Utterance structure in initial L2 acquisition

Jacopo Saturno

Eurosla Studies

No series description provided

Utterance structure in initial L2 acquisition

Jacopo Saturno

Saturno, Jacopo. 2020. *Utterance structure in initial L2 acquisition* (Eurosla Studies). Berlin: Language Science Press.

Contents

Contents

Contents

Acknowledgments

1 Introduction: Rationale and research questions

The purpose of this work is to analyse the strategies of input processing employed by beginner learners of Polish in the earliest stages of second language acquisition (SLA), with a particular focus on morphosyntax. It targets a minimal fragment of grammar, i.e. a subset of the singular number of the paradigm of feminine nouns in -*a* (e.g. *siostra* 'sister'). The morphosyntactic opposition of interest contrasts the nominative case (NOM), encoded by the ending [a] <a>, and the accusative case (ACC), encoded by the ending [e] <ę>, which respectively correspond to the subject (SUBJ) and object (OBJ) syntactic functions (1).

(1) siostr-a   lubi Warszaw-ę
    sister-NOM likes Warsaw-ACC
    '(my) sister likes Warsaw'

Case morphology makes it possible to vary word order as required for pragmatic purposes, as in (2), in which word order is manipulated to topicalise the OBJ.

(2) Warszaw-ę lubi siostr-a,   nie brat-∅
    Warsaw-ACC likes sister-NOM not brother-NOM
    'it is (my) sister who likes Warsaw, not (my) brother'

This is the target system which the learners in the present book will be called to master.

The data are a subset of the results of the VILLA project (Dimroth et al. 2013), a large SLA experiment devoted to the earliest stages of the acquisition of L2 Polish. 188 participants divided into five L1 groups (English, German, Dutch, French, Italian) were selected based on their lack of experience in Slavic languages, so as to make sure that the acquisition process started approximately from the same baseline for all of them, i.e. from scratch. Each L1 group took part in a two-week, 14-hour L2 Polish course taught by a specially trained native speaker, who delivered the same material in the various editions of the project. The course was designed

in such a way as to expose the participants to input carefully planned in terms of both range of lexical and grammatical structures and frequency; moreover, learners were often asked to engage in simple question-and-answer exchanges with the teacher as well as among themselves. The development of their interlanguage was observed through several tasks targeting various layers of language, such as phonology, morphosyntax and pragmatics.

The teacher's speech and the PowerPoint slides she used during classes represented the only sources of Polish input available to the learners. This was entirely recorded and transcribed, so that it is possible to search for regular correspondences between learner performance and the relevant features of teacher input.

The processing of the target structure is investigated in three different tasks: a comprehension task, an elicited imitation task, and a semi-spontaneous production task. This approach makes it possible to obtain a wider picture as to learner skills in the interlanguage than is typically possible in a typical psycholinguistic experiment, while at the same time maintaining the same amount of control over a wide range of variables. In addition to providing a comprehensive picture of learner skills in the manipulation of target morphosyntax, this work also attempts to identify a scale of task difficulty, defined as the interaction of the skill required to perform the exercise (comprehension, production), the target structure (SO vs. OS sentences) and the context in which the elicitation of the data takes place (structured test vs. semi-spontaneous production). The analysis is pursued on the basis of a fairly large data-set collected in a methodologically thorough manner, which makes it possible to exclude an uncontrolled effect of such variables as the amount and quality of the input received and the existing skills in the target language. Indeed, it has long been argued that acquisition success (and the related notion of acquisition difficulty) results from the complex interactions of a wide range of factors (Housen & Simoens 2016). The present work aims to experimentally exclude or control as many variables as possible in order to focus on the three main factors mentioned above.

This introductory chapter aims to contextualise this work against the wider picture of SLA studies, with particular respect to the initial stages of acquisition. The main research questions may be summarised as follows:

- What principles of utterance organisation do initial learners of L2 Polish adopt in order to identify and express syntactic functions?

- Do principles of utterance organisation vary depending on the communicative situation in which the target language is processed? Can a scale of task difficulty be identified?

- Does the learner's L1 influence the speed or accuracy of the acquisition process?

- Is there a relation between selected input parameters and the observed L2 output? Three frequency-based descriptors of the controlled VILLA input are considered, i.e. the strength of the association between inflectional endings and the corresponding grammatical meaning, the type frequency of SUBJ and OBJ and the frequency of transitive structures described in terms of word order and referent animacy and gender.

1.1 Word order vs. inflectional morphology

The general research question pursued in this work is whether and to what extent learners employ inflectional morphology to decode and encode utterance meaning in comprehension and production. To exemplify, a simple SO utterance like (3) may be successfully interpreted by relying on at least two processing strategies, lying at the opposite poles of a continuum.

(3) dziewczynk-a ciągnie portugalk-ę
little.girl-NOM pulls   Portuguese.woman-ACC
'the little girl pulls the Portugues woman'

First, learners may adopt a morphosyntactic principle, i.e. derive syntactic functions from inflectional morphology. This requires that case endings are categorised into paradigms depending on several features of the individual item, such as inflectional class, gender, animacy and number, as the same morph - understood here as a sound or string of sounds - may vehiculate different meanings in different inflectional paradigms: for instance, even within the limited set of VILLA lexical items considered in the present work, the ending *-a* may encode meanings such as NOM of feminine nouns or ACC of masculine animate nouns. Thus, in order to derive meaning from inflectional endings, the learner should know what inflectional class a given lexical item belongs to.

Alternatively, the utterance may be interpreted based on a positional principle, whereby words are assigned syntactic functions depending on their relative position in the utterance, i.e. the order in which they appear in the string. In this respect, various sources suggest that the SO word order should be considered the unmarked option. Firstly, SO is the dominant constituent order in all L1s involved in the VILLA project, although with various degrees of rigidity. Secondly, typological research indicates that this constituent order is by far more widespread

among the languages of the world than OS orders (Dryer 2013a). The reasons for such biased distribution, in turn, are generally believed to be of a cognitive nature (Siewierska & Bakker 2008.

The original research question may thus be reformulated as whether or not learners will be able to process an utterance's syntactic structure based on inflectional morphology, instead of relying on a default constituent order.

The skill referred to is sometimes labelled in the literature as the ability to process grammatical form, in addition to lexical and pragmatic meaning. While most studies point to the fact that initial learners tend to focus on lexical morphemes and ignore grammatical ones (e.g. VanPatten's (1996) PRIMACY OF MEANING principle, or Klein's (1986) and Rast's (2008) experimental evidence from Elicited Imitation tasks), cases in which learners attend to grammatical meaning (i.e. form) first are also documented. Park (2013) and Han & Peverly (2007) reported that their beginner learners of Korean and Norwegian, respectively, employed a form-based approach, supposedly contrary to the primacy of meaning principle. In both cases, it could be supposed that this was the case because the target language was lexically so different from any languages known to the participants that no processing for meaning seemed possible. As a result, learners turned to the analysis of formal regularities in the text.

As far as the extraction of grammatical meaning is concerned, word order lies at the core of another of VanPatten's principle, namely 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, provided that common sense does not suggest an alternative interpretation based on the learner's world knowledge, the situation in question or the nature of the referents involved (e.g. an inanimate noun is unlikely to be the subject of a sentence, even though it does appear in utterance-initial position).

In the absence of any context, as is the case in the VILLA experiment described in this book, grammatical meaning can only be identified based on word order or inflectional morphology; in the case of syntactically marked structures which depart from the basic SO word order, only the latter principle will lead to the correct interpretation. For this reason, the manipulation of word order is a key diagnostic tool to detect the learner's processing strategy.

As explained in the previous section, even in a morphologically rich language such as Polish, inflectional morphology is normally not the only cue to grammatical meaning, which may be suggested – although not with full certainty – by other hints such as word order and the semantics of the lexical items involved. The structured tasks employed in the present work aim to eliminate all these ancillary resources, in order to make inflectional morphology the only source of

information as to grammatical meaning. The manipulation of word order then becomes a crucial diagnostic tool of learner morphosyntactic skills: if in SO targets meaning can be identified independently of inflectional morphology by relying on the default relative order of SUBJ and OBJ, this is not possible in the case of OS targets, in which the same approach would lead to an incorrect interpretation of the utterance. This rationale is applied to the structured tests described in Chapter 4 (the Elicited Imitation task) and Chapter 5 (the Comprehension task).

In the case of the latter task, the research question is fairly straightforward: if learners manage to correctly identify the syntactic structure of OS targets, one can hypothesise that they are able to associate case endings to the corresponding syntactic meaning.

The situation is more complex as far as the production task is concerned, in which learners are asked to listen to a stimulus question and repeat it as accurately as possible following a distracting pause. The rationale of the test is that it does not require learners to repeat the target as a string of sounds, but rather to decode it and then re-produce it based on the present state of the interlanguage grammar (Chapter 4). In addition to a comprehension stage, then, this task also implies production, which raises an additional question: provided that learners correctly identify the syntactic structure of the target, will they rely on inflectional morphology to express it, or will they fall back on default word order? In addition, learners are required not only to understand the target, but also to repeat it as accurately as possible, which naturally includes inflectional morphology. One may therefore investigate whether the word order configuration of the stimulus question has an impact on the ability of learners to correctly inflect nouns for case.

Regarding the development of inflectional paradigms, some authors signal a phase of non-basic marking, in which a sort of mini-paradigm (Bittner et al. 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 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'

1 Introduction: Rationale and research questions

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

Similar observations were made concerning the acquisition of L1 Polish. Łuczyński (2002; 2004; 2010) shows that paradigms start off with three forms marking three grammatical functions, namely nominative, accusative and vocative. However, because of frequent instances of syncretism, some of these functions are performed by the same form, e.g. *dom*, 'home'[NOM/ACC] as opposed to *chłopak*, 'boy'[NOM] vs. *chłopak-a*, 'boy'-ACC. 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 begins, in which words appear in two forms, one of which is modelled on the nominative and the other simply contrasts with it. Further, Dziubalska-Kołaczyk (1997), following Dressler & Karpf (1995), 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. Proto-morphology marks the beginning of the morphological system of the language according to the principles of Natural Morphology (Dressler 1985; 1987; Dressler 2011; Wurzel 1989; Crocco Galeas 1998). The phase of morphology proper, finally, entails the full development and completion of the inflectional and derivation systems of the target language.

