearing in elliptic utterances are therefore excluded. For a detailed discussion of the sentence models investigated, please see section 6.2.3, p. 223.

Graph 25 shows the relative frequency of the lexical items instantiating the object function at T1 (9 hours).

Graph 25⁹⁰



The picture is clearly dominated by a single item, *język* "language". Its popularity is due to two main facts. The first is that nationalities were a key topic in the VILLA course: they commonly occur in copular structures, to which in turn were devoted some of the main research questions of the experiment. As a result, all characters are associated to a nationality, and consequently to a language (48).

(48) *Ewa jest Polką i zna język polski.*
 Ewa-NOM is Polish-INS and knows language.ACC
 Polish.ACC

⁹⁰ This graph was created using the statistical software R (R Core team 2014) and the packages *wordcloud* (Fellows 2014) and *extrafont* (Chang 2014). In this and the following two graphs, the horizontal axis has no numerical meaning and is only used to graphically present the lexemes.

"Ewa is Polish and speaks Polish"

Moreover, most of the course characters are exceptional polyglots, and speak several languages (49):

- (49) *Julia zna język polski język angielski i język włoski.*
Julia knows language.ACC Polish.ACC language.ACC English.ACC and language.ACC Italian.ACC
"Julia speaks Polish, English and Italian"

No wonder then that this lexical item should be so common. So common, perhaps, as to be treated as an unanalysed chunk in learner output, or perhaps as a verb, as in (50):

- (50) *kim sna jeski kristina?*⁹¹
Who.INS knows language.ACC Cristina-NOM⁹²
"What language(s) does Cristina speak?"

This lone lexical item is then followed by several other words with intermediate frequency, which differ from each other in terms of gender and animacy, a question which will be analysed in detail in section 6.2.3, p. 223. We finally encounter a great number of low-frequency words. Still, it appears that in spite of the extraordinary diffusion of the word *język*, the object function exhibits a comparatively high type frequency, which caused learners to be exposed to a significant number of different lexical items. These, in turn, are represented by different inflectional endings on the basis of gender and animacy (Table 57).

⁹¹ Example from Saturno (in prep).

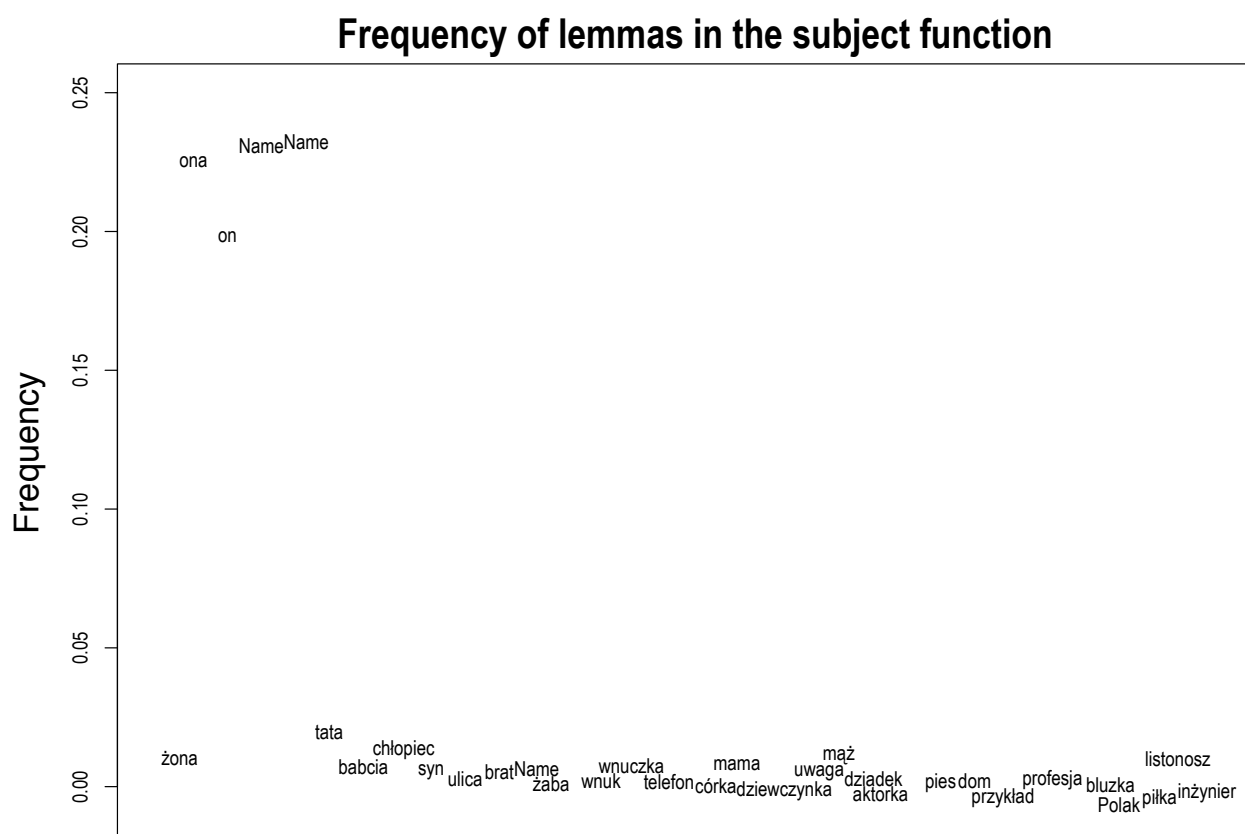
⁹² Morphological marking may look inappropriate, since we are considering the output of an initial learners. Their only function is to signal the most similar word-form of the target language.

Table 57:accusative and nominative marking by animacy and gender

	ANIMATE		INANIMATE	
	ACC	NOM	ACC	NOM
MASCULINE	<i>-/a/: kot-a</i>	<i>kot</i> "cat"	<i>-/∅/: dom</i>	<i>dom</i> "home"
FEMININE	<i>-/e/: mam-ę</i>	<i>mama</i> "mum"	<i>-/e/: kaw-ę</i>	<i>kawa</i> "coffee"

The relatively high type frequency of the object function becomes quite evident by comparing it to the data relative to the subject function, for which the picture is not as varied (Graph 26).

Graph 26⁹³



⁹³ This graph was created using the statistical software R (R Core team 2014) and the packages *wordcloud* (Fellows 2014) and *extrafont* (Chang 2014).

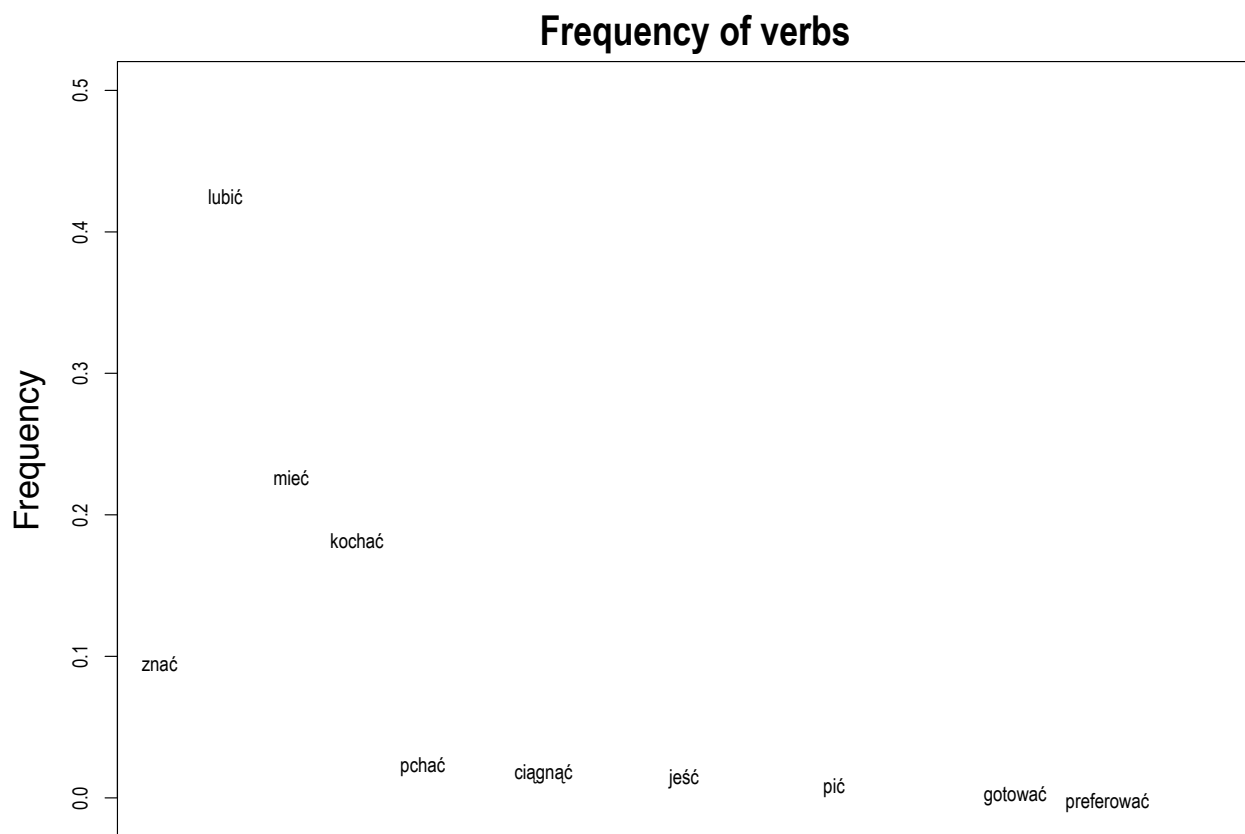
The distribution is dominated by only four types, namely the two personal pronouns in the nominative case, *on* and *ona*, respectively "he" and "she", together with masculine and feminine person names. All other nouns, among which the common nouns employed in the structured tests discussed in this work, account for a proportion of the input which is close to 0. Indeed, this trend will become apparent in the semantic analysis provided in section 6.2.3, p. 223. Curiously, what the four most frequent types have in common is that they do not represent common nouns like those which appear in the structured tests. It is interesting to note, therefore, that the nominative case is never processed incorrectly, which witnesses to the fact that the basic word-form of a lexeme is modelled on the nominative case in a manner which is to a certain extent independent of the frequency of its various word-forms in the input. Further, almost all nouns in the graph are animate nouns, which is hardly surprising. The few exceptions only occur in highly specific, infrequent constructions (51).

(51) *jaki* *kolor* *ma* *twoja* *bluzka?*
 which.ACC colour.ACC has your-NOM blouse-NOM?
 "what colour is your blouse"

In sum, the subject function is almost exclusively instantiated by animate nouns, which is not unexpected, and exhibits very low type frequency, so much so that the distribution is not even compatible with Zipf's law.

We finally turn to the type frequency of verbs (Graph 27).

Graph 27⁹⁴



The graph shows a very clear-cut picture, decidedly dominated by *lubić*, "to like", which alone accounts for more than 40% of all verbs found in transitive sentences. It is directly followed by *mieć*, "to have", and *kochać*, "to love", with a frequency of about 20% each. Just below 10% we then find *znać*, "to know". Finally comes a series of six verbs whose frequency is comprised between 0 and 5%, thus accounting for only a minimal proportion of the verbs encountered in this type of sentences. As a whole, the pattern reminds us of Zipf's law.

It is interesting to note that while *język* exhibited the highest type frequency among all nouns occurring in the object (Graph 25), the verb exclusively associated with it, *znać* "to know", only accounts for 10% of the verbs. This is because *lubić* associates with numerous nouns individually characterised by a

⁹⁴ This graph was created using the statistical software R (R Core team 2014) and the packages *wordcloud* (Fellows 2014) and *extrafont* (Chang 2014).

lower type frequency, but which together represent the bulk of transitive sentences. Further, as shown in example (49) on p. 218 above, the word *język* may be repeated several times in the same utterance, whereas *znać* typically occurs only once per utterance.

The verbs *lubić* "to like" and *kochać* "to love", in spite of their similar meaning, correspond to very different frequency values. We signal that while the former mainly associates with inanimate objects, the latter is used almost exclusively with human nouns and person names, another clear hint to the close relation between semantics and syntactic function.

The frequency of *lubić* "to like", *mieć* "to have" and *kochać* "to love" is accounted for by the variety of lexical items which can function as their object, that of *znać* "to know" by its frequent use to characterise course characters in terms of the languages they speak. In contrast, most of the lower-frequency verbs are only used in very specific contexts. First, most verbs, like *gotować* "to cook" in example (52), were often used intransitively when introducing and practising the new lexeme. When used in this manner, they were not included in the calculations presented above, which only consider transitive sentences.

- (52) *TEA: *a on co on robi Gina?*
 and he what he does Gina (student name)?
 "and him, Gina, what is he doing?"
- *TEA: *gotuje super on gotuje bardzo dobrze.*
 Cooks super he cooks very good
 "he cooks, super; he cooks, very good"

Second, verbs like *ciągnąć* "pull" and *pchać* "push" occurs with virtually a single object, *wózek* "cart", as a result of the situations treated throughout the course (53):

- (53) a. *Filip* *ciągnie* *wózek.*
 Filip.NOM pulls cart.ACC
 "Filip pulls the cart"
- b. *ona* *пча* *wózek.*
 she.NOM pushes cart.ACC
 "She pushes the cart"

This said, let us summarise our analysis of the VILLA input in terms of type frequency:

- the object function exhibits relatively high type frequency;
- the distribution of verbs is reminiscent of Zipf's law, whereby type frequency is inversely proportional to frequency rank;
- the subject function is almost exclusively instantiated by only four macro-types (personal pronouns and person names).

6.2.3. *Token frequency*

Token frequency regards the number of times a given form or structure occurs in the input. Though certainly intuitive, its effects are nowadays considered to be overestimated and more precisely attributable to other factors, from type frequency to semantic scope and others (see Ellis (2002) for a review). This parameter is central to the methodology of the VILLA project, however, with particular regard to the frequency of both lexical items and syntactic structures: accordingly, in this paragraph we will be concerned with the frequency of SO and OS transitive structures. We will not limit ourselves to a mere measure of frequency, though, but will take into account several grammatical and semantic properties of the lexical items which instantiate our target structures. The resulting measure should be relevant in terms of both token frequency and semantic scope.

Before proceeding with the quantitative analysis, we will briefly present the parameters on which we are going to focus.

In the simplified VILLA input, transitive structures are composed of two nominals (54a) or groups of nominals (54b), inflected in the nominative and accusative form, as well as a verb, typically positioned between the nominals.

- (54) a. *Juli-a ma lalk-ę*
Julia-NOM has doll-ACC
"Julia has a doll"
- b. *dobrze Juli-ę kocha brat Filip.*
good Julia-ACC loves brother.NOM Filip.NOM
"good, Julia loves (her) brother Filip"

The first parameter which we will use to classify such structures is the order in which the two noun phrases occur in the utterance, i.e. constituent order. For our purposes, we only consider two values, namely SO (54a) and OS (54b)⁹⁵.

Second, transitive structures by definition require a bivalent verb. Given the very limited verb set of the VILLA project and their relative uniformity, for our purposes it will be sufficient to distinguish between bivalent verbs and all others, typically copulas and monovalent verbs.