Concerning the role of word order in the acquisition of inflectional morphology, studies conducted within the framework of Processability Theory (Pienemann 1998; 2015; Di Biase & Bettoni 2015) show quite clearly that accusative case marking first emerges in syntactically unmarked SVO sentences, in which the marked, non-nominative marking appears in post-verbal position, as shown by Artoni & Magnani (2015: 190) (see example (5a) below). Indeed, this tendency is so strong as to generalise to contexts which do not require accusative marking, like the post-verbal SUBJ in the German L2 examples in (5b) (Baten 2011: 490) and 5c (Diehl et al. 2000: 235).

(5) 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, while at the same time clearly marking syntactic functions.

These observations highlight a further reason why the OS word order may be considered as the marked, more demanding option compared to its SO counterpart. Building on Levelt's (1989) speech model and lexico-functional grammar (Bresnan 2001), Processability theory postulates that a dealignment between the semantic level (argument structure), the syntactic level (syntactic functions) and the actual order in which the arguments appear (constituent structure) is associated with a higher processing cost. To exemplify, placing in sentence-initial position an argument other than the semantically prominent agent argument, in turn associated to the subject syntactic function, requires a disruption of the default alignment which requires time and practice to be mastered (see Bettoni & Di Biase 2015 for a comprehensive description of this theoretical approach).

1.2 Task effects: structured tests vs. semi-spontaneous production

While structured tasks undoubtedly provide the researcher with a fully controlled environment to test linguistic hypotheses, it can be argued that they hardly resemble any realistic communicative situation, so that making general claims as to the learner's linguistic skills on the basis of structured tests alone may not be an unproblematic operation (Ellis 1985: 289–290). This criticism is not new, from Krashen's (1981) distinction between acquisition and learning to research conducted within the Learner Variety approach (Perdue 1993; 1996; Starren 2001; Bernini 2003; Giacalone Ramat 2003), which is almost entirely based on spontaneous production data, to Processability Theory, whose claims are programmatically founded on the quantitative analysis of spontaneous speech.

After studying the learners' morphosyntactic skills through structured tasks, questions of a more applied nature emerge, such as to what extent input really becomes intake, i.e. is sufficiently acquired and automatized to be ready for use when needed for communication. Is it possible that input is assimilated to a degree sufficient to use the target structure in a given context (such as a structured

task), but not others? These questions are pursued in the present work by comparing the learners' performance on the same target structure in two different contexts, i.e. two structured tasks as opposed to semi-spontaneous production, in which participants are required to talk to each other in Polish in order to solve a practical extra-linguistic task. In addition to lexical and grammatical accuracy, the learner here has to pay attention to discourse structure and to the development of the interaction with the interlocutor.

Such semi-spontaneous production is seen here as a concrete test of morphosyntactic skills previously observed in the controlled, yet artificial, environment of the structured tests. The question one asks at this stage is "given what learners can do in a laboratory context, what will they prove able to do when using language not just to perform an exercise, but to actually communicate a message?". To this end, the results of the structured tests will be compared to performance in semi-spontaneous production in order to highlight any systematic discrepancies in morphosyntactic accuracy, and perhaps even a threshold in the structured test score which learners have to meet in order to be able to produce inflectional morphology in semi-spontaneous production.

Comparisons between task types are often encountered in the debate on task effects (Révész et al. 2016; Plonsky & Kim 2016; Sasayama 2016) and linguistic-cognitive complexity, defined as the mental resources allocated and cognitive mechanisms deployed in processing and using a given structure (Housen & Simoens 2016). Although a detailed discussion of these topics lies beyond the scope of this work, it is worthwhile to briefly sketch why some tasks may seem harder than others.

Skehan and Foster's (Skehan 2009; Skehan & Foster 2001) Limited Attentional Capacity model advocates that the amount of information (in terms of data and goals) one can keep track of is limited. When performing a task, various components compete with each other for attention. This leads to trade-off effects, as only those processes which are allocated sufficient attention will be performed at the optimal level; performance in all others will inevitably decline. Crucially, if there is an extra-linguistic communicative objective, this receives priority. In other words, learners first aim to express their message in an effective, though not necessarily accurate manner. If enough attentional resources are left, they can be destined to objectives such as complexity, accuracy and fluency (Skehan & Foster 2007).

This view is not shared by another influential approach, namely Robinson's (2001; 2005; 2015) Cognition Hypothesis, whereby trade-off effects do not necessarily occur because different processes may draw from different attentional pools. Decreases in performance only take place when task complexity increases

in terms of resource-dispersing factors (as opposed to resource-directing), such as reduced planning time. In contrast, increasing complexity in terms of resource-directing factors may actually result in production which is both more accurate and more complex, if required to reach the communicative goal. Indeed, this is partly the case of the tasks considered in this work, in which the production task requires a much wider range of lexical items and grammatical structures than the structured tests, although, as stated before, some particularly complex structures (like OS transitive sentences) may be avoided because they are either too difficult or simply unnecessary.

By comparing learner performance in two very different contexts, i.e. structured tests and semi-spontaneous production, the analysis presented in this book aims to approximate a comprehensive view of learner skills as far as NOM/ACC case marking is concerned. The two structured tests make it possible to explore what learners are able to do in comprehension and production in the best possible conditions, i.e. in a laboratory setting and with a task of limited complexity. The semi-spontaneous production component shows what the same learners can do in a realistic, complex communicative situation.

1.3 L1 influence

It may be hypothesised that speakers of specific L1s may be advantaged in the processing of the target structure. The rationale behind this claim is that the processing of the L2 target structure rests on mechanisms which are similar to those of the L1 and are consequently available to the speaker (Tokowicz & MacWhinney 2005; Ellis 2006a). Effectively, it appears that speakers of morphologically complex languages are more at ease when processing target languages characterised by complex morphology. **EllisSagarra2010b**; Ellis & Sagarra (2010; 2011)) studied the acquisition of temporal reference in Latin after only one hour of input exposure. While they found that focusing learner attention to verbs or adverbs orients their processing strategies towards that category, they also highlighted important L1 effects, such as the fact that speakers of morphologically poor languages such as Chinese and English tended to rely more heavily on lexical cues than did speakers of morphologically more complex L1s, such as Spanish and Russian. When paradigm complexity increased, however, all learners seemed 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.

In the context of the present work, L1 interference is relevant in light of the fact that the speakers who were exposed to the same Polish input and took the same experimental tasks were speakers of five different L1s. Crucially, only German behaves similarly to Polish in terms of the morphological expression of case and word order manipulation for pragmatic purposes: all other languages only encode fragments of case in the pronominal paradigm and tend to adhere to a default SO word order, although other word orders are also possible for pragmatic purposes. If the rationale of the hypothesis is correct, then the German learners should prove faster and more accurate in the processing of Polish morphosyntax. More generally, differences in learner output are expected which can be attributed to an L1 effect.

1.4 Input control

Studying the output of learners confronted with a completely novel language may be illuminating with regard to the general mechanisms of input processing (**RastEtAl2001**; Perdue 2002), provided that it is possible to have full control over the input and correlate its relevant input parameters with learner output. In the present work, input is defined as any item of the target language that learners are exposed to through any channel.

Everyone roughly agrees that input should play some role in SLA, but opinions start to differ broadly as soon as the debate moves to the way in which input, "what is available to go in" is processed and transformed into intake, "what goes in", in Corder's (1967: 165) words. Following **MacWhinney2010**; **MacWhinney2015** and Hulstijn (2015), two main streams of theories may be identified in the prolific literature that has developed around this topic, namely the generative position and a constellation of emergentist approaches.

In the generative framework (see Rankin & Unsworth 2016 for a recent review), input is mainly seen as the activator of an innate mechanism (Chomsky 1981). An innate capacity for language must be postulated, since language acquisition cannot be solely based on the input received because the learners' input is deficient, an argument known as the poverty of the stimulus (Chomsky 1980). The language faculty is thus seen as an innate system, input only providing a few examples which the language processor will take care to shape and systematise through the acquisition process.

In contrast, emergentist approaches maintain that input contains a wealth of information which learners are equipped to analyse using a variety of cognitive

processes (Tomasello 2005), including the statistical search for form-function associations. In this respect, 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".

Within the emergentist universe, the Learner Variety approach seems particularly important in the architecture of the present work. Input is seen as a wealth of linguistic material which learners interpret and shape based on their provisional interlanguage grammar. Moreover, the L2 learner is seen as a proficient speaker of at least another language, an expert communicator ready to employ all known strategies to transmit the intended message. In an attempt to do so, linguistic elements may be reinterpreted and assigned meaning which they do not possess in the native variety. To exemplify, Bernini (2018b) and Dimroth (2018: 28–33) suggest that in some VILLA data the instrumental word form *strażakiem* contrasts with its nominative equivalent *strażak* in that with some learners it seems to express plurality.

But in the absence of clear, reliable data on the input received, 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 the various parameters under investigation, whether related to the input or not. Full control, in turn, requires that the target language should be completely unknown to the learner, so that input effects may be teased apart from existing knowledge. But since most of the commonly investigated languages are relatively widespread, it is usually hard to find learners who have never had any exposure to the target language, however minimal.

One possibility is to employ short samples of "exotic" (Gullberg et al. 2010; Carroll 2012a,c,b; Carroll & Widjaja 2013; Carroll 2014) or artificial languages (Hulstijn 1997; Williams 2010). Both solutions make it possible to perfectly tune the target language to the desired research questions. In addition, since the target language can only be learned during the experiment itself, every learner's learning experience is necessarily identical. For these reasons, numerous studies have made this choice. However, the language samples considered typically lack the complexity and idiosyncrasies of natural languages, so that the ecological validity of such studies may be questioned (Hulstijn 1989; Robinson 2010).

The VILLA project was designed to draw a clearer picture of input processing based on the results of the research just referred to. Thanks to its methodology, learner performance can be directly checked against the input received through two fundamental methodological steps. First, the learners were selected in such a

way as to make sure that they had no previous knowledge of the target language. The choice of an uncommonly taught language like Polish facilitated their recruitment. This approach ensured that all learners began the acquisition process from the same baseline. Second, input was entirely controlled throughout the experiment. Therefore, it may be argued that one of the general research questions pursued in this book is whether or not a relation may be identified between the features of the input received by learners and their own output. Input control also makes it possible to verify if learners can generalise the patterns contained in the input to target structures which differ in various respects.

A very natural question concerns the effect of additional exposure to the input. Clearly, such effect is expected to be positive: more precisely, it may also be argued that the closer the learner variety gets to the target language, the more learners will be able to rely on form-based processing, i.e. on grammar, 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 (Bardovi-Harlig 2000; Lee 2002; Leeser 2004; Ellis & Sagarra 2010), while 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 1992; Skiba & Dittmar 1992; Dietrich et al. 1995; Starren 2001; Bardovi-Harlig 2000; Parodi et al. 2006; Bordag & Pechmann 2008). To verify this claim, most VILLA tasks were repeated several times throughout the course in order to monitor the progress of the interlanguage. This is also the case of the two structured tests considered in the present work, which were repeated twice with a 4:30 hour lag. This makes it possible to verify whether additional, albeit limited, exposure to the input contributed to modify the learners' strategies of input processing in any way, though presumably towards the target-like morphosyntactic principle.

In addition to experimentally isolating the target variables, input control may directly contribute to explaining learner errors and shedding light on the main issue investigated in this book. The following sections detail two input parameters which appear to be particularly relevant for the study at hand.

1.4.1 Markedness: frequency and form-function association

The first question is whether or not the statistical distribution in the input of the target endings [a] NOM and [e] ACC may favour any of the two endings, or, said otherwise with an ambiguous (Haspelmath 2006), yet practical terminology, if

any of the two terms may be considered as a marked alternative. This topic is pursued in terms of form-function association, with two predictions:

a) the form which is most closely associated to the corresponding meaning will be more likely to be over-generalised, if errors indeed occur;

b) more speculatively, it may be argued that if the function of one of the forms under consideration is not easy to identify, learner preferences may shift towards a principle of utterance organisation in which that form is not required. In the present context, that implies the positional principle, whereby grammatical meaning is independent of inflectional morphology and only relies on the relative position of nouns within the utterance.

Form-function association refers to the strength of the link between a given linguistic meaning and the forms which express it: said otherwise, how frequently and unambiguously a given form is used to convey a given function, and vice versa. The rationale has been developed in different theoretical frameworks, such as Competition model (MacWhinney & Bates 1987) and Natural Morphology (Dressler 1987). Research applied to several L1s has shown that the degree to which a given form suggests the corresponding function varies across 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 et al. 1984). Further, the strength of the form-function association is a powerful predictor of acquisition success and rapidity. Kempe & MacWhinney (1998) demonstrated that the Russian case system, although much more complex than its German equivalent, is more rapidly acquired because of the more systematic relationship between case endings and grammatical meaning.

The analysis presented in this book attempts to calculate the strength of the association between the two target endings [a] and [e] and the corresponding grammatical meaning (SUBJ and OBJ, respectively). The relative strength of the form-function association should be a particularly good predictor of what form is selected as the basic word form of the learner variety.

1.4.2 Generalisability of input models

The second question regards the learners' ability to generalise target structures as they occur in the input to other models of utterances, differing with respect to a few characteristics of the lexical items involved, like animacy, gender etc. This question is of great relevance for the more general debate on the role of input, especially with regard to the generativist and usage-based perspectives. More specifically, if learners only prove able to process the target structure in the same type of utterances encountered in the input, one may consider the acquired

construction as a chunk. The analysis of semi-spontaneous production further makes it possible to verify whether or not learners choose to adhere to the input model when given a choice.

For the purposes of a comparison with learner performance, the input is characterised in terms of two factors, namely the type frequency of the SUBJ and OBJ syntactic functions, on the one hand, and the token frequency of transitive structure models defined in terms of word order, gender, animacy and word class.

Type frequency refers to the number of lexical items which occur in a construction. The literature on this topic maintains two positions which may seem mutually exclusive. On the one hand, several researchers argue that learners might benefit from high type frequency, whereby the same construction is instantiated by a greater number of types (Bybee 1985; Bybee 1995; Bybee & Thompson 2000a; Bybee 2006; Goldberg et al. 2004; Onnis et al. 2008), especially as far establishing abstract patterns is concerned (Mcdonough & Kim 2009). Type frequency ensures productivity, as hearing several different lexical items in a certain context makes it less likely that that construction may become specifically associated with any of them. Further, if a construction is instantiated by many items, it is probably quite general in meaning and easily generalisable to other items. Conversely, Kruschke & Blair (2000) argue that learning that a particular stimulus is associated with a particular outcome hinders the association of the same outcome with another stimulus at a later time, a phenomenon knows as *blocking* (Ellis 2006a). Finally, high type frequency ensures frequent use in speech (Bybee & Thompson 2000b).

Other researchers claim that highly skewed distributions may be just as beneficial. In a skewed distribution, the vast majority of the occurrences of a given grammatical constructions is instantiated by a small number of lexical items. The rationale is that construction learning is a process of categorisation (Goldberg et al. 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 learners are indeed facilitated in the construction of categories by low-variance input (Gentner et al. 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 (1935) law, whereby the frequency of a given word is inversely proportional to its rank in a frequency table. As a result, a small number of very common words account for a substantial proportion of all tokens in a text (Mintz et al. 2002). There may be various reasons for this, usually linked

with the semantics of the words involved (Kidd et al. 2006; Thompson 2002; Ellis & Ferreira–Junior 2009).

The comparison of learner performance under different input skewedness conditions shows that its beneficial effects are not completely clear (Borovsky & Elman 2006; Casenhiser & Goldberg 2005). In most cases the results do not point to a single, univocal predictor of acquisition success, but rather suggest that all the parameters considered jointly drive acquisition (Year & Gordon 2009; Wulff et al. 2009).

In the present book, the analysis of type frequency is designed to quantify the extent to which target structures encountered in the input are associated with particular lexical items. In the case where structures are strongly associated with a limited number of items in the input, one can assume that applying a construction to a different set of lexical items in the structured tests will require a degree of abstraction and generalisation.

With the same aim, the input analysis presented in the book also consists in searching for the most common models of transitive utterances in terms of word order, noun animacy and noun gender. The purpose of this step is to compute the number of input examples corresponding to the test target structures to which the learners were exposed throughout the course.

The comparison of a prototypical input transitive sentence (6a) with a target sentence of the Elicited Imitation task (6b) highlights the fact that although the target structure is arguably the same, at least from a morphosyntactic point of view (the expression of SUBJ and OBJ through [a] NOM and [e] ACC), the two utterances differ in several respects. In addition to word order (SVO vs OVS), which is a variable controlled for experimentally, notable differences may be noted in terms of semantics, whereby the OBJ function is instantiated by an inanimate noun in (6a), but by a common nationality noun in (6b); by the same token, the SUBJ is a person name in (6a) and again a common noun in (6b). The verbs also differ in terms of argument structure, so that *ciągnie* 'pulls', but not *lubi* 'likes' may be considered as a prototypical transitive verb. This in turn is defined here as a verb in which the syntactic subject performs the semantic role of agent. Thus, *lubi* 'likes' is clearly transitive from a syntactic point of view, because it requires a subject (marked as NOM) and a direct object (marked as ACC); however, it is not prototypically transitive because the syntactic functions SUBJ and OBJ do not correspond to the semantic roles AGENT and PATIENT.

(6) a. Mari-a   lubi kaw-ę.
      Maria-NOM likes coffee-ACC
      'Maria likes coffee.'

    b. Portugalk-ę   ciągnie dziewczynk-a.
        Portuguese.woman-ACC pulls   little.girl-NOM
        'The little girl pulls the Portuguese woman.'

The research question thus investigates whether or not learners can identify the morphosyntactic structure of interest in the input and apply it to somewhat different sentence models, in which semantics is of no help to the expression of grammatical meaning.

1.5 A note on labels and notation

Polish examples in this book are normally transcribed in standard Polish orthography if they were uttered by a native speaker or if they are used within a theoretical argumentation. A guide to reading Polish orthography is provided in the Appendix.

A broad IPA transcription (Landau et al. 1999; Jassem 2003) is used to transcribe utterances produced by learners, in order to avoid any undue morphosyntactic interpretation of the raw data (**Saturno2015**). The rationale for this decision is as following. Because of the rich inflectional system of Polish, grammatical meaning is often indicated by a single word-final sound (7a), potentially in addition to a stress shift (7b) due to the substitution of a zero morph with a vocalic ending (the lexical stress of virtually all Polish words falls on the penultimate syllable).

(7)   a. *studentk-a* 'female student-NOM.SG' vs. *studentk-ę* 'female student-ACC.SG'

        b. *strażak-∅* 'fireman-NOM.SG' vs. *strażak-a* 'fireman-GEN/ACC.SG'

On the one hand, learner varieties typically exhibit very conspicuous phonological deviations from their native target. On the other hand, phonological variability may occur in the earlies stages, albeit with no functional value: "There is no inflection in the B[asic] V[ariety] [...]. Thus, lexical items typically occur in one invariant form. [...] Occasionally, a word shows up in more than one form, but this (rare) variation does not seem to have any functional value: the learners simply try different phonological variants". (Klein & Perdue 1997: 311). Broeder et al. (1993: 160–161) hypothesise that "random variation on the phonetic and phonological level at the first stages of second language acquisition is gradually replaced by variation produced by the acquisition of proper morphological rules", especially with regard to verbs, the most highly inflected word class in the target languages of the ESF project (Perdue 1993). Indeed, the VILLA production data

too contain examples which suggest a productive, systematic use of phonological variation to express grammatical meaning. Because of their fluid state, though, it is often problematic to distinguish phonological variability from contrasts reflecting an opposition in meaning (Bernini 2018a,c; Dimroth 2018). It is argued that a phonetic transcription of learner output leaves the question open for analysis, without imposing an *a priori* interpretation which may later condition the discussion of the results.

A few words should be spent to clarify what labels will be used in order to refer to the functions that nouns may perform in an utterance. Within this study, a prototypical transitive sentence is composed of two noun phrases (NP) in utterance-initial (NP1) and utterance-final (NP2) position, as well as a bivalent verb in utterance-medial position. It can be argued that the two NPs tend to concentrate three functions belonging to the layers of information structure, syntax and semantics.

First, in unmarked transitive structures the NP1 is part of the topic, i.e. what is talked about in the utterance, while NP2 is part of the comment, i.e. what is said about it. Further, Since Polish is a predominantly SO language, NP1 is most often the grammatical subject of the sentence, identified by noun-verb agreement. It follows that NP2 must be the grammatical object.

From the semantic point of view, NP1 is usually characterised by a higher degree of agency, which in terms of semantic roles corresponds to agent or experiencer (Table 1.1). It is no coincidence that more often than not the referent of NP1 is animate; for the same reason, NP2 is typically inanimate (Chapter 3).