Nominals require a more detailed analysis. To start, they may be represented by either pronouns or nouns: as the former only appear in their nominative form, they are limited to the subject function only. Regarding nouns, we distinguish two sub-classes, namely person names, like *Andrzej*, and common nouns. While the former always refer to animate entities (most commonly human), common nouns may refer to about anything, including people, animals, things, as well as abstract ideas. We therefore need yet another subdivision, this time according to the semantic parameter of animacy: we distinguish animate nouns, referring to people and animals, and inanimate ones,

⁹⁵ Note however that slight differential processing effects were observed in the comprehension test between OSV and OVS structures (see section 3.4.1, p. 88). We will not pursue the question further here, as OSV structures are extremely rare in the input.

referring to things. We need not concern ourselves with abstract ideas because no such word is comprised in the VILLA input.

This taxonomy of nominals in transitive sentences seems necessary to determine whether the frequency of the target structures of our test, and indeed, of transitive sentences in general, is determined by any of the lexical (gender) and semantic (animacy, common nouns vs. person names) parameters discussed above.

The rationale for this analysis is that from the learner point of view, a particularly strong link between a certain kind of nominal and transitive sentence may have significant implications for the establishing of form-function associations. In a heavily inflected language like Polish, nominals take on different forms depending on their grammatical function: in what follows we will attempt to verify whether grammatical functions in turn are influenced by semantics. If so, then we may simplify the logical argument above by saying that semantics influences the form in which nominals appear. We have seen in the preceding sections that this is indeed the case: the most common object types are represented by inanimate nouns, and conversely the subject function is an almost exclusive dominion of pronouns and person names. In this section we will attempt to combine this information and perform a similar analysis for full-fledged structures, in which verb, subject and object occur together. We will thus be able to ask such questions as "does the semantics of the subject influence that of the object, and vice versa?"

Table 58 below shows the possible alignment patterns of grammatical category, animacy, gender and case for any given nominal.

Table 58: combinations of grammatical category, animacy, gender and case

	COMMON NOUN	PERSON NAME	PRONOUN
Animacy	ANIMATE/INANIMATE	ANIMATE	ANIMATE
Gender	MASC/FEM	MASC/FEM	MASC/FEM
Case	NOM/ACC	NOM/ACC	NOM

As can be seen, the widest range of values is available for common nouns, while pronouns may only vary in their gender, which is lexically determined and does not vary contextually. Case, in contrast, is determined by the syntactic context. Common nouns and person names occur in both the nominative and accusative case, whereas pronouns were only presented in their nominative form. Regarding animacy, person names and pronouns always refer to animate (human) entities, while common nouns may indicate both animate and inanimate referents. Table 59 presents examples for all combinations.

Table 59: combinations of grammatical category, animacy, gender and case: examples

GRAMMATICAL CLASS	GENDER	ANIMACY	NOM	ACC
Pronoun	MASC	ANIMATE	<i>on</i>	-
Pronoun	FEM	ANIMATE	<i>ona</i>	-
Name	MASC	ANIMATE	<i>Karol</i>	<i>Karola</i>
Name	FEM	ANIMATE	<i>Sabina</i>	<i>Sabinę</i>
common noun	MASC	ANIMATE	<i>strażak</i>	<i>strażaka</i>
common noun	FEM	ANIMATE	<i>kucharka</i>	<i>kucharke</i>
common noun	MASC	INANIMATE	<i>samochód</i>	<i>samochód</i>
common noun	FEM	INANIMATE	<i>piłka</i>	<i>piłkę</i>

We searched for all grammatically possible combinations of the five parameters listed in Table 53, organised in both SO and OS structures. This resulted in a total of 96 combinations. Table 60 and Table 61 provide a schematic view of the distribution of SO and OS patterns, respectively. For each cell, the table shows the relative frequency across the input of all five editions as observed at T2. A complete table providing information as to all patterns for which at least one example was found across the entire corpus is presented on p. 260. The underlined figures are relative to the ten most frequent patterns, whereas those in shaded cells refer to the patterns encountered in the structured tests discussed in this work, which will be discussed in detail in section 6.3, p. 231.

The number between brackets refers to the corresponding example, discussed below in the text.

Table 60: transitive sentence patterns, SO

				NP2 (Acc)					
				Animate				Inanimate	
				name		Noun		Noun	
				M	F	M	F	M	F
NP 1 (Nom)	Animate	Name	M					6% (56b)	4% (61)
			F		4% (60)		7% (56a)	3% (62)	
		PN	M				11% (55a)	5% (57)	
			F		5% (58)		11% (55b)	4% (59)	
	Noun	M			2% (73a)				
		F		2% (60a)	-				
	inanimate	Noun	M						
			F						

Table 61: transitive sentence patterns, OS

				NP2 (Nom)							
				Animate				Inanimate			
				name		PN		Noun		Noun	
				M	F	M	F	M	F	M	F
NP 1 (Acc)	Animate	name	M								
			F								
		noun	M					0.5% (73c)			
			F				0.2% (73d)	-			
	inanimate	noun	M								
			F								

Several observations are possible. First, it is quite obvious that SVO structures are disproportionately more common than their OVS equivalent. Second, the

targets of the structured tests (shaded cells) concentrate only in a restricted area of the table, which witnesses to their sharing numerous semantic traits. The same is true for the most common structures, which however do not coincide with the targets of the structured tests.

Below we present a brief overview of the ten most frequent constructions, all of which exhibit SO word order.

By far the most frequent types present the feminine (55a) or masculine (55b) personal pronouns in the subject function, coupled with a masculine inanimate noun in the object function.

- (55) a. *ona ma rower.*
 she has bike.ACC
 "she has a bike"
- b. *on zna język norweski.*
 he knows language.ACC Norwegian.ACC
 "he speaks Norwegian"

The following patterns are nearly identical, the only difference being that instead of personal pronouns, the subject function is instantiated by feminine (56a) or masculine (56b) person names.

- (56) a. *Ewa ma duży niebieski samochód.*
 Ewa-NOM has big.ACC blue.ACC car.ACC
 "Ewa has a big blue car"
- b. *czy Filip pcha wózek?*
 INTERR Filip.NOM pushes cart.ACC
 "is Filip pushing the cart?"

Next, we find sentences in which masculine (57a) or feminine (57b) personal pronouns perform the subject function, while the object is represented by inanimate feminine (57a) or animate masculine (57b) nouns.

- (57) a. *on bardzo lubi muzykę.*
 he very likes music-ACC
 "he really likes music"
- b. *ona ma brata.*
 she has brother-ACC
 "she has a brother"

Note that the *-/a/* ending corresponding to the accusative case of masculine animate nouns also instantiates the genitive case within the same paradigm. This in turn is the default case of direct objects under the scope of negation (58a). In the feminine paradigm, on the contrary, the genitive (58b) and the accusative endings are clearly distinct, so that direct objects are marked by different case endings depending on whether or not their verb is negated. Finally, we point out that animate nouns in Polish include not only human nouns, but also animals (58c).

- (58) a. *nie ona nie kocha dziadka Karola.*
 no she not loves grandpa-GEN Karol-GEN
 "no she doesn't love grandpa Karol"
- b. *on nie lubi muzyki.*
 he not likes music-GEN
 "he doesn't like music"
- c. *ona lubi kota tak.*
 She likes cat-ACC yes.
 "she likes (her) cat yes"

Moving further, we find sentences with feminine personal pronouns as subjects and inanimate feminine nouns as objects (59).

- (59) *tak ona ma lalk-ę.*
 yes she has doll-ACC
 "yes she has a doll"

The next type is instantiated by animate feminine nouns as subjects and animate masculine nouns as object (60a). We signal here that some animate masculine nouns are inflected following the feminine declension. The fairly common noun *tata*, "dad", is a good example (60b). In view of their lack of alignment between inflectional paradigm and grammatical gender (let alone semantic sex), these idiosyncratic lexical items may well represent some difficulty for learners as they try to establish their first form-function associations. Note that among the ten most common patterns found in the input, this is the only one which is directly represented in the tests.

- (60) a. *czy Juli-a ma brat-a?*
 INTERR Julia-NOM has brother-ACC
 "does Julia have a brother?"
- b. *córk-a Juli-a kocha tat-ę.*
 Daughter-NOM Julia-NOM loves dad-ACC
 "the daughter Julia loves (her) dad"

The following pattern is nearly identical, except that the subject is masculine and the object is feminine (61).

- (61) *Leon lubi herbat-ę.*
 Leon.NOM likes tea-ACC
 "Leon likes tea"

Finally, yet another variant presents animate feminine nouns as subjects and inanimate feminine nouns as objects (62). Note, however, that common nouns

such as *babcia* "grandma" are often accompanied by person names like *Maria*, too.

- (62) *tak* *babci-a* *Mari-a* *lubi* *kuchni-ę* *włosk-ą*.
 Yes grandma-NOM Maria-NOM likes cuisine-ACC Italian-ACC
 "yes grandma Maria likes Italian cuisine"

6.3. TEST TARGET STRUCTURES

We mentioned in the preceding section that the patterns encountered in the target sentences of the repetition and comprehension tests hardly occur in the input, at least in terms of the syntactic and semantic parameters we focussed on. In this paragraph we will look into this question in somewhat greater detail. Specifically, we will manipulate the parameters according to which we classified nominals - word class, animacy and gender - to identify those which have the greater impact on token frequency in the input.

6.3.1. Repetition test

We remind the reader that both the subject and the object of target sentences in the Repetition test are instantiated by animate feminine common nouns, as exemplified in (63).

- (63) a. *dziewczyn-a* *ciągnie* *portugalk-ę*
 girl-NOM pulls Portuguese-ACC
 "the girl pulls the Portuguese woman"
- b. *portugalk-ę* *ciągnie* *dziewczyn-a*
 Portuguese-ACC pulls girl-NOM
 "the girl pulls the Portuguese woman"

Remarkably, such structures are completely absent from the input in both their SO and OS form. There exists, however, a construction which is formally identical to the targets of the test, although different semantically. In (64), both the subject and the object are instantiated by nouns inflected according to the feminine paradigm, namely *córka*, "daughter" and *tata*, "Dad". Unfortunately, the latter is semantically masculine.

- (64) *córk-a* *Juli-a* *kocha tat-ę*.
 Daughter-NOM Julia-NOM loves dad-ACC
 "(the) daughter Julia loves (her) father"

While masculine nouns present different accusative marking depending on their animacy, this is not true of the feminine nouns in *-a* comprised in the VILLA input, whose accusative is marked by *-e* irrespectively of animacy. Person names also inflect according to the same paradigm (65).

- (65) a. *matk-a* mother-NOM; *matk-ę* mother-ACC
 b. *ryb-a* fish-NOM; *ryb-ę* fish-ACC
 c. *Mart-a*, Marta-NOM; *Mart-ę*, Marta-ACC

This said, we start by manipulating the semantic values of the object. We first relent our constraint on object animacy, which seems a legitimate operation thanks to the identical accusative marking just referred to. By so doing we obtain a proportion of 0.4%⁹⁶ for SO sentences, but still no OS example (66).

- (66) *babci-a* *lubi* *literatur-ę*
 grandmother-NOM likes literature-ACC
 "grandma likes literature"

⁹⁶ All proportions are computed on the total number of transitive sentences across all editions of the VILLA project, namely 1817. See the table on p. 273 for details on the absolute values.

Adding names in the object function results in a proportion of 0.7% for SO structures (67a) and 0.5% for OS ones (67b) hits. In these few examples the common nouns are always accompanied by a person name, which further differentiates these utterances from the target sentences of the repetition test.

- (67) a. *mam-a* *Ew-a* *kocha Juli-ę*
 mum-NOM Ewa-NOM loves Julia-ACC
 "mum Ewa loves Julia"
- b. *Juli-ę* *kocha* *mam-a* *Ew-a*
 Julia-ACC loves mum-NOM Ewa-NOM
 "mum Ewa loves Julia"

Only if we further include names in the subject function does relative frequency rise to 5.4%⁹⁷ for SO and 3.5% for OS (68). Note that while common nouns are always accompanied by a person name, as pointed out above, the reverse is not true.

- (68) a. *Juli-a* *ma* *pilk-ę.*
 Julia-NOM has ball-ACC
 "Julia has a ball"
- b. *lalk-ę* *ma* *Juli-a.*
 doll-ACC has Julia-NOM
 "Julia has a doll"

Finally, adding pronouns in the subject function results in a relative frequency of 10.2% for SO sentences, while OS figures do not vary⁹⁸ (69).

⁹⁷ This value should be compared to the frequency of the most common structure in the input, i.e. 11.3%.

⁹⁸ This means that pronouns never occur in post-verbal position: utterances like *herbat-ę lubi ona*, tea-ACC likes she "it is she who likes tea", though possible in native Polish, never occur in the VILLA corpus.

- (69) a. *ona lubi herbat-ę*
 she likes tea-ACC
 "she likes tea"

It is evident that pronouns and person names represent the lion's share as far as the expression of the subject function is concerned, while objects are mainly instantiated by inanimate nouns. This is not surprising if one considers the topics treated throughout the course as well as the typical teacher-learner interactional dynamics. First, pronouns are commonly used to refer to entities which have been previously introduced using a person name, as is the case with *ona* and *Julia* in example (70).

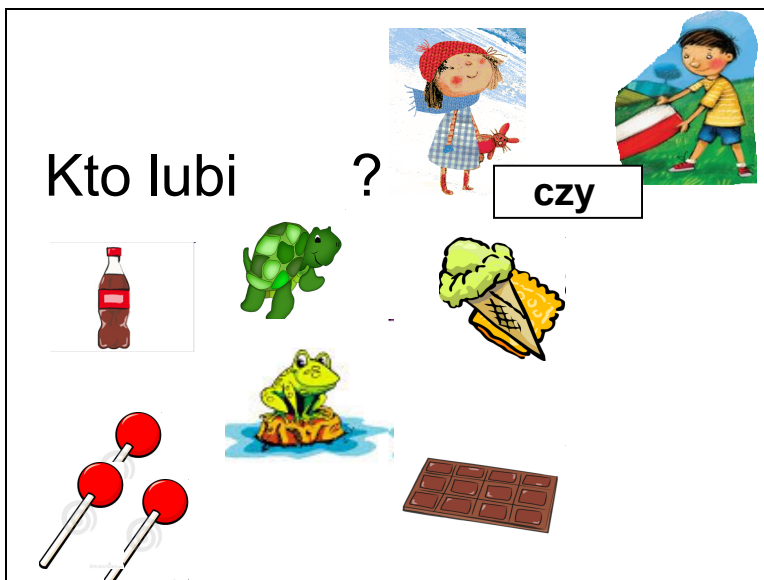
- (70) *TEA: *Juli-a lubi żółwi-a.*
 Julia-NOM likes turtle-ACC
 "Julia likes the turtle"
- *TEA: *ona lubi lizaki lubi lizaki.*
 She likes lollipops.ACC likes lollipops.ACC
 "she likes lollipops"
- *TEA: *ona lubi żółwi-a i lubi czekolad-ę.*
 She likes turtle-ACC and likes chocolate-ACC
 "she likes turtle and chocolate"

Further, it can be shown that the same entities are mainly introduced using person names rather than common nouns for reasons related to discourse. Consider example (71):

- (71) *TEA: *kto lubi czekolad-ę?*
 Who likes chocolate-ACC?
 "who likes chocolate?"
- *TEA: *eh Gaston?*
- *TEA: *tak Juli-a lubi czekolad-ę.*
 Yes Julia-NOM likes chocolate-ACC
 "yes Julia likes chocolate"
- *TEA: *czekolad-ę lubi Juli-a.*
 Chocolate-ACC likes Julia-NOM
 "Julia likes chocolate"

This sequence occurs in a context in which the class is working on the powerpoint slide in Figure 9, trying to decide what course character likes or owns each of the objects depicted thereby.