Table 1.1: alignment of information structure, syntax and semantics

Mari-a	pije	herbat-ę
Maria-NOM	drink.3SG	tea-ACC
TOP	COMMENT	
SUBJ		OBJ
A		P

If the whole VILLA input were composed of prototypical transitive sentences, the three labels would appear to be interchangeable. That is not the case, however: although N1 and NP2 tend to be characterised by a coincidence of functions in terms of information structure, syntax and semantics, there may be occasions in which that arrangement is disrupted. It is therefore desirable to identify the label which best describes the function of the two NPs independently of their position in the sentence.

1 Introduction: Rationale and research questions

Clearly TOP does not suit this purpose because it is rigidly linked to the utterance-initial position. In marked word orders such as OS, NP1 is still the TOPIC of the sentence but it encodes a syntactic function other than SUBJ, such as OBJ (Table 1.2).

Table 1.2: disalignment of information structure, syntax and semantics

herbat-ę	pije	Mari-a
tea-ACC	drink.3SG	Maria-NOM
TOP		COMMENT
OBJ		SUBJ
P		A

Semantic roles may appear to be more intuitive to the linguistically untrained learner. The notion of subject after all rests on noun-verb agreement, which is a meta-linguistic concept. Semantic roles, on the other hand, may be thought to more faithfully reflect the role of arguments (i.e. referents) in a given situation, which does not seem to imply any meta-linguistic reasoning. However, this is only true with respect to prototypically transitive verbs, i.e. verbs whose first argument can be identified as the agent. In other cases, it may be difficult to clearly identify a true agent and a true patient. Indeed, this case is fairly common in the VILLA input, some representative examples of which are presented in (8). The first argument of these verbs is better described as experiencer (8a and 8b) or possessor, although syntactically (i.e. based on noun-verb agreement) it is clearly the subject of the utterance.

(8) a. student-∅   zna   język-∅   polsk-i
      student-NOM know.3SG language-ACC Polish-ACC
      'the student speaks Polish'

    b. siostr-a   lubi   brat-a
      sister-NOM like.3SG brother-ACC
      'the sister likes the brother'

    c. dziewczynk-a ma   balonik-∅
      little.girl-NOM have.3SG balloon-ACC
      'the little girl has a baloon'

Similar considerations apply to the labels controller and controllee, often employed in the literature related to the Learner Variety approach (Klein & Perdue

1992; Perdue 1993) to refer to the "argument of a verb by the greater or lesser degree of control that its referent exerts, or intends to exert, over the referents of the other argument(s)" (Klein & Perdue 1997: 314).

To summarise, a VILLA transitive structure may always be described with reference to the syntactic functions SUBJ and OBJ, whereas A and P are not always appropriate because the two syntactic functions (especially SUBJ) may correspond to more than one semantic role. Moreover, SUBJ and OBJ suggest the role that the corresponding constituent would presumably fulfil in the target language, whose principles of utterance organisation are of a syntactic nature. On the one hand, this argument reminds Bley-Vroman's (1983) comparative fallacy, whereby learner output is interpreted in light of the target model, rather than in its own right: in this respect, the comparative fallacy is frowned upon in SLA studies, because it obscures the internal structure of the interlanguage and introduces a clearly evaluative (rather than descriptive or interpretative) approach to L2 data. On the other hand, at times the use of the labels SUBJ and OBJ in the description of interlanguage output may represent a useful terminological shortcut to indicate the meaning which the learners would presumably express if they mastered the target grammar sufficiently. Although learner output may be organised around functions which do not play the same role in native varieties (such as controller or topic, as argued above), for practical purposes it may be useful to refer to their intended function in the target language, which (like the learners' L1s) is based on syntactic categories.

It may be anticipated that the target sentences of the structured tasks described in chapters 4 and 5 do contain prototypically transitive verbs such as *push*, *pull*, *call*, *cheer*, so that the argument stated above may appear not to be particularly influential. This is not the case in the semi-spontaneous interaction described in Chapter 7, in which participants were free to use the whole range of known lexical items, which indeed includes non-prototypical transitive verbs like *love*, *have*, *know* etc.

For these reasons the labels SUBJ and OBJ will be used throughout this study.

2 The VILLA Project: Methodology

The present chapter presents an introduction to the VILLA project, with a specific focus on those aspects which are directly relevant for the object of the volume. A more detailed description of all other aspects may be found in Dimroth et al. (2013) and the forthcoming VILLA manual (Watorek et al. in prep).

2.1 The course

The objective of the VILLA project was to observe the very earliest stages of the acquisition of a morphologically complex language in light of the input received. It follows that the input is a particularly crucial component of the project, as not only did it contain the raw material for language acquisition, but was also a carefully controlled independent variable in the experiment design.

In order to maximise learner engagement and provide a realistic environment, the VILLA input was provided in the shape of a 14-hour interactive 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: Nijmegen (the Netherlands), Osnabrück (Germany), Paris VIII (France), York (UK), Pavia, Bergamo (Italy).

The research question of the VILLA project required that input should be thoroughly controlled for. Moreover, for the purposes of cross-linguistic comparison, it had to be kept as constant as possible across the various editions. To this end, input was planned in advance, and a course schedule was prepared for the teacher to follow in all editions. The course thus had a very precise structure, detailing the topics to cover, the vocabulary to introduce, the activities to perform, and, crucially, the frequency with which lexical items had to occur during classes.

Although some slides included a few written words in Polish orthography, Polish orthographic conventions were never introduced, so that learners would have been hardly able to autonomously read them in a target-like manner. Nevertheless, it is quite possible that they tried to pair the words they heard in the input to their written representation in some of the slides used in the course (the only available source of written input).

As far as contents are concerned, the VILLA course describes a handful of characters (the members of the Kowalski family) in terms of nationality, address, family links, profession, likes and dislikes, home furniture etc. A specific section is devoted to the description of a map and to giving route directions. An unexhaustive list of the grammatical structures practices throughout the course includes copular structures, transitive constructions, prepositional phrases, nominal paradigms, various verb endings (infinitive, 3sg) etc.

It is important to point out that the VILLA input is much more varied than the target structure examined in this work, which is but one of the many grammatical constructions included in the course. Said otherwise, VILLA is not a psycholinguistic experiment devoted exclusively to transitive structures, in which the input was only meant to provide the learners with the necessary examples. On the contrary, from the learner perspective the VILLA course was primarily a language course (albeit with a few peculiarities, such as the prohibition to take notes: see below) containing a wide variety of lexical items and grammatical structures, some of which were later tested in some of the project tasks.

In each country except Germany (where child and adult acquisition were compared), the same contents were organised into two versions of the input, namely meaning-based (MB) and form-based (FB). Although only the results relative to the former (including the German adult group) are discussed in this work, it is worthwhile to briefly describe both types of input so as to highlight the main differences, designed in order to pursue research questions concerning the factors influencing input saliency.

The purpose of the MB input is to avoid drawing the learner's attention on any particular feature of Polish. Figure 2.1 shows a typical slide from the MB input course focussing on transitive structures such as *Eryk lubi kawę* 'Eryk.NOM likes coffee.ACC'. As can be seen, no written word-forms or hints of any kind as to the target structure (the accusative case) are presented. Learners of the MB editions could only rely on their own processing skills in order to identify and organise any formal regularities of the aural and written input.

In contrast, the FB groups were exposed to enhanced input (Sharwood-Smith 1993), designed to highlight specific formal features of the input. This was mainly achieved through focus-on-form activities (Doughty & Williams 1998) and corrective feedback. Linguistic information was visually presented in a more explicit fashion, as exemplified in Figure 2.2.

The slide focuses on transitive constructions, but the written word-forms of target items are presented on screen and graphically highlighted (the −ę ending, here circled in red) together with other relevant information, such as the grammatical gender of the referent (the "feminine" icon, also circled in red). Thus,
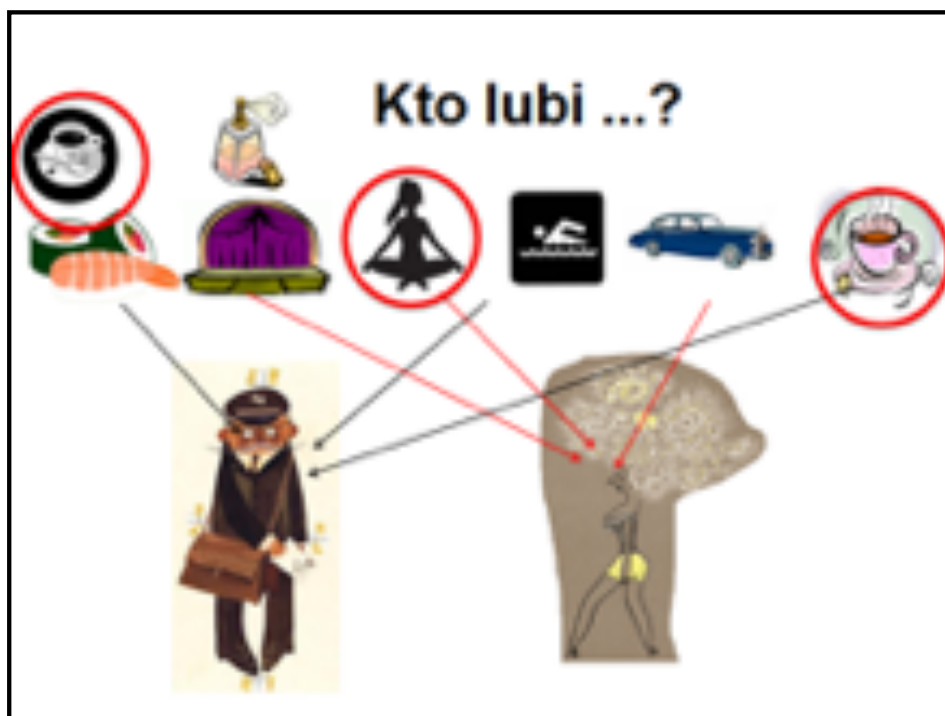
Figure 2.1: MB input example slide

learners are provided with a complete set of meta-linguistic information. Nevertheless, even the FB input includes no metalanguage and students were never provided with explicit *explanations* as to the target grammar.

Beside these differences, the input contents were the same for both groups. In addition ot the range of lexical items and grammatical structures, their frequency and order of appearance in particular were kept as uniform as possible.