Figure 9



Based on information provided previously during the lesson, they can decide between Julia and Filip (top right), both well known course characters. It follows from the structure of this communicative context that the objects represent the discourse topic, while the two children are in the focus position.

In the example, the teacher first asks a specific learner who likes chocolate. The input transcription available at this stage only concerns the teacher's speech and does not comprise the learner's response, but only teacher feedback to it⁹⁹. Judging on the teacher's third turn, the learner's answer must have been appropriate, though perhaps only semantically; in any case, the teacher repeats the response, if we assume it to be grammatically correct, or recasts it, in the opposite case. This utterance is what interests us the most. Even though its topic, *czekoladę* "chocolate-ACC" performs the object function, thus licensing syntactically marked word order, the native speaker at first prefers to produce a syntactically unmarked SO sentence, in which pragmatic markedness is expressed prosodically through the stressing of *Julia*, in initial position, which highlights her as the sentence focus. Only later, and perhaps for experimental reasons only, does the teacher produce the equivalent OS utterance.

This example is precious to understand two important points. The first concerns the relative rarity of OS sentences compared to SO ones: not only are they marked, but their purpose can be easily (and perhaps, preferably) fulfilled by other strategies to mark departures from the default alignment between the syntactic and pragmatic structure of the utterance (topic-subject; focus-object).

On the other hand, example (71) quote above should make it clear why person names should be so much more frequent than common nouns in transitive structures. Teacher speech is mainly based on powerpoint slides depicting the same characters over and over again. Thus, even if each of them is identified by a particular nationality and profession, expressed in turn by common nouns, the characters depicted become so familiar to the learners (and to the teacher) that it would seem unnatural to refer to them otherwise than by their name, for instance by saying *dziewczynka* "little girl" instead of just *Julia*. In contrast, the target sentences of the repetition test required learners to process common

⁹⁹ See section (5.2, p. 122), however, for a discussion of the transcription of learner speech during classes.

nouns in the absence of any context, something which they could arguably be ill-equipped to do at such early stage of acquisition.

6.3.2. *Comprehension test input patterns*

The target sentences of the comprehension test belong to four patterns, exemplified in (72). Only two nouns are used throughout the test, namely *brat*, "brother" and *siostra*, "sister", which alternatively perform the functions of subject and object in both SO and OS structures.

- (72) a. *brat* *wołá* *siostr-ę*
 brother.NOM calls sister-ACC
- b. *siostr-ę* *wołá* *brat*
 sister-ACC calls brat.NOM
- c. *siostr-a* *wołá* *brat-a*
 sister-NOM calls brother-ACC
- d. *brat-a* *wołá* *siostr-a*
 brother-ACC calls sister-NOM

As pointed out during the brief overview of the most common input patterns, the structures listed above are extremely rare. Most of them occur only a handful of times and not necessarily in all editions of the VILLA project¹⁰⁰. Not surprisingly, OS structures (73c and 73d) are even rarer than their SO equivalents (73a and 73b).

¹⁰⁰ We remind the reader that all the input figures provided in this chapter are computed on the cumulative input of all five MB editions at T2 (13:30 hours).

- (73) a. *mój brat ma żon-ę* (1.7%)
 my brother-NOM has wife-ACC
 "my brother has a wife"
- b. *i żon-a kocha męż-a.* (1,7%)
 and wife-NOM loves husband-ACC
 "and (the) wife loves (her) husband"
- c. *dziadk-a Karola kocha wnuczk-a* (0.5%)
 grandpa-ACC Karol-ACC loves niece-NOM
 "the niece Julia loves grandpa Karol"
- d. *babci-ę Ann-ę kocha dziadek Karol.* (0.2%)
 grandma-ACC Anna-ACC loves grandpa.NOM Karol.NOM
 "Grandpa Karol loves grandma Anna"

Again, we are going to progressively give up the constraints regarding nouns in the object and subject functions, in order to identify the parameters which affect token frequency the most. We start with structures with a masculine object, like *siostra woła brata*, "the sister calls (her) brother".

Ignoring the animacy contrast is more complicated here than it was in previous sub-paragraph, in which we showed that this operation has no consequence on the morphological marking of feminine nouns. When dealing with masculine nouns, instead, we need to take into account the fact that inanimate items, unlike the word *brat* "brother" employed in the test, do not distinguish between the nominative and accusative case. As a consequence, sentences which only differ with regard to the animacy of the object exhibit a rather different structure (74):

- (74) a. *siostr-a wola brat-a*
 sister-NOM calls brother-ACC
 "(the) sister calls (her) brother"
- b. *Julia ma czerwony balonik.*
 Julia-NOM has red.ACC balloon.ACC
 "Julia has a red balloon"

For this reason, we are not going to manipulate the animacy parameters of masculine objects. Instead, we first allow the object function to be instantiated by a person name, but it turns out that no example following this new pattern can be found.

However, if we allow the subject to be instantiated by a person name, too, proportions increase to 5.8% for SO (75a) and 1.7% for OS (75b).

- (75) a. *tak Ann-a kocha Karol-a.*
 Yes Anna-NOM loves Karol-ACC
 "yes Anna loves Karol"
- b. *i Karol-a kocha Juli-a*
 and Karol-ACC loves Julia-NOM
 "and Julia loves Karol"

If we further include subject pronouns, the relative frequency of SO targets rises to 10.8% (76), while the OS word order remains virtually unaffected.

- (76) *ona kocha Karol-a tak*
 she loves Karol-ACC yes
 "she loves Karol yes"

In sum, it appears that as far as masculine animate objects are concerned, the most common structures include a person name or a pronoun in the subject function, unlike in the test structure.

We now apply the same analysis to structures with feminine objects and masculine subjects, like *brat woła siostrę*, "the brother calls (his) sister". The frequency of the original test sentence is 1.7% for SO and only 0.2% for OS ones. Allowing for person names in the object function modifies the values to 2.7% for SO (77a) and 1.2% for OS (77b).

- (77) a. *tata* *kocha* *Julię*
 dad.NOM loves Julia-ACC
 "dad loves Julia"
- b. *Julię* *kocha* *brat*
 Julia-ACC loves brother.NOM
 "(the) brother loves Julia"

It must be said that as far as the lexical items involved are concerned, most of these utterances are identical to the examples presented above.

Allowing for person names in the subject function raises the frequency count to 7.2% for SO (78a) and 3.4% for OS (78a).

- (78) a. *ale tutaj Filip* *kocha Julię.*
 but here Filip.NOM loves Julia-ACC
 "but here Filip loves Julia"
- b. *Julię* *kocha Leon.*
 Julia-ACC loves Leon.NOM
 "Leon loves Julia"

Finally, adding subject pronouns produces a final figure of 9.3% for SO structures (79). Just like we observed above with regard to feminine pronouns, these elements hardly ever occur in post-verbal position, so that the frequency of OS sentences remains unaffected.

(79) *on kocha Ewę*
he loves Ewa-ACC
"he loves Ewa"

In sum it appears that again, the sentences in which human nouns occur in the object function mostly present a pronoun or a person name as their subject. This is in contrast with the target elements of the structured tests. These, therefore, may be thought to probe the learners' ability to generalise a morphosyntactic pattern to nominals characterised by different semantic traits.

6.4. CONCLUSION

In this chapter we have attempted to apply a quantitative methodology to the controlled input of the VILLA project in order to identify regularities which may at least partially explain the results obtained in the structured tests and in the learners' spontaneous output. To this purpose, we first introduce the system of automatic morphological glossing on which our methodology is based. We then used it to investigate three input parameters which appeared particularly relevant to our work, namely form-function association, type frequency, and token frequency.

We first quantified the strength of the form-function association between case endings and the corresponding syntactic functions, using methodological tools derived from the theoretical frameworks of Natural Morphology and the Competition model. The findings show that indeed, the association between the ending *-/a/* and the subject function is stronger than that between the ending *-/e/* and the object function. This might play a role in justifying why learners hardly ever process the nominative case incorrectly, while errors concerning the accusative case are quite common. However, we suspect this explains learner behaviour only partially, if at all: the nominative case of feminine nouns may be favoured by a stronger form-function association, but what really matters is that it is selected as the basic word form of lexical items. The

criteria which affect the choice of a particular form of a paradigm are not entirely clear to date: while one may intuitively call upon frequency, other factors are probably involved. Among these, we believe the citation form is particularly important. To exemplify, *kuchnia* "cuisine" is a noun which due to its semantics tends to occur in the accusative case. Yet, its basic word form is modelled on the nominative case. In fact, a qualitative analysis of the input shows that whenever the word is used out of context, it occurs precisely in that case. In example (80), the teacher first uses the noun in the accusative case, then asks the class to repeat it aloud in the nominative.

- (80) *TEA: *ona lubi kuchni-ę włosk-q.*
 she likes cuisine-ACC Italian-ACC
 "she likes Italian cuisine"
- *TEA: *proszę mówić kuchnia włoska.*
 please say cuisine-NOM Italian-NOM
 "please say Italian cuisine"

Further, we turned to type frequency, namely the number of lexical items which instantiate a given construction. We computed different values relative to subjects, objects, and verbs.

Not surprisingly, the analysis shows that the subject and object functions are mainly represented by animate and inanimate nouns, respectively. While the object function was characterised by relative high type frequency, though, the subject function is instantiated by only four macro-types, namely the two personal pronouns *on* and *ona*, "he" and "she", and masculine and feminine person names. While the subject case is not in the focus of our analysis and is usually processed correctly, the high type frequency of the object function might explain why some learners managed to correctly inflect in the accusative case even nouns which never appeared in that form in the input, like *rodzina* "family". Since the VILLA project was not designed to investigate this particular research question, we can hardly pursue it experimentally based on the present

data. These insights do represent a further piece of a complex mosaic, however, which could be usefully exploited in the future to contribute to the lively debate presented in the literature.

We then turned to token frequency, that is, the mere number of times a given item or construction occurs in the input. We applied this analysis to full transitive sentence, composed of two nominals, performing the subject and object functions respectively, as well as a bivalent verb. We scanned the input for 96 sentence models, described in terms of the various combinations of the following parameters:

- word order: SO vs. OS;
- subject word class: personal pronoun, person name, common noun;
- object word class: person name, common noun;
- subject and object gender: masculine vs. feminine;
- animacy (animate vs. inanimate).

Most of the resulting models were completely absent from the input. We presented examples from the 10 most frequent structures, which confirm the trends highlighted by the analysis of type frequency: the subject tends to be instantiated by personal pronouns or person names, while the object shows a privileged association with inanimate nouns. Further, semantic and syntactic cues, such as animacy and the unmarked SO word order, are almost always available to disambiguate sentences which may sound potentially ambiguous with respect to morphology. Utterances cannot be processed on the basis of semantics only when they include two nouns which do not differ in animacy, as is the case in the structured tests. Such utterances, however, are rare or absent altogether from the VILLA corpus. Moreover, in most cases ambiguity can easily be resolved on the basis of unmarked SO word order or common sense. In sum, inflectional morphology, while a characteristic and obligatory trait of the target language, is by no means necessary to interpret transitive sentences. No wonder, then, that learners can easily do without it and still

communicate effectively. One should also add the effects of context, which offers perhaps the most effective clues to disambiguate potentially ambiguous sentences.

The models corresponding to the target sentences of the structured tests were either absent or extremely rare, precisely because they did not fit in the dominant trends referred to above. The learner free production, in contrast, is in full accordance with them. It would be hard, and perhaps inappropriate to distinguish between discourse context, semantics, and input processing in accounting for such behaviour: the three could be better seen as different aspects of a single phenomenon. Specifically, the particular associations of semantic and syntactic categories referred to above are privileged in the input simply because of common sense, as animate nouns are more likely to be the controller of a situation, while inanimate entities are more likely to be themes. Second, in a context in which the same characters are mentioned over and over again, it is appropriate to refer to them using personal pronouns or their names, rather than a common noun. In fact, since the VILLA learners are all adult, competent speakers in at least one language, we could argue that their experience in terms of pragmatics and world knowledge may sometimes prevail on the input received: they often choose to ignore input patterns, in spite of their high frequency, and develop their own structures instead. While presentative structures with *to* "this" are extremely common in the input, for example, they appear to be disfavoured both in structured tests (Saturno 2015) and in spontaneous interaction. In the latter context, learners creatively elaborate on structures available in the input, like "his/her name is" (81), or invented new, ungrammatical constructions (82). It cannot be excluded that the source language played a role in their shaping, a question which could be pursued in the future through the qualitative analysis of free production elicited in the various VILLA editions.

- (81) *STU: *ona ma na imje lutfia.*¹⁰¹
 She has on name Lucia
 "her name is Lucia"
- *STU: *ona jest wo wo wojsko wojskon i artiskom.*
 She is Italian and artist
 "she is Italian and an artist"
- (82) *STU: *ona jest dzovanna.*¹⁰²
 She is Giovanna
 "she is Giovanna"
- *STU: *dzovanna est nautfitfelkon.*
 Giovanna is teacher
 "Giovanna is a teacher"

These last examples are particularly precious to understand that input, while surely an important and necessary factor, does not explain everything in SLA. Even in the VILLA project, in which this factor is thoroughly controlled, different acquisitional outcomes are observed that cannot be explained by exclusively referring to it. Other parameters, such as the L1, learning style, psychometric characteristics and many more are probably involved. At times, one might even get the discouraging impression that most of what can be observed is up to the individual learner, in the end.

¹⁰¹ 5115

¹⁰² 5112

7. CONCLUSION

In the preceding chapters we have described and discussed in detail the results of two structured tests as well as learner semi-spontaneous interaction: we have then analysed the input from both a quantitative and qualitative point of view in order to find regularities that could explain learner behaviour, with partial success. All our work was guided by the general research questions outlined in the introduction as well as the more specific predictions detailed in each section. In this final chapter, we would like to summarise the results of our work so far and ask one more simple question: what have we learned from the VILLA project about the processing of complex morphosyntax in the earliest stages of L2 acquisition? What do we know now that we did not know when we started, and what can we retain for the future?

7.1. AN OVERALL SUMMARY

The methodological heart of both the repetition and the comprehension test is the simple intuition that the processing of inflectional morphology can be studied by manipulating word order. While SO targets can be processed by linking case endings to the corresponding functions or by relying on unmarked word order (83a), only the former strategy will work with OS targets (83b).

- (83) a. *dziewczynk-a* *ciągnie* *portugalk-ę*
 little girl.NOM pulls portuguese woman.ACC
 "the little girl pulls the Portugues woman"
- b. *portugalk-ę* *ciągnie* *dziewczynk-a*
 portuguese woman.ACC pulls little girl.NOM
 "the little girl pulls the Portugues woman"

Above chance accuracy scores in the processing of OS targets, therefore, should represent evidence that the learner has established a solid form-function association between case endings and the corresponding syntactic function. Our tasks only considered the opposition between nominative and accusative within the paradigm of feminine nouns in -a, instantiated by the endings -/a/ and -/e/, respectively, so that only two forms and two functions are involved. But the VILLA project is not a laboratory experiment exclusively targeting this point, and Polish is not an artificial language: this task, while apparently easy, had to be performed while breaking into a completely new language, characterised by exotic phonology and vocabulary and dozens of forms and functions to match to each other. That learners should be able to do that was not obvious, and indeed not everyone succeeded in the task.

What has been said can be applied directly to the comprehension test, in which learners were only required to listen to a target sentence and choose the interpretation which in their opinion best described it (Chapter 3). In addition to that, the repetition test (Chapter 2) involves a production component, so that learner output is observable directly. A superficial analysis of the data shows that the nominative case in -/a/ is hardly ever repeated incorrectly, whereas the accuracy rate for the accusative case varies greatly. When the latter is not repeated in a target-like manner, it is typically substituted by the -/a/ ending (84a, target in 84b).

- | | | | | |
|------|----|--------------------------|---------------------|--|
| (84) | a. | [artistk-a
artist-NOM | posdravja
cheers | tumatʃk-a] ¹⁰³
interpreter-NOM |
| | b. | /artistk-e
artist-ACC | pozdravja
cheers | twumatʃk-a/
interpreter-NOM |

This witnesses to the fact that if lexical items occur in only one form, which we will call "basic word form", this is going to be modelled on the nominative case,

¹⁰³ 2117, T2.

confirming the findings of other studies regarding the L1 or L2 acquisition of morphologically complex languages (see chapter 1).

We did not evaluate the results of the tests based on group means, as is usually the norm in this kind of studies. Instead, we preferred to compute the probabilities that learners obtained their scores in the absence of a rational, consistent principle, which in the case of OS targets, our main diagnostic tool, necessarily had to be of a morphosyntactic nature. Learners whose results exhibited sufficiently low probabilities of being due to guessing or a positional principle could be thought to have acquired the target opposition. Based on this methodology, we were able to combine the results of the two tests and compute a hierarchy of contexts (Chapter 4), defined as the interaction between task and word order. From hardest to easiest, this reads as following:

OS repetition \supset OS comprehension \supset SO repetition \supset SO comprehension

Evidently, OS targets and the repetition task make the learners' work harder. Based on the literature reviewed in chapter 1, word order effects are not unexpected. In fact, there are several alternative or complementary explanations for the learners' bias towards SO, such as the processing constraints hypothesised by Processability Theory (Pienemann 1998; Baten 2011; Artoni & Magnani 2015), VanPatten's (1996) First Noun principle, and Klein and Perdue's (1997) principles of utterance organisation in the Basic Variety. The last theoretical framework does not rely on syntactic categories like subject and object, but its interaction of phrasal, pragmatic and semantic constraints often results in utterances in which the first noun is both the topic and the entity exerting the greatest control over the situation, just like in the SO structures considered in this work.

Regarding the repetition test, it can be argued that it is more complex than the comprehension test in that it completely encompasses it, while at the same time exerting other demands on the learner. According to most of the literature available, the repetition test requires learners not just to repeat a string of

sounds, but to decode it (just like in the comprehension test) and then reproduce it, both operations being performed on the basis of the present stage of interlanguage development. Despite some points left unclear by this test, the implicational hierarchy above seems to make sense.

The learner sample who participated is fairly large (90 students), although they belonged to five different L1 groups. One of our general predictions (Chapter 1) based on the existing literature (e.g. Sagarra & Ellis 2013) regarded precisely the role of the native language, arguing that speakers of a morphologically complex language should be facilitated in the processing of a complex target morphological system. This turned out to be the case, as the German learners exhibited overall higher scores in both tests. Cross-linguistic interference turned out to be more complex than hypothesised, though, as the Italian speakers surprisingly performed almost just as well, despite the fact that their language does not express case on full nouns. The key seems to be in the fact that they showed exceptionally good repetition skills, sometimes unrelatedly to the corresponding processing abilities. We hypothesise that the Italian lexical stress might play an important role in this respect: while often found on the penultimate syllable, it is in principle free, which in turn could clear the learners from L1-induced bias in segmenting speech. This information leads to two interesting observations. First, it highlights the importance of perception and perceptual prominence in the (perhaps apparent) processing of morphology (Gallimore & Tharp 1981; Peters 1985). Second, it raises stimulating doubts as to the nature as well as the validity of the repetition test for our research purposes (Vinther 2002; Erlam 2006; Van Moere 2012).

We then turned to verifying the learners' processing of the same target structure in a very different context, namely semi-spontaneous interaction in pairs or small groups (Chapter 5). In describing characters to each other, learners produced numerous instances of transitive sentence, some of which contained feminine nouns in *-a*, the same targets employed in the structured tests (85).

- (85) *dzulj-alubi herbat-e kol-a i tfokolad-a*¹⁰⁴
Giulia-NOM likes tea-ACC coke -NOM and chocolate-NOM
"Giulia likes tea, coke and chocolate"

The analysis of the data show two interesting facts. First, accuracy rates for the accusative case marking, computed with the same statistical approach described above, are markedly lower than in the structured tests. Task effects (Skehan & Foster 1997; Michel 2011) were not unexpected, as it can be argued that interaction is a much more complex context than a rather abstract task conducted in a laboratory setting: in addition, or better prior to grammatical correctness, the person engaged in an interaction needs to keep track of discourse, elaborate his own messages, decode those produced by the interlocutor, take care of pragmatics and so on. No wonder then that such an accessory level as inflectional morphology should not be given priority (Klein 2002).

That inflectional morphology should be considered as an accessory structure is further evidenced by the qualitative analysis of learner output. While not all learners produce inflectional morphology in an accurate and consistent manner, as observed, the meaning of utterances is usually retrievable through alternative means (Siewierska & Bakker 2008), such as semantic and syntactic principles. The former relies on the fact that the nouns involved in transitive sentences usually differ in their animacy: animate nouns, specifically, have greater probabilities of performing the subject function, while the object function is more likely for inanimate nouns. In the rare cases in which both nouns are animate or inanimate, meaning is retrievable through the default SO word order. In fact, not a single OS utterance was produced in the whole data set of spontaneous learner output.

This casual observation represents a turning point for our argumentation. In fact, there may be two possible reasons why no OS structures are encountered. The first one would be that learners cannot yet produce this kind of structures

¹⁰⁴ 5114

in the present stage of interlanguage development, and intentionally avoid them: after all, in such a loose design, speakers are free to choose what structures and vocabulary to use. The data from the structured tests, truly, witness to the contrary: at least some learners can accurately and consistently manipulate inflectional morphology in both SO and OS sentences. We have already argued, however, that interaction is a much more complex and demanding environment in which to use language than a structured test. For the time being, then, we feel we cannot safely say much about the learners' ability to spontaneously produce OS structures.

Independently of whether this is the case or not, we can also argue that OS structures are not produced because they are not *necessary* in the context of the interactional task. The main purpose of these marked structures is to topicalise the object and focalise the subject, thus disrupting the unmarked alignment of pre-verbal position, topic and subject, on the one hand, and post-verbal position, focus and object, on the other hand¹⁰⁵. In this respect, the interactional task under consideration does not offer room for any such contrasts: participants simply describe a set of characters, based on the information provided or on their fantasy, and ask questions to elicit additional information. In such a context, pragmatically marked OS structures would be pragmatically inappropriate. Although learners may not be at ease with Polish OS structures, as adult speakers they are certainly familiar with the rules of information structure and pragmatics, so that they know that pragmatically and syntactically unmarked SO structures are just what they need.

In sum, it seems that the interactional task does not make it possible to observe whether or not learners can produce inflectional morphology in OS structures because it is not appropriate to elicit the obligatory contexts in which they would be required. This is the beginning of a long and complex debate about test design, characterised by a constant and unavoidable tension between

¹⁰⁵ In the VILLA project, due to the absence of passive constructions, the subject of transitive sentences is also the agent, while the object is always the patient.

realism and ecological validity, on the one hand, and control over the target structure, on the other hand.

At the one end of the continuum, eliciting spontaneous speech constrains learner output the least and may best approximate the speech that might be produced in a real-life communicative situation. For research purposes, however, it offers insufficient control as to what structures are going to be elicited. While this methodology may be perfectly appropriate to study common and obligatory structures like, say, articles in English, it would make it hard to elicit rarer or optional structures, like for example third type conditional clauses. It may be necessary to record huge amounts of data to only collect a few occurrences of the desired target. Moreover, there is no guarantee that these will appear at all.

At the other end of the continuum we find structured tests, which do offer full control over the target structure, but at the cost of creating an artificial and unnatural context which may not be representative of real-life communicative situations. Moreover, the risk is always present that if you look for something hard enough, you will find it eventually: in other words, tests may be structured in such a way as to condition learner output in a manner that does not necessarily reflect their actual competence.

In between these two extreme positions we find a variety of so-called "communicative tasks", designed to make the use of a given target structure more likely than in unstructured free production, while at the same time maintaining a certain degree of ecological validity. While this intuitively seems to be the right way to go, striking a balance between the two opposite requirements may be harder than it seems. This is a great lesson we learnt from the VILLA project: investigating the same target structures from two radically different perspectives made it evident to what extent the results obtained depend on the elicitation technique and on the communicative situation.

The analysis of learner spontaneous output, for example, highlighted semantic and syntactic regularities which made us suspicious as to whether they could be present in the input, too. After all, teacher speech was the only source of

information about Polish available to the learners, so that it is possible that it may have influenced the very structure of learner output, working alongside other universal, general constraints deriving from world knowledge and experience.

With this in mind, we morphologically tagged the entire input and analysed the transitive sentences comprised in it in terms of morphosyntax and semantics: specifically, we classified utterances in terms of word order, and individual words in terms of grammatical class, case, gender and animacy (Chapter 6). This enriched dataset allowed us to compute an index of form-function association for case ending and syntactic functions (Kempe & MacWhinney 1998) as well as type and token frequency (Ellis 2002) for a variety of constructions, including subjects, objects and transitive sentences manipulated for word order and referent semantics. The results show that indeed, strong tendencies do exist: the vast majority of subjects are instantiated by personal pronouns or person names; most objects are represented by inanimate nouns; structures with referents not differing in animacy are rare or absent altogether, and so on. All of this sounds rather intuitive, but the very fact that common-sense semantics strongly correlates with morphosyntax has far-reaching implications for test design and interpretation. The structures employed in our tests, for instance, comprising two animate nouns, are hardly present in the input. This does not affect case marking directly, as the feminine paradigm on which we focussed does not distinguish animacy. The finding does tell us, however, that the test results cannot be a mere reflex of the input, but instead witness to some degree of generalisation involving new lexemes and semantic classes. Another observation leads to the same conclusion. Some lexical items, like *matematyka*, "maths" only occurred in the input in their accusative form, due to their object-like semantics. Yet, learners often produce them in an invariable word-form modelled on the nominative case, just like all other lexical items belonging to the same inflectional class. This information is a precious piece of the debate as to the factors affecting the choice of the basic

word-form of a lexical item, and indeed on the development and complexification of learner varieties on the basis of the input (Hulstijn 2015).

If we had to summarise the results of our study, we could do that as follows: while several learners proved able to process inflectional morphology in a structured test after only a few hours of exposure to the input, much fewer could do the same in the context of spontaneous interaction. The lack of functional case marking had no effect on the efficacy of communication, though, as meaning was effectively transmitted through semantic and syntactic means, like animacy contrasts and default word order. The same is true of the input, which, though experimentally manipulated, only presented a handful of examples in which meaning could be only extracted on the basis of inflectional morphology. On the basis of these considerations, we could define "potential competence" what the learner is able to do in the aseptic, meaningless and highly abstract context of the structured test; "actual competence", in contrast, represents what the learner can do under pressure in a real communicative situation. This distinction is useful in order to identify structures which are necessary to survive in a communicative situation, and other which are only necessary to speak correctly or express sophisticated meaning. In order to express syntactic functions in an unmarked pragmatic context, inflectional morphology is largely superfluous. Two reasons may prompt learners to complexify their learner variety in that direction. First, they may want to sound more alike native speakers: this is what we would normally call grammatical correctness. Second, there may be functional reasons (Klein & Perdue 1997): for instance, expressing specific pragmatic nuances precluded by the unmarked SO word order, which in turn is inevitable in the absence of inflectional morphology.

7.2. A PROVISIONAL BALANCE

Let us come back to the two questions we asked at the beginning of this last chapter: what have we learned from the VILLA project about the initial

processing of morphosyntax? What do we know now that we did not know when we started, and what can we retain for the future?

We have learnt something as to the factors which affect the learners' ability to produce inflectional morphology in two classical psycholinguistic tasks, namely word order and task type: these results, however, are mostly theoretical and do not tell us much as to the practical development of morphology. Rather, if we had to name one fact which the present study can proficuously contribute to the present knowledge of SLA, it would be the following: after around 8 hours of exposure to a completely novel, morphologically complex natural language, some learners managed to use inflectional morphology productively and with above chance accuracy in transitive sentences, characterised by unmarked SO word order and elicited in an interactional task. This says something as to the actual learning possibilities of adults: in Corder's (1967) terms, at least part of the input, "what is available for going in" (p.165) has become intake, "what goes in", even in such a complex subsystem as morphosyntax. The fact that only a small proportion of participants succeeded further opens interesting perspectives for future research: to what are these difference due? Is it innate talent, or a function of motivation, learning style, effort, L1, teaching methodology?

As far as we are concerned, the greatest lesson we learnt from writing this work is a methodological one, and we are grateful that it should be so. The lesson concerns elicitation techniques: the VILLA project was quite exceptional in its methodological ambitiousness, the enormous effort provided to control the input and correlate it to the development of the interlanguages, as well as the numerous perspectives from which target structures were investigated. After becoming so intimate with the data, though, we begin to think that the elicitation procedures were not always adequate to the target structures and to their communicative function in real life. It is not enough to verify if learners can produce accusative case marking in OS structures: one should verify if they can - and will - do so in a real communicative situation in order to express some meaning. The first results of this awareness are already coming out:

though the potential of VILLA is partly still to be exploited, specific research questions have been taken on by a child project: VILLA NOVA. While the full experiment is still in preparation, a pilot study has been run at Humboldt University in Berlin in October 2016. Both the input and the tests reflect the experience developed thanks to VILLA: the development of inflectional morphology is still investigated through word order manipulation, but this time in a meaningful, pragmatically motivated manner. By the same token, structured tests have been substituted by communicative tasks with a specific non-linguistic objective to reach. Learners are put in the condition that discourse pragmatics requires them to topicalise the object and focalise the subject: whether or not they will manage to do so, and how, is the research question we ask.