2.1.1 Input control

As stated in the introduction, one of the major objectives of the VILLA project was a detailed, fine-grained analysis of the relation between input and intake. It is clear that thorough input control is an essential prerequisite in order to pursue this research question. In the VILLA project this ambitious objective required several steps.

First, one had to control for the participants' existing experience of the target language. This was only possible through the exclusion of participants who

Figure 2.2: MB input example slide

declared any previous contacts with Polish or other Slavic languages. Second, it was necessary to make sure that exposure to Polish throughout the course would be limited to the experimental input, including teacher speech and a set of Power-Point slides. The choice of Polish as a target language made it rather unlikely that participants could accidentally be exposed to it outside classes. In any case, all participants signed a contract to the effect that they would not intentionally look for additional information on Polish. Clearly it is impossible to verify whether or not this requirement was respected.

In order to make the experimental input as uniform as possible in terms of both quantity and quality, learners were asked not to take notes during classes. The rationale behind this decision is that the quality of learners' notes as well as their effort would very likely differ from person to person, thus introducing an undesirable variable beyond methodological control. For the same reason, no homework was assigned and individual practice outside classes was discouraged.

Finally, input was carefully planned in terms of topics, vocabulary and frequency of both lexical items and syntactic structures. This resulted in a general

scheme which the teacher replicated with remarkable accuracy throughout the various editions of the course. Classes would never be identical to each other, as they were not recorded but performed live. Nevertheless, since frequency was one of the variables controlled for in the tests, it was vital that the target linguistic items should occur an equal number of times in each edition. To this aim, classes were monitored in real time by a team member, who signalled to the teacher whether a given word had appeared too rarely or too often in relation to its planned frequency. An *a posteriori* analysis on the input corpus showed that lexical frequencies were 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.

2.1.2 Input transcription

By far the most thorough tool of input control was input recording and transcription, so that it can now be accessed and studied in a written format. The teacher wore a portable wireless microphone which recorded her speech. The resulting tracks were subsequently transcribed[1] 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 2.2). Transcription is produced from left to right along the horizontal axis; participants are assigned different tiers which are listed from top to bottom.

   To separate input addressed to all learners from comments aimed at individual learners or groups during interactional games, teacher speech was transcribed on two different tiers, labelled *TEA and *TEB respectively. Since it represents the vast majority of utterances, only the former is considered in this work. The ELAN files were converted into a text-like top-bottom format using the CHAT/CLAN (MacWhinney 2000) suite of software (Figure 2.1).

   A CHAT-CLAN automatic morphological tagging system was developed by Christine Dimroth and Roman Skiba, with a small contribution by the present author (Figure 2.2).

   On the dependent tier %mor, each word is morphologically tagged with the appropriate values of the relevant grammatical categories, depending on the word class 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

---

[1]The input for Italian and English editions was transcribed by Jacopo Saturno; the French, German and Dutch editions were transcribed by members of the corresponding research teams.

2 The VILLA Project: Methodology
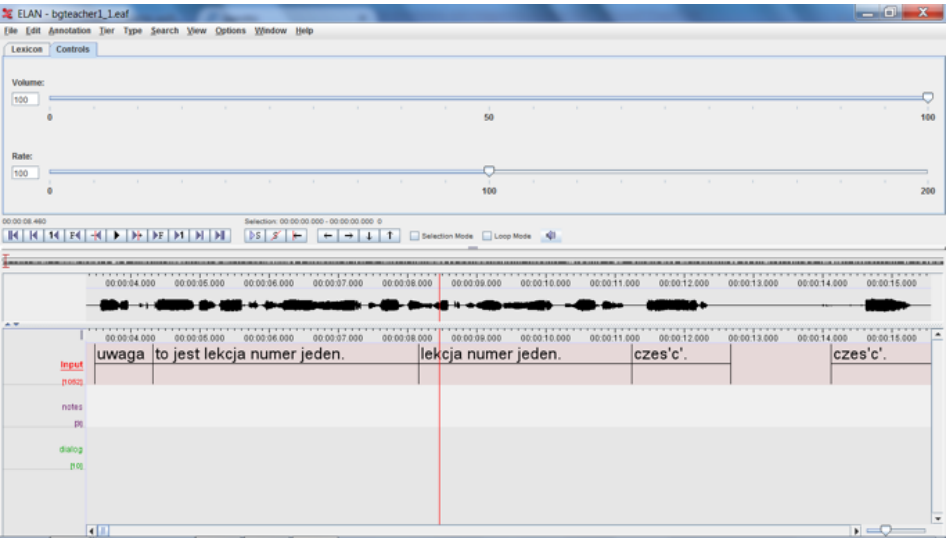
Figure 2.3: transcription of the VILLA input with ELAN

Table 2.1: transcription of the VILLA input in CHAT

```
6    *TEA:   uwaga
7    *TEA:   to jestlekcjanumerjeden .
8    *TEA:   lekcjanumerjeden .
9    *TEA:   cześć .
10   *TEA:   cześć .
11   *TEA:   cześć .
12   *TEA:   cześć .
13   *TEA:   bardzodobrze .
14   *TEA:   cześć .
15   *TEA:   to jestkontaktnieoficjalny .
16   *TEA:   kontaktnieoficjalny .
17   *TEA:   cześć .
18   *TEA:   cześć .
```

Table 2.2: automatic morphological tagging of the VILLA input in
CHAT

```
*TEA:   uwaga
%mor:   Noun|uwaga:Nom-fem-Sg=attention
*TEA:   to jestlekcjanumerjeden .
%mor:   Pron|ten:demon-3-Sg-neutr-Nom=this Cop|być:3-Sg=be
        Noun|lekcja:Nom-fem-Sg=lessonNoun|numer:Nom-masc-Sg=number
        Num|jeden:cardinal=one .
*TEA:   lekcjanumerjeden .
%mor:   Noun|lekcja:Nom-fem-Sg=lessonNoun|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 .
```

CLAN Mor programme associates the appropriate gloss, retrieving it from a spe-
cially designed lexicon. In case a given form corresponds to more than one tag,
which because of widespread morphological syncretism is a fairly common case
in Polish, all tags are presented subsequently. Glosses were not disambiguated
in any way.

Building on that basis a similar yet separate system was developed for the
purposes of the present work by adapting the same principle to a different tool,
namely the software R (R Core Team 2017) and its package *stringr* (Wickham
2017). New labels (in Italian) were also devised for all grammatical categories,
such as verbs, pronouns and adjectives. Compared to the system presented above,
this new version facilitates frequency searches for morphosyntactic patterns in
several technical respects.

Again, tags may include more than one possible grammatical meaning, as ex-
emplified in (1).

(1)   *balonik*:sostantivoIN_Acc_Mas_Sg//sostantivoIN_Nom_Mas_Sg:balonik

The input transcript, once glossed, can be searched for appropriate patterns
through regular expressions. A search for SVO sentences with animate masculine
nouns as subject and inanimate feminine nouns as object, for example, should
retrieve hits like *Leon lubi herbatę*, 'Leon likes tea'. In addition, it is possible to
identify all the instances of a given lexeme or grammatical value such as, for ex-
ample, 'nominative masculine singular'. This possibility will be widely exploited

in Chapter 3, in which searches were performed using R (R Core Team 2017) and its package *stringr* (Wickham 2017).

2.1.3 The VILLA input as a variety of Polish

While the VILLA project uses a natural language as input, it would be imprecise to claim 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. These constraints result in a language variety which at times may sound a little odd to a native speaker of Polish.

First, the dramatic competence gap between the teacher and the total beginner learners often results in a register definable as Teacher talk (Larsen-Freeman & Long 1991: 134–144) whose purpose is to simplify the input as much as possible while maintaining grammatical correctness. In fact, teacher speech within the VILLA project was extremely slow and hyperarticulated in an effort to make input more salient, i.e. more easily perceivable and segmentable.

Second, the choice was made to focus on a limited number of target structures, whose acquisition was later probed through the linguistic tasks. This caused them to be often produced with unnatural frequency, as is the case for the copula verb *jest* 'is'. Further, the frequency of syntactic structures directly conditions the frequency of the inflected word-forms belonging to the paradigm of a word. Compared to native varieties, input manipulation results in only a limited number of word-forms being represented in the VILLA input: plural forms for instance are completely absent for most nouns. Even within the singular number, the VILLA input is much more restricted than any L1 variety, being limited to only a couple of forms. Depending on the type of noun, the input might focus on the opposition between NOM and INS, for animate nouns, or between NOM and ACC, for inanimate ones. While most nouns only occurred in one or two forms, some did show a greater range of morphological endings.

Copular structures usefully illustrate another source of deviations from native varieties, namely pragmatics. Two main types of predicational copular clauses may be distinguished in Polish (Bondaruk 2013): In NOM-type structures (following the labels introduced in **Saturno2015**), the invariable pronoun *to* 'this' is supplied independently of referent gender and number, while the complement appears in the nominative form, e.g. *to jest Filip* 'this is Filip'. INS-type structures, in contrast, require the personal pronoun *on* 'he' or *ona* 'she', which specify the

gender of the corresponding referent, while the noun is provided in the instrumental case, e.g. *on jest studentem* 'he is a student'.

In native varieties of Polish, the two structures are pragmatically differentiated. NOM-type structures are mainly used deictically in order to introduce new referents in the discourse, e.g. *a to, kto to jest?* 'and this [person], who is this [person]?'. In contrast, the personal pronouns of INS-type structures typically refer to entities in an anaphoric manner, which means that the referent is already part of the discourse: the copular structure is used to provide additional details, e.g. *kim on jest?* 'who is he? [what's his job/nationality etc]?'. In contrast, in the VILLA input the two structures are used quite interchangeably in all contexts, so that no functional differentiation applies. Example (2), 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 (rhetorically) who Karol is using a an INS-type copular structure (2a), which calls for the same structure in the responses in (2b) and (2c). However, in (2d) 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, it would sound pragmatically inappropriate to a native speaker of Polish.

(2)   a. A   Karol,   kim   jest Karol?
         and Karol.NOM who.INS is   Karol.NOM?
         'And Karol, who is Karol?'

      b. Karol   jest strażakiem.
         Karol.NOM is   fireman.INS
         'Karol is a fireman.'

      c. On jest strażakiem.
         he is   fireman.INS
         'He is a fireman.'

      d. To   jest strażak.
         this is   fireman.NOM
         'This is a fireman.'