The work which ends here raised more questions than it offered answers. While it is commonplace that this should already be considered as a scientific result of some relevance, the VILLA project proves quite exceptional one last time and allows us a slight variation on that classic conclusion for a research paper. We feel quite privileged, and thankful, that we can say: this work aroused more questions than it offered answers, but showed us exactly where to go next.

Bergamo, 15.12.2016

PRONUNCIATION GUIDE

Below we provide a quick pronunciation guide to Polish standard orthography, useful for reading the examples produced by the native speaker. This section is only intended as a reading aid: for a detailed description of Polish phonology, see Gussman (2007).

Orth.	IPA	notes	example	IPA
a	a		<i>ale</i> "but"	/ale/
ą	ɔN ¹⁰⁶	before stops and affricates	<i>początek</i>	/poʦɔntek/
ą	ɔw̃	before fricative and word-final	<i>mama</i> "mum.ins"	/mamɔw̃/
b	b		<i>niebieski</i> "blue"	/nʲebo/
c	ts		<i>co</i> "what"	/tso/
ć	tɕ		<i>pić</i> "to drink"	/pitɕ/
ch	x		<i>dach</i> "roof"	/dax/
ci	tɕ		<i>ciągnie</i> "pulls"	/tɕɔŋgnʲe/
cz	tʂ		<i>czarny</i> "black"	/tʂarni/
d	d		<i>dom</i> "home"	/dom/
dz	dz		<i>bardzo</i> "very"	/bardʑo/
dź	dʑ		<i>dźwięk</i> "sound"	/dʑvʲenk/
dż	dʒ		<i>drożdże</i> "yeast"	/droʒdʑe/
dzi	dʒ		<i>gdzie</i> "where"	/gdʑe/
e	ɛ		<i>we</i> "in"	/ve/
ę	ɛN	before stops and affricates	<i>między</i> "between"	/mʲɛndʑi/
ę	ɛw̃	before fricative	<i>męża</i> "husband"	/mɛw̃ʒa/
ę	ɛ	word-final	<i>kawę</i> "coffee"	/kave/
f	f		<i>flaga</i> "flag"	/flaga/

¹⁰⁶ The nasal archiphoneme /N/ indicates that the nasal consonant is homorganic with the following segment, and may be realised by its alveolar, bilabial or velar allophones.

g	g		<i>gra</i> "game"	/gra/
h	x		<i>hotel</i> "hotel"	/xotel/
i	i		<i>i</i> "and"	/i/
j	j		<i>ja</i> "I"	/ja/
l	l		<i>kolor</i> "colour"	/kolor/
ł	w		<i>mały</i> "small"	/mawi/
m	m		<i>mama</i> "mum"	/mama/
n	n		<i>komin</i> "fireplace"	/komin/
ń	ɲ		<i>dzień</i> "day"	/ɕɛɲ/
ni	ɲ		<i>niemiec</i> "german"	/ɲem'ets/
o	ɔ		<i>to</i> "this"	/to/
ó	u		<i>mówić</i> "to say"	/muviɕ/
p	p		<i>pokój</i> "room"	/pokuj/
r	r		<i>rower</i> "bike"	/rover/
rz	ʒ		<i>dobrze</i> "good"	/dobʒe/
s	s		<i>jest</i> "is"	/jest/
ś	ɕ		<i>śpi</i> "sleeps"	/ɕpi/
si	ɕ		<i>siostra</i> "sister"	/ɕostra/
sz	ʃ		<i>proszę</i> "please"	/proʃe/
t	t		<i>tam</i> "there"	/tam/
u	u		<i>tu</i> "here"	/tu/
w	v		<i>w</i> "in"	/v/
y	ɨ		<i>ty</i> "you"	/ti/
ź	ʒ		<i>jeździ</i> "goes"	/jeʒɔɰi/
ż	ʒ		<i>mąż</i> "husband"	/mɔɲʃ/
zi	ʒ		<i>zielony</i> "green"	/ʒeloni/

NOTES

- Lexical stress always falls on the penultimate syllable, except in learned loanwords from Greek or Latin and when clitics are attached, e.g.

matematyka "mathematics" /mate'matika/; *chodziliśmy* "we went"
/xo'ɕiliçmi/;

- Nasals and stops followed by pre-vocalic /i/ are palatalised to various degrees, e.g. *niebieski* "blue" /nⁱebieski/. The letter <i> in this case effectively functions as a diacritic.

TRANSITIVE STRUCTURES IN THE INPUT - PATTERNS

Table 62: transitive structure patterns

	EN		FR		GER		IT		NL		tot		% T2
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	
PN ♀ + Acc IN ♂	30	44	16	18	30	36	41	48	46	60	163	206	11,3%
PN ♂ + Acc IN ♂	44	63	8	13	24	32	23	30	44	54	143	192	10,6%
Nom Name ♀ + Acc IN ♂	26	33	7	8	19	20	34	43	22	22	108	126	6,9%
Nom Name ♂ + Acc IN ♂	20	33	5	7	13	19	22	26	13	18	73	103	5,7%
PN ♂ + + Acc IN ♀	12	15	3	5	18	26	18	20	24	31	75	97	5,3%
PN ♀ + Acc AN ♂	14	18	5	6	19	20	15	18	24	29	77	91	5,0%
PN ♀ + Acc IN ♀	22	25	6	8	13	17	10	10	14	15	65	75	4,1%
Nom Name ♀ + Acc AN ♂	13	16	4	6	9	10	27	32	9	12	62	76	4,2%
Nom Name ♂ + Acc IN ♀	13	14	2	2	10	10	23	26	15	17	63	69	3,8%
Nom Name ♀ + Acc IN ♀	14	14	1	1	7	7	18	21	17	19	57	62	3,4%

PN ♂ + Acc AN ♂	13	29	0	0	5	11	4	10	12	16	34	66	3,6%
Nom Name ♂ + Acc AN ♂	11	19	0	1	8	11	10	14	10	11	39	56	3,1%
Nom Name ♂ + Acc AN ♀	9	9	3	4	11	11	10	11	11	13	44	48	2,6%
Acc IN ♀ + Nom Name ♀	7	7	2	2	6	6	11	13	9	11	35	39	2,1%
Acc IN ♀ + Nom Name ♂	6	6	0	0	8	9	9	11	7	10	30	36	2,0%
Nom Name ♂ + Acc Name ♀	6	6	2	5	6	7	9	9	4	6	27	33	1,8%
Nom AN ♀ + Acc AN ♂	7	8	0	0	9	9	7	8	4	5	27	30	1,7%
Acc IN ♂ + Nom Name ♀	3	3	0	1	6	10	5	8	7	12	21	34	1,9%
Nom AN ♂ + Acc AN ♀	5	6	2	3	4	5	5	6	7	10	23	30	1,7%
PN ♂ + Acc AN ♀	7	9	0	0	2	2	3	3	10	15	22	29	1,6%
Acc AN ♂ + Nom Name ♀	6	6	2	2	3	3	3	3	9	9	23	23	1,3%
Acc AN ♂ +	2	2	1	1	5	5	1	2	10	10	19	20	1,1%

Nom Name ♂														
Acc Name ♀ + Nom Name ♂	2	2	1	4	2	3	3	5	4	10	12	24	1,3%	
Nom AN ♂ + Acc Name ♀	7	7	1	3	1	1	3	3	3	5	15	19	1,0%	
Acc IN ♂ + Nom Name ♂	2	2	0	0	2	3	6	6	3	6	13	17	0,9%	
Nom Name ♀ + Acc AN ♀	4	4	0	0	1	1	5	6	2	3	12	14	0,8%	
Acc Name ♀ + Nom AN ♂	2	2	1	3	0	1	3	5	1	8	7	19	1,0%	
Acc AN ♀ + Nom Name ♂	3	3	2	2	1	1	3	4	2	5	11	15	0,8%	
Nom AN ♂ + Acc AN ♂	3	4	0	0	2	3	2	2	4	5	11	14	0,8%	
Nom IN ♀ + Acc IN ♂	2	3	0	1	0	1	0	0	5	11	7	16	0,9%	
Acc Name ♀ + Nom Name ♀	1	1	0	0	2	4	1	3	3	7	7	15	0,8%	
PN ♂ + Acc Name ♀	1	1	0	0	4	4	1	1	3	4	9	10	0,6%	
Acc AN ♂ +	2	2	0	0	3	3	3	3	1	1	9	9	0,5%	

Nom AN ♀														
Nom IN ♂ + Acc IN ♂	3	6	1	1	0	0	1	1	2	2	7	10	0,6%	
PN ♀ + Acc AN ♀	0	0	0	0	1	1	4	4	3	3	8	8	0,4%	
Acc IN ♂ + Nom IN ♂	3	5	1	1	0	0	1	1	2	2	7	9	0,5%	
Nom AN ♀ + Acc IN ♀	3	3	0	0	1	2	1	1	2	2	7	8	0,4%	
Nom Name ♀ + Acc Name ♀	2	2	0	1	2	2	1	1	0	4	5	10	0,6%	
Nom AN ♂ + Acc IN ♂	2	3	0	1	2	3	1	1	0	1	5	9	0,5%	
Acc Name ♀ + Nom AN ♀	1	1	0	0	0	1	1	3	1	4	3	9	0,5%	
Nom AN ♂ + Acc IN ♀	1	1	1	1	1	1	1	1	1	2	5	6	0,3%	
Acc AN ♂ + Nom AN ♂	2	2	1	1	1	1	0	0	1	1	5	5	0,3%	
Nom AN ♀ + Acc IN ♂	0	1	0	1	1	1	0	1	1	3	2	7	0,4%	
Nom AN ♀ + Acc Name ♀	1	1	0	1	0	0	1	1	0	1	2	4	0,2%	
Acc IN ♂ + Nom AN ♂	0	0	0	0	3	3	0	0	0	0	3	3	0,2%	
Nom IN ♀ + Acc AN ♂	0	1	0	1	0	0	0	1	0	1	0	4	0,2%	

Acc AN ♀ + Nom AN ♂	0	0	0	0	0	0	0	1	0	2	0	3	0,2%
Acc IN ♂ + Nom IN ♀	0	0	1	1	0	0	0	1	0	0	1	2	0,1%
Acc AN ♂ + Nom IN ♂	0	0	0	0	0	0	0	0	1	1	1	1	0,1%
Acc IN ♀ + Nom IN ♂	0	0	1	1	0	0	0	0	0	0	1	1	0,1%
Acc IN ♀ + on	0	0	1	1	0	0	0	0	0	0	1	1	0,1%
Acc AN ♂ + PN ♀	1	1	0	0	0	0	0	0	0	0	1	1	0,1%
Nom IN ♂ + Acc AN ♂	0	1	0	0	0	0	0	0	0	0	0	1	0,1%
Nom AN ♀ + Acc AN ♀	0	0	0	0	0	0	0	0	0	1	0	1	0,1%
PN ♀ + Acc Name ♀	0	0	0	0	0	1	0	0	0	0	0	1	0,1%
Acc AN ♀ + Nom Name ♀	0	0	0	0	0	0	0	0	0	1	0	1	0,1%

REFERENCES

- Amato, Michael S. & Maryellen C. MacDonald. 2010. Sentence Processing in an Artificial Language: Learning and Using Combinatorial Constraints. *Cognition* 116(1). 143–148.
- Ananiewa, Natalia. 1997. Zjawisko interferencji w początkowym etapie nauczania języka polskiego w środowisku rosyjskojęzycznym. In Władysław Miodunka (ed.), *Nauczanie języka polskiego jako obcego*, 197–202. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Andorno, Cecilia & Ada Valentini (eds.). forthcoming. Descrivere il processo: le specificità del processo di acquisizione di una L2. *Verso una nuova lingua. Capire l'acquisizione di L2*. Torino: UTET.
- Artoni, Daniele & Marco Magnani. 2015. Acquiring case marking in Russian as a second language: an exploratory study on subject and object. In Camilla Bettoni & Bruno Di Biase (eds.), *Grammatical development in second languages: Exploring the boundaries of Processability Theory*, 177–193. Amsterdam: The European Second Language Association.
- Baddeley, Alan. 2003. Working memory and language: an overview. *Journal of Communication Disorders* 36. 189–208.
- Baddeley, Alan, Susan Gathercole & Costanza Papagno. 1998. The Phonological Loop as a Language Learning Device. *Psychological review* 105(1). 158–173.
- Bardovi-Harlig, Kathleen. 1992a. The use of adverbials and natural order in the development of temporal expression. *IRAL - International Review of Applied Linguistics in Language Teaching* 30(4).
- Bardovi-Harlig, Kathleen. 1992b. The relationship of form and meaning: A cross-sectional study of tense and aspect in the interlanguage of learners of English as a second language. *Applied Psycholinguistics* 13(03). 25–278.
- Bardovi-Harlig, Kathleen. 2000. *Tense and Aspect in Second Language Acquisition: Form, Meaning, and Use*. Oxford: Wiley-Blackwell.
- Baten, Kristof. 2011. Processability Theory and German Case Acquisition. *Language Learning* 61(2). 455–505.
- Bates, Elizabeth & Brian MacWhinney. 1981. Second-Language Acquisition from a Functionalist Perspective: Pragmatic, Semantic, and Perceptual Strategies. In Harris Winitz (ed.), *Native Language and Foreign Language Acquisition*, 190–214. New York: New York Academy of Sciences.

- Behrens, Heike. 2006. The input–output relationship in first language acquisition. *Language and Cognitive Processes* 21(1-3). 2–24.
- Bernini, Giuliano. forthcoming. The first encounter with a norm: perception and analysis of L2 initial input. In Elena Pandolfi, Johanna Miecznikowski, Sabine Christopher & Alain Kamber (eds.), *Language norms in context*. Frankfurt am Main: Peter Lang.
- Bernini, Giuliano. 2016. De la phonétique à la phonologie. La forme des mots au stade initiale de l'apprentissage de langues secondes. *Revue Française de Linguistique Appliquée* 21(2). 139–151.
- Biase, Bruno Di, Camilla Bettoni & Lucija Medojević. 2015. The development of case: a study of Serbian in contact with Australian English. In Camilla Bettoni & Bruno Di Biase (eds.), *Grammatical development in second languages: Exploring the boundaries of Processability Theory*, 195–212. Amsterdam: The European Second Language Association.
- Biase, Bruno Di & Satomi Kawaguchi. 2002. Exploring the typological plausibility of Processability Theory: language development in Italian second language and Japanese second language. *Second Language Research* 18(3). 274–302.
- Bittner, Dagmar, Wolfgang Dressler & Marianne Kilani-Schoch (eds.). 2000. *First verbs : on the way to mini-paradigms*. Berlin: ZAS.
- Bordag, Denisa & Thomas Pechmann. 2007. Factors influencing L2 gender processing. *Bilingualism: Language and Cognition* 10(03). 299–314.
- Borovsky, Arielle & Jeff Elman. 2006. Language input and semantic categories: a relation between cognition and early word learning. *Journal of Child Language* 33(4). 759–790.
- Boyd, Jeremy K., Erin A. Gottschalk & Adele E. Goldberg. 2009. Linking Rule Acquisition in Novel Phrasal Constructions. *Language Learning* 59. 64–89.
- Bremer, Katharina, Celia Roberts, Marie-Thérèse Vasseur, Margaret Simonot & Peter Broeder. 1996. *Achieving Understanding: Discourse in Intercultural Encounters*. London: Longman.
- Broeder, Peter, Guus Extra & Roeland Van Hout. 1989. Processes in the developing lexicon of adult immigrant learners. In Paul Nation & Ronald Carter (eds.), *Aila Review*, 86–109.
- Broeder, Peter, Guus Extra & Roeland Van Hout. 1993. Richness and variety in the developing lexicon. In Clive Perdue (ed.), *Adult language acquisition: cross-linguistic perspectives*, vol. 1: Field Methods, 145–163. Cambridge: Cambridge University Press.
- Brugman, Hennie & Albert Russell. 2004. Annotating Multimedia/Multi-modal resources with ELAN. In Maria T. Lino, Maria F. Xavier, Fátima Ferreira, Rute Costa & Raquel Silva (eds.), *Proceedings of LREC 2004, Fourth*

- International Conference on Language Resources and Evaluation, 2065–2068*. Paris: ELRA.
- Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5). 425–455.
- Bybee, Joan L. 1985. *Morphology: A Study of the Relation Between Meaning and Form*. John Benjamins Publishing.
- Bybee, Joan & Sandra Thompson. 2000. Three Frequency Effects in Syntax. *Berkeley Linguistics Society* 23(1). 378–388.
- Carroll, Susanne. 2005. Input and SLA: Adults' Sensitivity to Different Sorts of Cues to French Gender. *Language Learning* 55(S1). 79–138.
- Carroll, Susanne E. 1999. Putting "input" in Its Proper Place. *Second Language Research* 15(4). 337–388.
- Carroll, Susanne E. 2012. Segmentation on first exposure to an L2: evidence for knowledge-driven, top-down processing. In K. Braunmüller, C. Gabriel & B. Hänel-Faulhaber (eds.), *Multilingual Individuals and Multilingual Societies*, 23–45. Amsterdam: John Benjamins.
- Carroll, Susanne E. 2013. Introduction to the special issue: Aspects of word learning on first exposure to a second language. *Second Language Research* 29(2). 131–144.
- Carroll, Susanne Elizabeth. 2001. *Input and Evidence: The raw material of second language acquisition*. Amsterdam: John Benjamins.
- Carroll, Susanne E. & Elizabeth N. Widjaja. 2013. Learning exponents of number on first exposure to an L2. *Second Language Research* 29(2). 201–229.
- Casasola, Marianella. 2005. When Less Is More: How Infants Learn to Form an Abstract Categorical Representation of Support. *Child development* 76(1). 279–290.
- Casenhiser, Devin & Adele E. Goldberg. 2005. Fast mapping between a phrasal form and meaning. *Developmental Science* 8(6). 500–508.
- Chaudron, Craig. 1985. Intake: On Models and Methods for Discovering Learners' Processing of Input. *Studies in Second Language Acquisition* 7(1). 1–14.
- Chomsky, Noam. 1965. *Aspects of a theory of syntax*. Cambridge (MA): MIT Press.
- Chomsky, Noam. 1980. *Rules and representations*. Oxford: Basil Blackwell.
- Chomsky, Noam. 1981. Principles and parameters in syntactic theory. In Norbert Hornstein & David Lightfoot (eds.), *Explanation in linguistics: the logical problem of language acquisition*, 123–146. London: Longman.
- Collins, Laura, Pavel Trofimovich, Joanna White, Walcir Cardoso & Marlise Horst. 2009. Some Input on the Easy/Difficult Grammar Question: An Empirical Study. *The Modern Language Journal* 93(3). 336–353.

- Corder, S. P. 1967. The significance of learner's errors. *International Review of Applied Linguistics* 5(4). 161–170.
- Costa, Albert, Martin J. Pickering & Antonella Sorace. 2008. Alignment in second language dialogue. *Language and Cognitive Processes* 23(4). 528–556.
- Crocco Galeas, Grazie. 1998. *The Parameters of Natural Morphology*. Padova: Unipress.
- Dąbrowska, Ewa. 2004. Rules or schemas? Evidence from Polish. *Language and Cognitive Processes* 19(2). 225–271.
- Dąbrowska, Ewa. 2005. Productivity and beyond: mastering the Polish genitive inflection. *Journal of Child Language* 32(2). 191–205.
- Dąbrowska, Ewa. 2006. Low-level schemas or general rules? The role of diminutives in the acquisition of Polish case inflections. *Language Sciences* 28(1). 120–135.
- Dąbrowska, Ewa. 2008. The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language* 58(4). 931–951.
- Dabrowska, Ewa & Marcin Szczerbinski. 2006. Polish children's productivity with case marking: the role of regularity, type frequency, and phonological diversity. *Journal of Child Language* 33(03). 559.
- DąBrowska, Ewa & Michael Tomasello. 2008. Rapid learning of an abstract language-specific category: Polish children's acquisition of the instrumental construction. *Journal of Child Language* 35(03).
- DeKeyser, Robert. 2001. Automaticity and automatization. In Peter Robinson (ed.), *Cognition and Second Language Instruction*, 125–151. Cambridge: Cambridge University Press.
- Diego-Balaguer, Ruth De & Diana Lopez-Barroso. 2010. Cognitive and Neural Mechanisms Sustaining Rule Learning From Speech. *Language Learning* 60. 151–187.
- Diehl, Erika, Sandra Leuenberger, Isabelle Pelvat & Thérèse Studer (eds.). 2000. *Grammatikunterricht, alles für der Katz? Untersuchungen zum Zweitspracherwerb Deutsch*. (Reihe Germanistische Linguistik 220). Tübingen: Niemeyer.
- Dietrich, Rainer, Wolfgang Klein & Colette Noyau. 1995. *The Acquisition of Temporality in a Second Language*. . Vol. 7. (Studies in Bilingualism). Amsterdam: John Benjamins Publishing Company.
- Dimroth, Christine. submitted. Utterance structure after controlled contact to a new language: The impact of Basic Variety principles and input properties on child and adult novice learners of Polish.

- Dimroth, Christine, Rebekah Rast, Marianne Starren & Marzena Wałtorek. 2013. Methods for studying the learning of a new language under controlled input conditions: The VILLA project. *EUROSLA Yearbook*, vol. 13, 109–138. Amsterdam: John Benjamins.
- Doughty, Catherine J. S & Jessica Williams. 1998. *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press.
- Dressler, Wolfgang & Annemarie Karpf. 1994. The theoretical relevance of Pre- and Proto- morphology in Language Acquisition. *Yearbook of morphology*. 99–122.
- Dressler, Wolfgang U. 1985. On the predictiveness of Natural Morphology. *Journal of Linguistics* 21(02). 321–337.
- Dressler, Wolfgang U. 1987. *Leitmotifs in Natural Morphology*. Amsterdam: John Benjamins.
- Dressler, Wolfgang U. 2011. The rise of complexity in inflectional morphology. *Poznan Studies in Contemporary Linguistics* 47(2). 159–176.
- Dryer, Matthew S. 2013a. Definite articles. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/37>.
- Dryer, Matthew S. 2013b. Order of Subject, Object and Verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/81>.
- Dryer, Matthew S. 2013c. Determining Dominant Word Order. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/s6>.
- Duff, Patricia A. 1993. Tasks and interlanguage performance: an SLA perspective. In Graham Crookes & Susan M. Gass (eds.), *Tasks and Language Learning: Integrating Theory and Practice.*, 57–95. Clevedon: Multilingual Matters.
- Dziubalska-Kołączyk, Katarzyna. 1997. Pre- and Proto- in Polish phonology and morphology and their interrelations. *Papers and Studies in Contrastive Linguistics* 33(9). 159–171.
- Ellis, N.C. & L. Collins. 2009. Input and Second Language Acquisition: The Roles of Frequency, Form, and Function Introduction to the Special Issue. *The Modern Language Journal* 93(3). 329–335.
- Ellis, N.C. & F. Ferreira-Junior. 2009. Construction Learning as a Function of Frequency, Frequency Distribution, and Function. *The Modern Language Journal* 93(3). 370–385.

- Ellis, Nick. 2002. Frequency effects in language processing. *Studies in Second Language Acquisition* 24(02). 143–188.
- Ellis, Nick C. 2006a. Selective Attention and Transfer Phenomena in L2 Acquisition: Contingency, Cue Competition, Salience, Interference, Overshadowing, Blocking, and Perceptual Learning. *Applied Linguistics* 27(2). 164–194.
- Ellis, Nick C. 2006b. Language Acquisition as Rational Contingency Learning. *Applied Linguistics* 27(1). 1–24.
- Ellis, Nick C. & Nuria Sagarra. 2010a. Learned Attention Effects in L2 Temporal Reference: The First Hour and the Next Eight Semesters. *Language Learning* 60. 85–108.
- Ellis, Nick C. & Nuria Sagarra. 2010b. The bounds of adult language acquisition. *Studies in Second Language Acquisition* 32(04). 553–580.
- Ellis, Nick C. & Nuria Sagarra. 2011. Learned attention in adult language acquisition. *Studies in Second Language Acquisition* 33(04). 589–624.
- Ellis, Rod. 1985. *Understanding Second Language Acquisition*. Oxford: Oxford University Press.
- Ellis, Rod. 2005. Measuring implicit and explicit knowledge of a second language: a psychometric study. *Studies in Second Language Acquisition* 27(02).
- Endress, Ansgar D. & Luca L. Bonatti. 2007. Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition* 105(2). 247–299.
- Erlam, Rosemary. 2006. Elicited Imitation as a Measure of L2 Implicit Knowledge: An Empirical Validation Study. *Applied Linguistics* 27(3). 464–491.
- Ferguson, Charles A. 1975. Toward a Characterization of English Foreigner Talk. *Anthropological linguistics* 17(1). 1–14.
- Flege, James. 2009. Give input a chance! In Thorsten Piske & Martha Young-Scholten (eds.), *Input Matters in SLA*, 175–190. Bristol: Multilingual Matters.
- Folia, Vasiliki, Julia Uddén, Meinou De Vries, Christian Forkstam & Karl Magnus Petersson. 2010. Artificial Language Learning in Adults and Children. *Language Learning* 60. 188–220.
- Gallimore, Ronald & Roland G. Tharp. 1981. The Interpretation of Elicited Sentence Imitation in a Standardized Context. *Language Learning* 31(2). 369–392.
- Gaonac'h, Daniel. 1991. *Théories d'apprentissage et acquisition d'une langue étrangère*. Paris: Editions Didier.
- Gass, Susan M. 1997. *Input, Interaction, and the Second Language Learner*. Lawrence Erlbaum Associates.

- Gass, Susan, Alison Mackey & Lauren Ross-Feldman. 2011. Task-Based Interactions in Classroom and Laboratory Settings: Task-Based Interactions in Two Settings. *Language Learning* 61. 189–220. (1 November, 2016).
- Gentner, Dedre, Jeffrey Loewenstein & Barbara Hung. 2007. Comparison Facilitates Children's Learning of Names for Parts. *Journal of Cognition and Development* 8(3). 285–307.
- Gentner, Dedre & José Medina. 1998. Similarity and the development of rules. *Cognition* 65(2–3). 263–297.
- Giacalone Ramat, Anna. 1992. Grammaticalization Processes in the Area of Temporal and Modal Relations. *Studies in Second Language Acquisition* 14(03). 297–322.
- Givón, Talmy. 1985. Function, structure, and language acquisition. In Dan Slobin (ed.), *The Crosslinguistic Study of Language Acquisition*, 1008–1025. Hillsdale: Lawrence Erlbaum.
- Goldberg, Adele E., Devin Casenhiser & Nitya Sethuraman. 2004. Learning argument structure generalizations. *Cognitive Linguistics* 15(3). 289–316.
- Goldberg, Adele E., Devin Casenhiser & Tiffani R. White. 2007. Constructions as categories of language. *New Ideas in Psychology* 25(2). (Modern Approaches to Language). 70–86.
- Goldschneider, Jennifer & Robert DeKeyser. 2002. Explaining the “Natural Order of L2 Morpheme Acquisition” in English: A Meta-analysis of Multiple Determinants. *Language Learning* 51(1). 1–50.
- Grüter, Theres, Moti Lieberman & Andrea Gualmini. 2008. A Test Case for L1 versus UG as the L2 Initial State: The Acquisition of the Scope Properties of Disjunction by Japanese Learners of English. In Roumyana Slabakova, Jason Rothman, Paula Kempchinsky & Elena Gavrusseva (eds.), *Cascadilla Press: 9th Generative Approaches to Second Language Acquisition Conference (GASLA 2007)*, 47–56. Somerville: Cascadilla Proceedings Project. <http://cascadillapress.blogspot.it/2008/02/9th-generative-approaches-to-second.html> (17 May, 2015).
- Gullberg, Marianne, Leah Roberts & Christine Dimroth. 2012. What word-level knowledge can adult learners acquire after minimal exposure to a new language? *IRAL* 50(4). 239–276.
- Gullberg, Marianne, Leah Roberts, Christine Dimroth, Kim Veroude & Peter Indefrey. 2010. Adult Language Learning After Minimal Exposure to an Unknown Natural Language. *Language Learning* 60, suppl. 2. 5–24.
- Gülzow, Insa & Natalia Gagarina (eds.). 2007. *Frequency Effects in Language Acquisition: Defining the Limits of Frequency as an Explanatory Concept*. Berlin, New York: De Gruyter Mouton.