Similar structures were produced intentionally for didactic purposes, i.e. in order to show learners that predications about referents may be expressed through different syntactic constructions. This contrast was also the target structure of several tasks. Most VILLA research questions are concerned with morphosyntax, which in light of the constraints imposed by a first exposure study necessarily led

to a partial neglect of pragmatics and information structure. While this state of things seemed inevitable for methodological reasons, it has two negative consequences: first, as mentioned, the input at times might seem unnatural to a native speaker of Polish; second, the two contrasting structures end up to express exactly the same meaning, so that the contrast loses any functional motivation.

Similar arguments may be made concerning a crucial point of the present work, namely the order of subject and object in transitive sentences. Although both SO and OS orders are possible in Polish, this does not mean that different versions carry an identical meaning, at least from a pragmatic point of view (Siewierska 1993). In general terms, 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 to treating the subject as the focus and the object as the topic. This is not necessarily the case in the VILLA input, as example (3) makes clear.

(3)   a.   Filip   wózek   ciągnie.
           Filip.NOM cart.ACC pulls

           'Filip pulls the cart.'

   b.   Filip   ciągnie wózek   tak.
           Filip.NOM pulls   cart.ACC yes.

           'Filip pulls the cart.'

   c.   Wózek   ciągnie Filip.
           cart.ACC pulls   Filip.NOM

           'Filip pulls the cart.'

Compared to L1 practice, several facts are a little odd. First, the same referent is verbalised three times using maximally explicit means such as a person name. Second, based on the pragmatics of native Polish, one should conclude that (3a) focuses the verb, as in 'Filip doesn't push the cart, he pulls it'; (3b), at least in the absence of specific intonational patterns, should be interpreted as unmarked; in (3c), finally, the focus would be on the subject, as in 'it's not Julia who pulls the cart: it is Filip'. However, none of these interpretations is warranted in the example in question.

In fact, this obvious manipulation of syntax served the sole purpose of showing the learners that Polish word order is free. Unfortunately, given the very limited scope and vocabulary range of the VILLA project, at times it was impossible to do so in a pragmatically meaningful way, and the important link between information structure and syntax had to be sacrificed. On the one hand, word order

manipulation 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?". On the other hand, it could be argued that the purpose of word order manipulation, i.e. the expression of pragmatically marked meaning, could not be adequately inferred from the examples provided.

In spite of these small differences with native varieties, it would be inappropriate to state that the VILLA Polish differs significantly from native varieties, or that it is not a natural language. The VILLA input retains numerous idiosyncrasies which are likely to create difficulties to the learners, and at the same time resemble the difficulties often encountered in SLA. Alongside a few instances of *pluralia tantum*, several nominal paradigms present a range of idiosyncrasies and specificities which may not be immediately easy for the learner to grasp, such as the animacy-based differential object marking found in the paradigm of masculine nouns (see next section). To summarise, the VILLA input represents a specific variety of Polish, which retains the complexities and idiosyncrasies of a natural language, although it does present a limited range of lexical and grammatical items as well as a few instances in which syntax and pragmatics were somewhat bent to the need of didactics.

2.2 The target language and the VILLA L1S

2.2.1 Inflectional morphology

In addition to its low availability outside the language classroom in the countries where the VILLA courses were held, Polish was chosen as the target language of the experiment because it typologically differs from the VILLA L1s in various respects. One key feature is its rich and complex system of nominal morphology, contrasting two numbers, three genders in the singular and two in the plural, and crucially, as many as seven cases. Table 2.3 shows the paradigms which appear in the VILLA input. Virtually all nouns appeared in the singular number only, so much so that plural forms may be considered exceptional (they are mainly limited to a few *pluralia tantum*) and will not be considered in this volume.

The selection of the ACC ending depends on the interaction of animacy and grammatical gender, defined by noun-adjective agreement and NOM ending. Animacy is not relevant in the case of neuter and feminine nouns (4a and 4b), but it determines whether the ACC of masculine nouns is identical to the NOM in non-palatalized consonant, as for inanimate nouns (4c), or, in the case of animate nouns (4d), to the genitive in -*a* (4e).

2 *The VILLA Project: Methodology*

Table 2.3: Polish nominal paradigm, singular

	M ANIM	F ANIM	M INANIM	F INANIM	NEU
NOM	strażak	mam-a	balonik	kaw-a	biurk-o
GEN	strażak-a	mam-y	balonik-u	kaw-y	biurk-a
DAT	strażak-owi	mam-ie	balonik-owi	kaw-e	biurk-owi
ACC	strażak-a	mam-ę	balonik	kaw-ę	biurk-o
INS	strażak-iem	mam-ą	balonik-iem	kaw-ą	biurk-iem
LOC	strażak-u	mam-ie	balonik-u	kaw-ie	biurk-u
VOC	strażak-u	mam-o	balonik-u	kaw-o	biurk-o
	'fireman'	'grandmother'	'balloon'	'coffee'	'desk'

(4) a. Jan-∅   lubi kaw-ę
         Jan-NOM likes coffee.F-ACC

     b. Jan-∅   kocha mam-ę
         Jan-NOM loves sister.F-ACC

     c. Jan-∅   ma balonik-∅
         Jan-NOM has balloon.M-ACC

     d. Jan-∅   zna   strażak-a
         Jan-NOM knows fireman.M-ACC

     e. to   jest samochód-∅ strażaka
         this.NOM is   car-NOM   fireman.GEN

Syncretism may obscure the one-to-one pairing of form and function. Even in a subset of Polish as is represented in the VILLA input, the inflectional ending -a expresses at least two types of grammatical meaning in the nominal domain, i.e. NOM.F and GEN/ACC.M (the latter for animate nouns only), in addition to the 3sg of some verbs (e.g. *on zna* 'he knows'). Whenever instances of such categories co-occur, they present the same ending (5). It follows that in order to apply the morphosyntactic principle of utterance decoding, the listener needs to be aware of the grammatical gender of the nouns involved as well as their inflectional paradigm.

(5) siostr-a   lubi brat-a
      sister-NOM likes brother-ACC

In terms of inflectional morphology, the VILLA L1 closest to Polish is certainly German. This is the only L1 to express case on nouns, while all others only distinguish case on some personal pronouns. However, case in German is mainly signalled on the determiner (completely absent in Polish), while the inflectional paradigm is characterised by diffused syncretism, whereby several functionally differentiated word-forms are formally identical (Table 2.4). Indeed, it has been shown (Kempe & MacWhinney 1998; 1999 for Russian) that the Slavic case is a much better cue to the identification of the sentence agent than it is in German, for both native speakers and learners.

Table 2.4: German nominal inflection

		M	F
SING	NOM	der Feuerwehrmann	die Köch-e
	GEN	des Feuerwehrmann-s	der Köch-e
	DAT	dem Feuerwehrmann	den Köch-en
	ACC	den Feuerwehrmann	die Köch-e
PLUR	NOM	die Feuerwehrmänn-er	die Köchinn-en
	GEN	der Feuerwehrmänn-er	der Köchinn-en
	DAT	der Feuerwehrmänn-ern	den Köchinn-en
	ACC	die Feuerwehrmänn-er	die Köchinn-en
		'fireman'	'cook(F) '

The four remaining L1s, namely Dutch (Table 2.5), Italian (Table 2.6), French (Table 2.7) and English (Table 2.8) lack case altogether as far as nouns are concerned. These only present two forms, corresponding to singular and plural.

Table 2.5: Dutch nominal inflection

Singular		Plural	
M	F	M	F
de kok	de kokkin	de kok-s	de kokkin-s, kokkinn-en
'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: the only morphological cue to number is the determiner (Table 2.7).

Table 2.6: Italian nominal inflection

	Singular		Plural
F	M	F	M
il pompier-e	la cuoc-a	i pompier-i	le cuoch-e
'the fireman'	'the cook.ғ'	'the firemen'	'the cooks.ғ'

Table 2.7: French nominal inflection

	Singular		Plural
M	F	M	F
le pompier	*la cuisinière*	*les pompiers*	*les cuisinières*
/lə pɔ̃.pje/	/la kɥi.zi.njɛʁ/	/le pɔ̃.pje/	/le kɥi.zi.njɛʁ/
'the fireman'	'the cook.ғ'	'the firemen'	'the cooks.ғ'

Gender is not morphologically encoded in English nouns, which only distinguish a singular and a plural form. Even when a noun is characterised in terms of intrinsic sex, this category is only visible through anaphoric reference.

Table 2.8: English nominal inflection

	Singular		Plural
the fireman	*the cook*	*the firemen*	*the cook-s*

As far as word order is concerned, Polish is a predominantly SVO language (Dryer 2013b), although the encoding of case makes it possible to freely manipulate it for pragmatic purposes. Word order in German may also be manipulated for pragmatic reasons without the need for specific syntactic devices (such as cleft sentences), because syntactic functions are made explicit by case marking. It is worth noting that, unlike Polish, German word order is constrained by the obligatoriness of the finite verb in second position.

In all VILLA L1s except German, the lack of explicit case marking results in syntactic functions being assigned based on the default SO word order. Departures from this pattern (6a) are possible, but require marked phonological or syntactic means, particular intonational contours, cleft sentences (6b) or dislocations (6c).

Of course, all these possibilities are also available in languages with more complex morphology, like Polish (6d). Word order manipulation in these languages is simply an extra resource to explicitly mark information structure (6e), but this does not immediately translate into their being more flexible, or favouring more variable word orders.

(6) a. il   gatto insegue il   cane
         ART.M.SG cat chases ART.M.SG dog

     b. è il   gatto che insegue il   cane
         is DET.M.SG cat   REL chases ART.M.SG dog

     c. il   cane lo   insegue il   gatto
         ART.M.SG dog PRO.ACC chases ART.M.SG cat

     d. to kot   goni   psa
         it cat.NOM chases dog.ACC

     e. psa   goni   kot
         dog.ACC chases cat.NOM

Because of the lack of articles (Dryer 2013b), however, word order manipulation in Slavic languages is one of the main means to express definiteness (Jacennik & Dryer 1992; Siewierska 1993).

As far as the lexicon is concerned, although numerous words belonging to international lexicon may be found, most Polish vocabulary is of Slavic origin and therefore fairly opaque to the VILLA learners. Lexical stress is fixed on the penultimate syllable, with the partial exception of learned loanwords from Latin or Greek as well as elements to which clitics are attached (both virtually absent from the experimental input). Among the VILLA L1s such rigid pattern is only found in French, where the stress is fixed on the last syllable of the intonation unit.