- Gussman, E. 2007. *The Phonology of Polish*. Oxford: Oxford University Press.
- Hadley, Pamela A., Matthew Rispoli, Colleen Fitzgerald & Alison Bahnsen. 2011. Predictors of Morphosyntactic Growth in Typically Developing Toddlers: Contributions of Parent Input and Child Sex. *Journal of Speech Language and Hearing Research* 54(2). 549.
- Håkansson, Gisela. 1989. The Acquisition of Negative Placement in Swedish. *Studia Linguistica* 43(1). 47–58.
- Haman, Ewa. 2003. Early productivity in derivation: A case study of diminutives in the acquisition of Polish. *Psychology of Language and Communication* 7(1). 37–56.
- Hamann, S. 2002a. Postalveolar Fricatives in Slavic Languages as Retroflexes. In S. Baauw, M. Huiskes & M. Schoorlemmer (eds.), *OTS Yearbook 2002*, 105–127. Utrecht: Utrecht Institute of Linguistics.
- Hamann, S. 2002b. Retroflexion and Retraction revised. *Papers on Phonetics and Phonology: The Articulation, Acoustics and Perception of Consonants.*, 13–26. (ZAS Papers in Linguistics 28).
- Hamann, S. 2004. Retroflex Fricatives in Slavic Languages. *Journal of the International Phonetic Association* 34(01). 53–67.
- Hamann, Silke. 2003. *The phonetics and phonology of retroflexes*. LOT.
- Han, ZhaoHong & Rebekah Rast (eds.). 2014. *First Exposure to a Second Language: Learners' Initial Input Processing*. Cambridge: Cambridge University Press.
- Han, ZhaoHong & Yayun Anny Sun. 2014. First exposure: a replication of Han and Peverly (2007). In Zhaohong Han & Rebekah Rast (eds.), *First Exposure to a Second Language*, 7–39. Cambridge: Cambridge University Press.
- Harley, Birgit & Merrill Swain. 1984. The interlanguage of immersion students and its implications for second language teaching. In A. Davies, C. Cripser & A. Howatt (eds.), *Interlanguage*, 291–311. Edinburgh: Edinburgh University Press.
- Haspelmath, Martin. 2008. Terminology of case. In Andrej L. Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*, 505–517. 1st ed. Oxford: Oxford University Press.
- Hentschel, G. 1990. Język polski w obcych ustach, uszach i umysłach. Esej na temat fonetyczno-fonologicznych przyczyn leksykalnych, kategorialnych i konstrukcyjnych uników w uczeniu się języka obcego. In W.T. Miodunka (ed.), *Język polski w świecie*. Warszawa-Kraków: Państwowe Wydawnictwo Naukowe.
- Hinz, Johanna, Carina Krause, Rebekah Rast, Ellenor M. Shoemaker & Marzena Watorek. 2013. Initial processing of morphological marking in

- nonnative language acquisition: Evidence from French and German learners of Polish. *EUROSLA Yearbook* 13. 139–175.
- Hopp, H. 2013. The development of L2 morphology. *Second Language Research* 29(1). 3–6.
- Houston, Tony. 1997. Sentence processing in Spanish as a second language: a study of word order and background knowledge. In Ana T. Pérez Leroux & William R. Glass (eds.), *Contemporary perspectives on the acquisition of Spanish*, 123–134. Somerville: Cascadilla Press.
- Hulstijn, Jan. 1989. Implicit and incidental second language learning: Experiments in the processing of natural and partly artificial input. In Hans-Wilhelm Dechert & Manfred Raupach (eds.), *Interlingual Processes*, 49–73. Tübingen: Narr.
- Hulstijn, Jan & Robert DeKeyser (eds.). 1997. Testing sla theory in the research laboratory. *Studies in Second Language Acquisition* 19(s2).
- Hulstijn, Jan H. (ed.). 1997. Second language acquisition research in the laboratory: possibility and limitations. *Studies in Second Language Acquisition* 19(02). 131–143.
- Hulstijn, Jan H. 2015. Discussion: How Different Can Perspectives on L2 Development Be?: Perspectives on L2 Development. *Language Learning* 65(1). 210–232.
- Jessop, Lorena, Wataru Suzuki & Yasuyo Tomita. 2007. Elicited Imitation in Second Language Acquisition Research. *Canadian Modern Language Review* 64(1). 215–238.
- Jezek, Elisabetta. 2016. *The lexicon: an introduction*. (Oxford Textbooks in Linguistics (otl)). New York, NY: Oxford University Press.
- Juffs, Alan & Michael Harrington. 2011. Aspects of working memory in L2 learning. *Language Teaching* 44(2). 137 - 166.
- Kempe, Vera & Brian MacWhinney. 1998. The acquisition of case marking by adult learners of Russian and German. *Studies in Second Language Acquisition* 20(3). 543–587.
- Kempe, Vera, Nada Ševa, Patricia J. Brooks, Natalija Mironova, Angelina Pershukova & Olga Fedorova. 2009. Elicited production of case-marking in Russian and Serbian children: Are diminutive nouns easier to inflect? *First Language* 29(2). 147–165.
- Kern, Sophie, Frédérique Gayraud & Florence Chenu. 2014. The role of input in early first language morphosyntactic development. *Language, Interaction and Acquisition* 5(1). 1–18.
- Kidd, Evan, Elena Lieven & Michael Tomasello. 2006. Examining the Role of Lexical Frequency in the Acquisition and Processing of Sentential Complements. *Cognitive Development* 21(2). 93–107.

- Kittleson, Megan M., Jessica M. Aguilar, Gry Line Tokerud, Elena Plante & Arve E. Asbjørnsen. 2010. Implicit language learning: Adults' ability to segment words in Norwegian. *Bilingualism* 13(4). 513–523.
- Klein, Wolfgang. 1986. *Second Language Acquisition*. Cambridge University Press.
- Klein, Wolfgang. 2002. Why case marking? In Ingrid Kaufman & Barbara Stiebels (eds.), *More than words: Festschrift for Dieter Wunderlich*, 251–273. Akademie Verlag. Berlin.
- Klein, Wolfgang & Clive Perdue. 1997. The Basic Variety (or: Couldn't natural languages be much simpler?). *Second Language Research* 13(4). 301–347.
- Krajewski, Grzegorz. 2005. The role of grammatical gender in the acquisition of noun inflection in Polish. *Psychology of Language and Communication* 9(1). 75–84.
- Krashen, Stephen D. 1981. *Second language acquisition and second language learning*. Pergamon Press.
- Krashen, Stephen D. 1985. *The input hypothesis: issues and implications*. Longman.
- Kruschke, John K. & Nathaniel J. Blair. 2000. Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review* 7(4). 636–645.
- LaBrozzi, Ryan M. 2012. The Role of Study Abroad and Inhibitory Control on Processing Redundant Cues. In Hispanic Linguistics Symposium, Kimberly L. Geeslin & Manuel Díaz-Campos (eds.), *Selected proceedings of the 14th Hispanic Linguistics Symposium*. Somerville, MA: Cascadilla Proceedings Project.
- Lantolf, James P. & Xian Zhang. 2015. Response to Pienemann's Critique of Zhang and Lantolf (2015): Response to Pienemann. *Language Learning* 65(3). 752–760.
- Larsen-Freeman, Diane & Michael H. Long. 1991. *An introduction to second language acquisition research*. Longman.
- Latos, Agnieszka. 2014. The effects of meaning-based and form-based input on the initial acquisition of Polish verbal morphology. In Anna Bonola, Paola Cotta Ramusino & Liana Goletiani (eds.), *Studi italiani di linguistica slava: strutture, uso e acquisizione*, 391–419. (Biblioteca Di Studi Slavistici 24). Firenze, Italy: Firenze University Press.
- Latos, Agnieszka, Marzena Watorek, Rebekah Rast & Marie Durand. 2016. Badanie procesu akwizycji języka drugiego od podstaw: projekt VILLA i jego praktyczne implikacje w nauczaniu języka obcego w systemie klasowo-lekcyjnym. In Jolanta Klimek-Grądzka & Majewska-Wójcik (eds.),

- N auczanie języka polskiego. Tradycje i innowacje*, vol. 2, 68–94. Lublin: Wydawnictwo KUL.
- Lee, James. 1987. Morphological factors influencing pronominal reference assignment by learners of Spanish. In Terrell Morgan, James Lee & Bill VanPatten (eds.), *Language and language use: Studies in Spanish*, 221–232. Landham, MD: University Press of America.
- Lee, James. 2002. The incidental acquisition of Spanish future morphology through reading in a second language. *Studies in Second Language Acquisition* 24(1). 55–80.
- Leeser, Michael J. 2004. Learner proficiency and focus on form during collaborative dialogue. *Language Teaching Research* 8(1). 55–81.
- Lenzing, Anke. 2013. *The Development of the Grammatical System in Early Second Language Acquisition: The Multiple Constraints Hypothesis*. John Benjamins Publishing.
- Lieven, Elena. 2010. Input and first language acquisition: Evaluating the role of frequency. *Lingua* 120(11). 2546–2556.
- LoCoco, Veronica. 1987. Learner comprehension of oral and written sentences in German and Spanish: The importance of word order. In Bill VanPatten, Trisha Dvorak & James Lee (eds.), 119–129. Cambridge MA: Newbury House.
- Łuczyński, Edward. 2002. Fleksja języka polskiego z punktu widzenia ontogenezy mowy. *Biuletyn Polskiego Towarzystwa Językoznawczego* 58. 157–165.
- Łuczyński, Edward. 2004. *Kategoria przypadku w ontogenezie języka polskiego czyli o wchodzeniu dziecka w rzeczywistość gramatyczną*. Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego.
- Łuczyński, Edward. 2010. Akwizycja gramatyki języka polskiego. *Psychologia Rozwojowa* 15(1). 9–18.
- Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Machowska, Jolanta. 2006. *Nabywanie kategorii przypadku. Okres wczesnoszkolny*. Kraków: Wydawnictwo Naukowe Akademii Pedagogicznej.
- MacWhinney, Brian. 2000. *The CHILDES Project: Tools for analyzing talk*. III. Mahwah, NJ: Lawrence Erlbaum.
- MacWhinney, Brian. 2011. The logic of the Unified model. In Susan M. Gass & Alison Mackey (eds.), *The Routledge Handbook of Second Language Acquisition*, 85–112. New York: Routledge.
- MacWhinney, Brian & Elizabeth Bates. 1987. Competition, variation, and language learning. In Brian MacWhinney (ed.), *Mechanisms of Language Acquisition*, 157–193. Hillsdale, NJ: Lawrence Erlbaum.

- MacWhinney, Brian, Elizabeth Bates & Reinhold Kliegl. 1984. Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior* 23(2). 127–150.
- MacWhinney, Brian & Catherine Snow. 1985. The child language data exchange system. *Journal of Child Language* 12(02).
- Macwhinney, Brian, Michael Tomasello & Dan Slobin. 2005. New Directions in the Competition Model. *Beyond nature-nurture: Essays in honor of Elizabeth Bates*, 81–110. Mahwah: Lawrence Erlbaum.
- Magajewska, M. 2008. Problemy fonetyczne Włochów uczących się języka polskiego - ujęcie porównawcze. In W.T. Miodunka & A. Seretna (eds.), *W poszukiwaniu nowych rozwiązań: dydaktyka języka polskiego jako obcego u progu XXI wieku*, 185–198. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Maguire, Mandy J., Kathy Hirsh-Pasek, Roberta Michnick Golinkoff & Amanda C. Brandone. 2008. Focusing on the relation: fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science* 11(4). 628–634.
- Majewska-Tworek, A. 2005. Trudności fonetyczne i fonologiczne niemieckojęzycznych studentów uczących się języka polskiego jako obcego. In P. Gancarek (ed.), *Nauczanie języka polskiego jako obcego i polskiej kultury w nowej rzeczywistości europejskiej. Materiały z VI Międzynarodowej Konferencji Glottodydaktycznej*, 159–163. Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.
- Malovrh, Paul. 2006. L2 Sentence Processing of Spanish OVS Word Order and Direct Object Pronouns: An Analysis of Contextual Constraints. In Nuria Sagarra & Almedia Toribio (eds.), *Selected Proceedings of the 9th Hispanic Linguistics Symposium*, 169–179. Somerville, MA: Cascadilla Press.
- Mcdonough, Kim & Youjin Kim. 2009. Syntactic Priming, Type Frequency, and EFL Learners' Production of Why-Questions. *The Modern Language Journal* 93(3). 386–398.
- Michel, Marije C. 2011. Cognitive and interactive aspects of task-based performance in Dutch as a second language. Oisterwijk: BOXPress.
- Miller, George. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 101(2). 343–352.
- Mintz, Toben H., Elissa L. Newport & Thomas G. Bever. 2002. The distributional structure of grammatical categories in speech to young children. *Cognitive Science* 26(4). 393–424.
- Moere, Alistair Van. 2012. A psycholinguistic approach to oral language assessment. *Language Testing* 29(3). 325–344.

- Mueller, Jutta L. 2006. L2 in a Nutshell: The Investigation of Second Language Processing in the Miniature Language Model. *Language Learning* 56(issue supplement s1). 235–270.
- Ninio, Anat. 2005. Accelerated learning without semantic similarity: indirect objects. *Cognitive Linguistics* 16(3). 531–556.
- Norris, John M. & Lourdes Ortega. 2000. Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-analysis. *Language Learning* 50(3). 417–528.
- Okura, Eve & Deryle Lonsdale. 2012. Working memory's meager involvement in sentence repetition tests. In Naomi Miyake, David Peebles & Richard P. Cooper (eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, 2132–2137. Austin: Cognitive Science Society.
- Olma, Marcelli. 2007. Rozwój niektórych kategorii fleksyjnych rzeczownika I czasownika u dziecka w wieku do 5 lat. *Poradnik językowy* 8. 80–92.
- Onnis, Luca, Heidi R. Waterfall & Shimon Edelman. 2008. Learn Locally, Act Globally: Learning Language from Variation Set Cues. *Cognition* 109(3). 423–430.
- Pančikowá, Marta. 1997. Polsko-słowacka interferencja (Swoistość nauczania języka polskiego Słowaków). In Władysław Miodunka (ed.), *Nauczanie języka polskiego jako obcego*, 215–219. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Park, Eun Sung. 2013. Learner-Generated Noticing Behavior By Novice Learners: Tracing the Effects of Learners' L1 on Their Emerging L2. *Applied Linguistics* 34(1). 74–98.
- Parodi, Teresa, Bonnie D. Schwartz & Harald Clahsen. 2006. On the L2 acquisition of the morphosyntax of German nominals. *Linguistics* 42(3). 669–705.
- Perdue, Clive (ed.). 1993. *Adult Language Acquisition: Volume 1, Field Methods: Cross-Linguistic Perspectives*. Cambridge: Cambridge University Press.
- Peters, Ann M. 1985. Language segmentation: operating principles for the analysis and perception of language. In Dan I. Slobin (ed.), *The crosslinguistic study of language acquisition.*, vol. 2, 1029–1067. Hillsdale: Lawrence Erlbaum.
- Pęzik, Piotr. 2012a. Wyszukiwarka PELCRA dla danych NKJP. In Adam Przepiórkowski, Mirosław Bańko, Rafał Górski & Barbara Lewandowka-Tomaszczyk (eds.), *Narodowy Korpus Języka Polskiego*, 253–273. Warszawa: Wydawnictwo Naukowe PWN.
- Pęzik, Piotr. 2012b. Język mówiony w NKJP. In Adam Przepiórkowski, Mirosław Bańko, Rafał Górski & Barbara Lewandowka-Tomaszczyk (eds.),

- Narodowy Korpus Języka Polskiego*, 37–47. Warszawa: Wydawnictwo Naukowe PWN.
- Pica, Teresa. 1994. Research on Negotiation: What Does It Reveal About Second-Language Learning Conditions, Processes, and Outcomes? *Language Learning* 44(3). 493–527.
- Pickering, Martin J. & Simon Garrod. 2004. The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences* 27(02).
- Pickering, Martin J. & Simon Garrod. 2006. Alignment as the Basis for Successful Communication. *Research on Language and Computation* 4(2-3). 203–228.
- Pienemann, Manfred. 1984. Psychological Constraints on the Teachability of Languages. *Studies in Second Language Acquisition* 6(02). 186.
- Pienemann, Manfred. 1998. *Language Processing and Second Language Development: Processability Theory*. Amsterdam: John Benjamins Publishing.
- Pienemann, Manfred. 2015. An Outline of Processability Theory and Its Relationship to Other Approaches to SLA: An Outline of Processability Theory. *Language Learning* 65(1). 123–151.
- Piske, Thorsten & Martha Young-Scholten (eds.). 2008. *Input matters in SLA*. (Second Language Acquisition). Clevedon: Multilingual Matters.
- Plunkett, K. & V. Marchman. 1991. U-shaped learning and frequency effects in a multi-layered perceptron: implications for child language acquisition. *Cognition* 38(1). 43–102.
- Plunkett, K. & V. Marchman. 1993. From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition* 48(1). 21–69.
- Pösingerová, Katerina. 1997. Polsko-słowacka interferencja (Swoistość nauczania języka polskiego Słowaków). In Władysław Miodunka (ed.), *Nauczanie języka polskiego jako obcego*, 221–224. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Przepiórkowski, Adam, Mirosław Bańko, Rafał Górski & Barbara Lewandowka-Tomaszczyk (eds.). 2012. *Narodowy Korpus Języka Polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- Pullum, Geoffrey K & Barbara C Scholz. 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 18(1-2).
- Rankin, Tom. 2014. Word order and case in the comprehension of L2 German by L1 English speakers. *EUROSLA Yearbook* 14. 201–224.
- Rastelli, Stefano. 2014. *Discontinuity in second language acquisition: the switch between statistical and grammatical learning*. Bristol; Buffalo: Multilingual Matters.

- Rast, Rebekah. 2008. *Foreign Language Input: Initial Processing*. Clevedon: Multilingual Matters.
- Rast, Rebekah. 2015. Primi passi in un nuovo sistema linguistico. In Elena Favilla & Elena Nuzzo (eds.), *Atti del XIV congresso dell'Associazione Italiana di Linguistica Applicata*.
- Rast, Rebekah & Jean-Yves Dommergues. 2003. Towards a Characterisation of Saliency on First Exposure to a Second Language. *EUROSLA Yearbook*, vol. 3, 131–156.
- Rast, Rebekah, Marzena Wątopek, Heather Hilton & Ellenor Shoemaker. 2014. Initial processing and use of inflectional markers: Evidence from French adult learners of Polish. In Zhao Hong Han & Rebekah Rast (eds.), *First Exposure to a Second Language: Learners' Initial Input Processing*, 64–106. Cambridge: Cambridge University Press.
- R Core team. 2014. *R: A language and environment for statistical computing*. <http://www.R-project.org/>.
- Reber, Arthur S. 1967. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 6(6). 855–863.
- Rescorla, Robert A. & Allan R. Wagner. 1972. "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement,." In Abraham Prokasy & Black (eds.), *Classical Conditioning II: Current Research and Theory*, vol. Vol. 2, 64–99. New York: Appleton-Century-Crofts.
- Ristin-Kaufmann, Nuria & Marianne Gullberg. 2014. The effects of first exposure to an unknown language at different ages. *Bulletin Suisse de Linguistique Appliquée* 99. 17–29. (23 November, 2015).
- Robinson, Peter. 2010. Implicit Artificial Grammar and Incidental Natural Second Language Learning: How Comparable Are They? *Language Learning* 60. 245–263.
- Romanova, Natalia & Kira Gor. 2016. Processing of gender and number agreement in Russian as a second language. *Studies in Second Language Acquisition*. 1–32.
- Roy, Brandon, Michael Frank & Deb Roy. 2009. Exploring word learning in a high-density longitudinal corpus. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*.
- Sachs, Jacqueline S. 1967. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics* 2(9). 437–442.
- Saffran, Jenny R., Elissa L. Newport & Richard N. Aslin. 1996. Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language* 35(4). 606–621.

- Sagarra, Nuria. 2014. Absence of morphological transfer in beginners: evidence from eye tracking. In ZhaoHong Han & Rebekah Rast (eds.), *First Exposure to a Second Language: Learners' Initial Input Processing*, 139–170. Cambridge: Cambridge University Press.
- Sagarra, Nuria & Nick C. Ellis. 2013. From seeing adverbs to seeing verbal morphology. *Studies in Second Language Acquisition* 35(02). 261–290.
- Saturno, Jacopo. forthcoming. The elicited imitation test in between morphosyntactic processing and rote repetition.
- Saturno, Jacopo. in prep. Correzione tra pari in varietà di apprendimento iniziali: uno strumento appropriato?
- Saturno, Jacopo. in prep. Task type and the emergence of morphosyntax in initial Polish L2.
- Saturno, Jacopo. 2013. Towards an efficient approach to transcription: some tactics and strategies. Negotiating methodological challenges in linguistic research, ms. <http://irg2014.org/wp-content/uploads/2013/01/Saturno.pdf>.
- Saturno, Jacopo. 2014. Case ending processing in initial Polish L2: the role of frequency, word order and lexical transparency. In Jacek Witkoś & Sylwester Jaworski (eds.), *New Insights into Slavic Linguistics*, 343–356. Frankfurt: Peter Lang.
- Saturno, Jacopo. 2015a. Copular structures in Polish L2. *Linguistica e Filologia* 35. 69–98.
- Saturno, Jacopo. 2015b. Effects of input condition on case ending processing in initial Polish L2. In Anna Bloch-Rozmej & Karolina Drabikowska (eds.), *Within Language, Beyond Theories: Issues in Applied Linguistics*. Newcastle-upon-Tyne: Cambridge Scholars Publishing.
- Saturno, Jacopo. 2015c. La trascrizione di Varietà di Apprendimento iniziali. In Marina Chini (ed.), *Il parlato in [italiano] L2: aspetti pragmatici e prosodici/[Italian] L2 spoken discourse: pragmatic and prosodic aspects*, 117–138. Milano: Franco Angeli.
- Saturno, Jacopo. 2015d. Perceptual prominence and morphological processing in initial Second Language Acquisition. In Amedeo De Dominicis (ed.), *pS-prominenceS: Prominences in Linguistics. Proceedings of the International Conference*, 76–95. Viterbo: DISUCOM press.
- Saturno, Jacopo. 2016. Strategie di formazione delle parole in varietà iniziali di polacco L2. In Valentina Benigni, Lucyna Gebert & Julija Nikolaeva (eds.), 279–305. Firenze: Firenze University Press.
- Saturno, Jacopo & Marzena Watorek. forthcoming. The emergence of functional case marking in initial varieties of Polish L2. *LIA*.
- Savickienė, Ineta & Wolfgang U Dressler. 2007. *The acquisition of diminutives a cross-linguistic perspective*. Amsterdam; Philadelphia: J. Benjamins Pub.

- Schimke, Sarah. 2011. Variable verb placement in second-language German and French: Evidence from production and elicited imitation of finite and nonfinite negated sentences. *Applied Psycholinguistics* 32(04). 635–685.
- Schmidt, Richard W. 1990. The Role of Consciousness in Second Language Learning. *Applied Linguistics* 11(2). 129–158.
- Sharwood-Smith, Michael. 1993. Input enhancement in instructed SLA: theoretical bases. *Studies in Second Language Acquisition* 15(2). 165–179.
- Shoemaker, Ellenor. 2015. The Development of Perceptual Sensitivity to Polish Sibilants at First Exposure. *Proceedings of the Annual Meetings on Phonology 2*.
- Shoemaker, E. & R. Rast. 2013. Extracting words from the speech stream at first exposure. *Second Language Research* 29(2). 165–183.
- Siewierska, Anna. 1993. Syntactic weight vs information structure and word order variation in Polish. *Journal of Linguistics* 29(02). 233–265.
- Siewierska, Anna & Dik Bakker. 2008. Case and alternative strategies: word order and agreement marking. In Andrej L. Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*, 290–303. 1st ed. Oxford: Oxford University Press.
- Singleton, David. 2012a. The Critical Period Hypothesis: A coat of many colours. *IRAL - International Review of Applied Linguistics in Language Teaching* 43(4). 269–285.
- Singleton, David. 2012b. Ultimate Attainment and the Critical Period Hypothesis: Some Thorny Issues. In Marzena Watorek, Sandra Benazzo & Maya Hickmann (eds.), *Comparative perspectives on language acquisition: a tribute to Clive Perdue*, 188–203. (Second Language Acquisition 61). Clevedon: Multilingual Matters.
- Skehan, P. & P. Foster. 1997. Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research* 1(3). 185–211.
- Skiba, Romuald & Norbert Dittmar. 1992. Pragmatic, Semantic, and Syntactic Constraints and Grammaticalization. *Studies in Second Language Acquisition* 14(03). 323–349.
- Slobin, Dan I. 1985. Crosslinguistic evidence for the language-making capacity. In Dan I. Slobin (ed.), *The crosslinguistic study of language acquisition.*, vol. II, 1157–1256. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Smoczyńska, Magdalena. 1972. Przystwajanie form deklinacji rzeczownikowej przez dzieci. *Psychologia Wychowawcza* 5.
- Smoczyńska, Magdalena. 1985. The acquisition of Polish. In Dan Slobin (ed.), *The crosslinguistic study of language acquisition.*, vol. 1, 595–686. Hillsdale: Lawrence Erlbaum.

- Smoczyńska, Magdalena. 1997. Przystwajanie systemu gramatycznego przez dziecko. In Halina Mierzejewska & Maria Przybysz-Piwko (eds.), *Rozwój poznawczy i językowy dzieci z trudnościami w komunikacji werbalnej*. Warszawa: DiG.
- Smoczyński, Paweł. 1965. Trudności Niemców z wymową polską oraz związane z tym inne lingwistyczne problemy. In Jerzy Lewandowski (ed.), *Metodyka nauczania języka polskiego jako obcego: wybór artykułów*, 83–95. Warszawa: Państwowe Wydawnictwo Naukowe.
- Spreafico, Lorenzo. 2005. Lo sviluppo lessicale di un apprendente di italiano L2. Problemi e metodi di analisi quantitativa. In Giorgio Banti, Antonietta Marra & Edoardo Vineis (eds.), *Atti del 4° congresso di studi dell'Associazione Italiana di Linguistica Applicata*, 241–257. Perugia: Guerra.
- Starren, Marianne. 2001. *The Second Time: The Acquisition of Temporality in Dutch and French as a Second Language*. Utrecht: LOT.
- Szapkina, Olga. 1997. Podstawowe trudności w nauczaniu języka polskiego początkujących studentów rosyjskojęzycznych. In Władysław Miodunka (ed.), *Nauczanie języka polskiego jako obcego*, 203–208. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Tarone, E. 1983. On the Variability of Interlanguage Systems. *Applied Linguistics* 4(2). 142–164.
- Tarone, Elaine. 1979. Interlanguage as Chameleon. *Language Learning* 29(1). 181–191.
- Tarone, Elaine E. 1985. Variability in interlanguage use: a study of style-shifting in morphology and syntax. *Language Learning* 35(3). 373–403.
- Thompson, Sandra A. 2002. "Object complements" and conversation: Towards a realistic account. *Studies in Language* 26(1). 125–163.
- Tokowicz, Natasha & Brian MacWhinney. 2005. Implicit And Explicit Measures Of Sensitivity To Violations In Second Language Grammar: An Event-Related Potential Investigation. *Studies in Second Language Acquisition* 27(02). 173–204.
- Tomasello, Michael. 2005. Beyond formalities: The case of language acquisition. *The Linguistic Review* 22(2-4).
- Valentini, Ada & Roberta Grassi. 2016. Oltre la frequenza. L'impatto della trasparenza e dell'accento sull'apprendimento del lessico in L2. In Loredana Corrà (ed.), *Sviluppo della competenza lessicale. Acquisizione, apprendimento, insegnamento*, 125–143. Roma: Aracne.
- VanPatten, Bill. 1990. Attending to Form and Content in the Input. *Studies in Second Language Acquisition* 12(03). 287–301.
- VanPatten, Bill. 1996. *Input Processing and Grammar Instruction in Second Language Acquisition*. Norwood: Ablex.

- VanPatten, Bill. 2000. Thirty years of input (or intake, the neglected sibling). In Bonnie Schwierzbin, Frank Morris, Michael Anderson, Carol Klee & Elaine Tarone (eds.), *Social and cognitive factors in Second Language Acquisition: Selected proceedings of the 1999 Second Language research forum*, 287–311. Somerville: Cascadilla Press.
- VanPatten, Bill & Stefanie Borst. 2012. The Roles of Explicit Information and Grammatical Sensitivity in Processing Instruction: Nominative-Accusative Case Marking and Word Order in German L2. *Foreign Language Annals* 45(1). 92–109.
- VanPatten, Bill & Tony Houston. 1998. Contextual effects in processing L2 input sentences. *Spanish Applied Linguistics* 2. 53–70.
- VanPatten, Bill & Megan Smith. 2015. Aptitude as grammatical sensitivity and the initial stages of learning Japanese as a L2. *Studies in Second Language Acquisition* 37(01). 135–165.
- Vinther, Thora. 2002. Elicited imitation: a brief overview. *International Journal of Applied Linguistics* 12(1). 54–73.
- Vouloumanos, Athena. 2008. Fine-grained sensitivity to statistical information in adult word learning. *Cognition* 107(2). 729–742.
- Watorek, Marzena & Jacopo Saturno. 2016. La variabilité inter- et intra-individuelle dans le cas des apprenants débutants du polonais. Paper presented at the Journée d'Etude « La variabilité des parcours acquisitionnels en L2 », Université de Toulouse Jean Jaurès (FR).
- Wesche, M. 1994. Input and interaction in second language acquisition. In Clare Gallaway & Brian J. Richards (eds.), *Input and Interaction in Language Acquisition*, 219–249. Cambridge: Cambridge University Press.
- Wheeler, Max W. 1993. On the hierarchy of naturalness principles in inflectional morphology. *Journal of Linguistics* 29(01). 95–111.
- Wickham, Hadley. 2012. *stringr: Make it easier to work with strings*. <http://CRAN.R-project.org/package=stringr>.
- Williams, J.N. 2010. Initial Incidental Acquisition of Word Order Regularities: Is It Just Sequence Learning? *Language Learning* 60(suppl. 2). 221–244.
- Wulff, Stefanie, Nick C. Ellis, Ute Römer, Kathleen Bardovi-Harlig & Chelsea J. Leblanc. 2009. The Acquisition of Tense–Aspect: Converging Evidence From Corpora and Telicity Ratings. *The Modern Language Journal* 93(3). 354–369.
- Wurzel, Wolfgang U. 1989. *Inflectional Morphology and Naturalness*. Dordrecht: Kluwer Academic Publishers.
- Year, Jungeun & Peter Gordon. 2009. Korean Speakers' Acquisition of the English Ditransitive Construction: The Role of Verb Prototype, Input Distribution, and Frequency. *The Modern Language Journal* 93(3). 399–417.

- Zhang, Xian & James Lantolf. 2015. Natural or Artificial: Is the Route of L2 Development Teachable?: Natural or Artificial. *Language Learning* 65(1). 152–180.
- Zipf, Georg. 1935. *The psycho-biology of language: An introduction to dynamic philology*. New York: Houghton-Mifflin.
- Zygis, M. 2003. Phonetic and Phonological Aspects of Slavic Sibilant Fricatives. In T.A. Hall & S. Hamann (eds.), *ZAS Papers in Linguistics*, 175–213. (32).
- Żygis, Marzena & Jaye Padgett. 2010. A perceptual study of Polish fricatives, and its implications for historical sound change. *Journal of Phonetics* 38(2). 207–226. (10 June, 2012).
- Zygis, M. & S. Hamann. 2003. Perceptual and acoustic cues of Polish coronal fricatives. *Proceedings of the 15th International Congress of Phonetic Sciences*, 395–398.