2.3 The learners

Choosing an "exotic" language is surely a necessary step to run a first exposure study, but equally important is to make sure that the learners never had any experience of it. To this purpose, candidates to the VILLA project were first asked to fill 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 languages in which case is expressed morphologically, such as Greek, Latin or even German were also excluded. The reason for this is

that the ideal VILLA participant is a linguistically naïve speaker of any of the VILLA L1s, who (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 some languages, in contrast, inevitably implies some familiarity with this category. This might result in learners with that kind of experience processing Polish morphosyntax not just based on the input provided during the course, but rather thanks to their previous language skills. Fulfilling this criterion was particularly difficult in Italy and Germany, as many secondary school students take at least a year of Latin (Table 2.9).

Table 2.9: distribution of Latin skills by L1

| | L1 group | | | | | |
Latin	Fr	GE	IT	NL	UK	Total
Y	2	8	14	-	0	24
N	15	12	3	-	17	47
NA	-	-	-	20	-	20

The candidates who were selected based on their language profile further took a "language sensitivity test", in which they heard sentences in Polish, Russian and Finnish, and were asked whether or not they thought the sentences were in Polish. This was done in order to exclude people whose "intuition" appeared to be too good and could thus bias the results of the experiment.

The selection process took place identically in the five countries which participated in the initiative. For each L1 group, Table 2.10 reports the total number of learners who took part in the VILLA MB course as well as their distribution by sex and the group mean age. Because of occasionally missing data, slight discrepancies in the total number of participants considered in the analysis may occur throughout this book.

The vast majority of the VILLA participants were university students enrolled in a variety of degrees. Students of foreign languages, linguistics and psychology were excluded in order to avoid any potential bias related to a greater familiarity with the study of languages or the rationale of psycholinguistics experiments.

2.4 Learner data collection

The present section lists and describes the tasks which were used in the VILLA project to elicit linguistic data, with a particular focus on the tools which will be

Table 2.10: learners by L1, MB group

	EN	FR	GE	IT	NL	Total
n.	17	17	20	17	20	91
sex: F	13	13	12	10	9	57
sex: M	4	7	5	10	8	34
mean age	22	23	24	22	21	22

discussed in this book.

The main tool to elicit L2 data is represented by structured tests, schematically listed in Table 2.11. Although the focus is clearly on morphosyntax, several other levels of language were targeted as well.

Table 2.11: VILLA project, linguistic tasks

Test	Target
Phoneme Discrimination	phonology
Word Recognition	phonology
Lexical Decision	vocabulary
Grammaticality Judgement I (NOM/INS)	morphosyntax
Picture Production	morphosyntax
Comprehension (NOM/ACC)	morphosyntax
Elicited Imitation (NOM/ACC)	morphosyntax
Grammaticality Judgement II (NOM/INS)	morphosyntax
Written word order	syntax
Cloze test (Pronouns)	pro-drop
Route direction	free production: reference to space
Finite story (Dimroth 2012)	free production: film retelling

This book considers the results of two of these tasks, i.e. the Elicited Imitation (EI) and the Comprehension tasks. Precious linguistic data may be also extracted from a few interactional moments which took place during classes, in which learners were asked solve a simple communicative task in pairs or small groups. Indeed, one such instance will be used a source of semi-spontaneous production data in Chapter 7.

In order to control for specific learner attitudes which may have an impact on the tests above, individual difference measures were also taken (Table 2.12).

Although correlations of these measures with the results of the structured tests have been attempted within the VILLA project (Watorek & Saturno 2016; Saturno & Watorek 2020), these tasks will not be considered in the present work.

Table 2.12: 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

2.4.1 Transcription of the production tasks

Two of the sources of linguistic data described in this book, namely the EI test (§2.4.3) and semi-spontaneous production (§2.4.5), aimed to elicit oral production data from the learner. As these data needed to be made available in a written mode for further analysis, learner responses were digitally recorded and subsequently transcribed[2] using a combination of the ELAN (Brugman & Russell 2004) and CLAN (MacWhinney 2000) software. As explained in Chapter 1, in a first stage the production data were transcribed phonetically, using either the IPA (Landau et al. 1999) or the SAMPA (Wells 1997; 1995) phonetic alphabets. In preparation for analysis, the transcripts made by individual transcribers were then normalised to broad IPA. No effort was made to accurately transcribe a few subtle phonological contrasts of Polish phonology, such as that between the series of post-alveolar {ʃ,ʒ,ʧ,ʤ}[3] and pre-palatal {ɕ,ʑ,tɕ,dʑ} consonants, or that between the high front /i/ and the high central /ɨ/ vowels, because there are not

---

[2]Transcribers: Joanna Hinz (German corpus), Katarzina Loziczka (Dutch corpus), Jacopo Saturno (Italian, French and English corpora).

[3]These sounds are sometimes referred to as retroflex consonants and accordingly transcribed as {ʂ, ʐ, t͡ʂ, d͡ʐ}, although it can be argued that the notion of "retroflex" is quite problematic and

relevant for morphological analysis. When lexical stress is not specified, it is assumed that it falls on the penultimate syllable.

The production of centralised vowels by some learners required particular attention, as such sounds make it impossible to link the learner-produced form to one of the possible target word-forms. A word pronounced by a learner as [ˈpiwkə] for instance may correspond to both target [ˈpiwka] *piłk-a* 'ball-NOM.SG' and target [ˈpiwka] *piłk-ę* 'ball-NOM.SG'. For the present analysis target items in which the ending produced was not clearly identifiable as either -/a/ or -/e/ were discarded, which led to the exclusion of 334 items. The problem proved particularly severe in the case of the German data, in which 242 items were excluded, most probably because of transfer of the German phonological rule of word-final vowel centralisation.

Translations of learner output are not provided because of the difficulty to univocally ascertain what the learner really meant. Similarly, the gloss only indicates the input word-form which is closer to learner output, with no assumption that the indicated form was indeed the intended meaning.

2.4.2 Data analysis and visualisation

The data were entered into a spread-sheet format either manually (comprehension task) or semi-automatically using the export tools of ELAN (EIT and semi-spontaneous production). Descriptive and inferential statistics were then computed using the software R, version 3.6.0 (R Core Team 2017; generalised mixed linear models were fitted thanks to the package lme4, version 1.1-21 (Bates et al. 2015), while *stringr*, version 1.4.0 (Wickham 2017) proved essential for string manipulation. Figures were produced using the tools of base R as well as the packages *wordcloud*, version 2.6 (Fellows 2014) and *extrafont*, version 0.17 (Chang 2014).

2.4.3 The Elicited Imitation Test (EIT)

The VILLA EIT is a highly structured task which learners took on two occasions, namely after 9 hours of exposure to the input (T1) and after 13:30 hours (T2). It was administered individually on a computer screen; depending on the course edition, headphones or the integrated computer speakers were used.

---

may corresponds to different phonetic realisations across the languages of the world (Hamann 2002; 2003; 2004; Zygis 2003; Żygis & Hamann 2003; Padgett & Żygis 2007; Żygis & Padgett 2010). Throughout this book the symbols {ʃ,ʒ,ʧ,ʤ} will be used for reasons of readability.

The task was structured as follows. First, learners heard a short Polish transitive sentence, e.g. *dziewczynka ciągnie portugalkę*, 'little girl-NOM pulls Portuguese woman-ACC'. Subsequently, participants were required to draw on a separate answer sheet a simple geometric figure (exemplified in Figure 2.4) which appeared on screen.
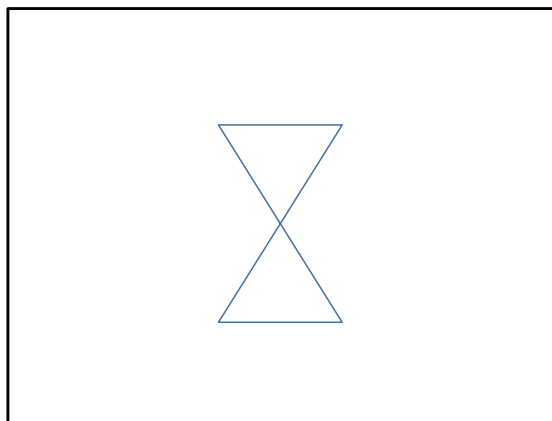
Figure 2.4: EI task, distractors

This step was included in the task in order 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. It should be noted here that the drawing task didn't involve articulatory suppression that might have disrupted subvocal rehearsal. Finally, learners were asked to repeat the target sentence as accurately as possible. Learner performances were not timed and no explicit time pressure was exerted.

All target sentences were 9 syllables long and had the structure Noun - Verb - Noun. Throughout the test, the two nouns always appeared in association with each other. 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'. While lexical transparency will not be considered in the analysis of the L2 data, this factor has been addressed in other works related to the VILLA project (Hinz et al. 2013; Saturno 2014; Rast 2015).

All targets were digitally recorded by the same female speakers. They were uttered with a slow speech rate and neutral intonation, so as to avoid any potential hints as to the pragmatic interpretation of the sentence.

All target nouns belonged to the paradigm of the feminine nouns in -*a*[4]. Each appeared in both the NOM and ACC case, instantiated by the endings [a] (<a>) and [e] (<ę>) respectively. Target sentences also varied with respect to constituent order, which could assume the values SVO or OVS. Since only the relative order of subject (S) and object (O) is relevant to the present analysis, henceforth SVO and OVS will be referred to as SO and OS, respectively, unless explicitly stated otherwise.

To summarise, each pair of nouns appeared in four target sentences, which makes it possible to isolate the parameters of case ending, word order and lexical transparency (Table 2.13). As there were 4 pairs of target nouns, the test included a total of 16 target sentences.

Table 2.13: EI task, target sentences

	SO	OS
NT - T	*dziewczynk-a woła portugalk-ę* little.girl-NOM calls portuguese.woman-ACC	*dziewczynk-ę woła portugalk-a* little.girl-ACC calls portuguese.woman-NOM
T - NT	*portugalk-a woła dziewczynk-ę* portuguese.woman-NOM calls little.girl-ACC	*portugalk-ę woła dziewczynk-a* 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 2.14.

The values assumed by the three parameters just discussed may combine in eight possible contexts (Table 2.15).

The test also included 16 filler sentences in the form of copular 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 were included in the task to make sure that all learners had correctly understood the procedure.

2.4.3.0.1 Theoretical premises   Unlike the Comprehension test and the spontaneous production task described later on in this chapter, the EIT requires a thorough discussion of its theoretical premises and underlying mechanisms. The

---

[4]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 2.14: EI task, parameters of obligatory occurrences

Kucharx-	/e/	*woła*	*Brazylijk-*	/a/	
cook	ACC	call 3SG	Brazilian woman	NOM	
	-/e/			-/a/	Target ending
	OS			OS	Constituent order
	NT			T	Lexical transparency

'the Brazilian woman calls the cook'

Table 2.15: EI task, 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

reason for this is that although it clearly is a highly structured test, it is often used (and indeed, it is used in this volume) as an approximation of spontaneous speech. Also known as "sentence imitation" or "sentence repetition", the EIT is a language assessment method whereby participants are asked to listen to a target sentence and repeat it as accurately as possible, usually after some distracting pause. The rationale underlying this procedure is effectively summarised by Buck (2001: 79):

"Although sentence repetition tasks work through listening, they require more than just listening skills. [...] As the sentences get [...] longer, it seems likely that chunking abilities and the ability to deal with reduced redundancy will begin to become more important and, as with dictation, these are closely related to general linguistic competence. [...] They are integrative tests in that they test the ability to use language rather than just knowing about it, but only as long as the segments actually challenge working memory capacity. And they do require speech production."

Said otherwise, test-takers can accurately repeat only the grammatical structures that are already part of their developing L2 grammar, here interpreted as the ability to identify "chunks" of language, which are stored in working memory not as mere strings of sounds, but as meaningful units which are subsequently re-encoded based on the present state of the interlanguage grammar. The task has

been successfully used to investigate the implicit competence of a wide range of populations, including L1 and L2 learners (both literate and illiterate, the task being administered orally) as well as patients affected by speech pathologies (Marinis & Armon-Lotem 2015; Armon-Lotem & Meir 2016).

On the practical side, the EI task offers numerous advantages to both language scientists and, to some extent, language testers (Brown & Abeywickrama 2010: 187–189): it is relatively quick and easy to administer, requires little equipment, and offers full control over the target structure (Van Moere 2012). This last point is certainly 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 will surface at all (Ferrari & Nuzzo 2009; Bettoni & Di Biase 2015). For instance, eliciting the OS structures on which the present work is based would imply waiting for the learner to spontaneously produce a structure which requires a certain degree of grammatical competence as well as the appropriate pragmatic context, in addition to the learner's intention to exploit it. In other words, even if test takers can be thought to be equipped with the required grammatical and pragmatic competence, there is no warranty that they will use it even in the appropriate context, simply because they are in control of their own speech. Things may be forced to a little extent, for instance by using tasks which make a specific structure particularly appropriate in a given context: for example, one 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 (7):

(7)   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 available to the learners. Speakers might just emphasise the subject pronoun using intonation, thus maintaining the default SO word order, or indeed produce an elliptic answer like *ja*, 'I (will)'. In sum, although (semi)spontaneous speech is undoubtedly the most authentic (Lewkowicz 2000) 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 obtained using this elicitation approach, it certainly is less practical than other elicitation techniques.

2 *The VILLA Project: Methodology*

The EIT is typically administered orally, ensuring that both the prompt and the repetition are produced at a rate close to that of spontaneous speech and, therefore, that only implicit competence can be accessed: test takers should not have the time to rely on explicit, declarative knowledge (Ellis 2005). This is a very important argument for those advocating the appropriateness of this task for the evaluation of the actual interlanguage state (Erlam 2006).

2.4.3.0.2 Memory in the EI test   Target sentence design is vital to ensure that test-takers cannot repeat target sentences *verbatim*. By relying on working memory (Baddeley 1986; 2003), in fact, it is usually possible to remember short strings of sounds for a handful of seconds and then repeat them with reasonable accuracy (Sachs 1967), even without necessarily understanding their meaning. This is due to the phonological loop (Baddeley et al. 1998), the device which makes it possible to mentally store and rehearse a chain of sounds for some time before it fades. While this mechanism is seen as crucial in language learning, it represents a serious methodological obstacle to the validity of EIT as a measure of implicit (i.e., automatized, non meta-linguistic) lexico-grammatical competence. A properly designed experimental protocol 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 block the exclusive reliance on phonological memory (Juffs & Harrington 2011). Advocates of the EI test argue that under these conditions participants can remember meaning and lexico-grammar, but not phonological forms, which have to be produced anew in repetition. The test becomes reconstructive in nature, as participants listen to targets, decode them, and then re-produce them on the basis of the current developmental stage of the interlanguage. Indeed, test-takers have been reported to systematically produce ungrammatical structures or, conversely, correct ungrammatical targets (Hamayan et al. 1977; Munnich et al. 1994). Håkansson (1989) states that up to a specific developmental stage, a three-year old Swedish child consistently reproduced a NEG-AUX structure instead of the target AUX-NEG structure of the L1. Such studies are interpreted as evidence that test-takers do not just repeat a string of sounds, but interpret and reproduce it "in their own way", often betraying the influence of factors such as markedness (as in Håkansson's study) or L1 transfer. In this perspective, the EI test makes it possible to investigate to what extent the learner is able to bypass the constraints which shape the developmental stages of language acquisition, such as for instance the first-noun principle.

From another perspective, Van Moere (2012: 325–326) builds on Skehan's (1998: 168) notion of "processing competence" to suggest that the EI test is particularly

apt to measure an under-researched, but vital skill such as processing efficiency, defined as "the speed and accuracy with which a learner orally processes familiar language", which in turn tends to "near effortless processing of language", or automaticity (DeKeyser 2001). Such position proposes a largely lexically-based view of language, whereby words tend to occur in meaningful chunks which the language user treats as a single unit (Pawley & Syder 1983; Ellis 2001) and indeed have been shown to be processed with greater efficiency by both native speakers and language learners (Conklin & Schmitt 2008). Within Construction Grammar (Gries & Wulff 2005; Hoffmann & Trousdale 2013) and usage-based approaches (Tomasello 2005; Cadierno & Eskildsen 2015; Tyler & Ortega 2016), chunks correspond to "constructions", "form-function mappings that are conventionalized as ways to express meanings in a speech community" (Wulff & Ellis 2018: 38), where meaning can vary greatly in its degree of abstraction (Goldberg 2006). The parallel is sometimes stated explicitly: "constructions can be viewed as processing units or chunks – sequences of words (or morphemes) that have been used often enough to be accessed together" (Bybee 2013: 51). Again, constructions are not seen as the product of rules, but as language units: "patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency" (Goldberg 1995: 5).

Against this background, the EI test is seen as a measure of acquisition in that it measures the learner's ability to repeat strings which are too long and complex to be stored in phonological memory, which is described as containing about 7 unrelated words or digits (Miller 1956) or two seconds worth of speech (Baddeley 1986). Effectively, it has been shown that test-takers perform much better when asked to repeat meaningful speech than non-words (Gathercole & Baddeley 2004). Other studies (Underhill 1987: 86; Buck 2001: 79) further showed that only the shortest targets can be processed as mere strings of sounds. Repetition of meaningful speech is only possible through chunking and processing for meaning (Radloff 1991: 9).

"As people become more familiar with a second language and more confident in manipulating its syntax, they are more able to pack the chunks full of information; and the more they control the morphology the better they are able to organize within chunks of syntax; and the more vocabulary they know the better they are able to hold on to the meaning until they can repeat the sentence."

As a result, "only test takers who have developed sufficient automaticity in processing linguistic information will perform successfully" (Van Moere 2012: 332).

The importance of processing for meaning as opposed to form is also underlined by Erlam (2006), who argues for the reconstructive nature of the test and

inserts a comprehension question as a pause to inhibit phonological memory. Her claim is based on Sachs's (1967) research, who demonstrated that while the exact lexical and morphosyntactic shape of a target sentence is lost soon after hearing it, memory for its general meaning lasts much longer.

2.4.3.0.3 Appropriateness of the EI test for language assessment Not all researchers would agree with the rationale just described, and the relation of the EIT with working memory is certainly complex. Many factors are thought to influence the engagement of working memory in the task, including, among others, the nature and length of the stimuli, the type of distractor, the target structure, the learner's proficiency level, and many more (see Vinther 2002 and Erlam 2006 for a review). For the present purposes, it is sufficient to say that some argue that the EIT has nothing to do with implicit linguistic competence, and only measures a learner's working memory capacity (Jessop et al. 2007), whereas others claim that working memory is only marginally involved if at all (Okura & Lonsdale 2012).

A good example of this debate is the controversy between Zhang & Lantolf (2015) and Pienemann (2015). Aiming to verify the Teachability Hypothesis (Pienemann 1984), Zhang & Lantolf exposed four English L1 learners of Chinese L2 to specially designed instruction. Learners were shown not only to be able to process structures deemed to be too advanced for their interlanguage, but also to skip developmental stages, which is excluded by Pienemann's Processability Theory. Pienemann (2015) questioned Zhang and Lantolf's results on various grounds, including their claim that they used the same elicitation methods and emergence criteria utilized in PT-inspired research: "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" (Pienemann 2015: 139). Indeed, a study by Pienemann et al. (2013) experimentally demonstrated that learners of L2 Swedish 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, while structures produced by the teacher, but beyond the learners' current developmental stage could only be repeated as unprocessed fixed formulas. 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

chapter, 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 in turn is defined as structures which are not copies of the previous utterance. 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 evidence that structures beyond the correct processability stage can indeed be repeated as formulaic items without being processed, hence Pienemann's scepticism with regard to the EIT. In their response, Lantolf & Zhang (2015) note that the method used to elicit repetition by Pienemann et al. (2013) is quite different from the typical EIT. Specifically, learners were asked to repeat teacher utterances straight after a stimulus sentence has been presented, whereas in Lantolf & Zhang's study (2015) they first had to perform a comprehension task. This is a sensitive point, as the design of EI tasks has been shown to have a direct and macroscopic impact on the kind of data it can produce.

The EIT was also criticised for its lack of authenticity by Chun (2006), who defines this construct as "the degree of correspondence of the characteristics of a given language test task to the features of a Target Language Use task" (Bachman & Palmer 2009: 23), i.e. to what extent the experimental task simulates a real communicative situation: "with task-based tests, the developers need to show that the content of the test tasks is representative of the demands of the corresponding task outside the test situation, and that the scoring reflects this" (Luoma 2004: 43). A commercial version of the EI task, Ordinate Corporation's PhonePass Spoken English Test–10 (now marketed as Versant by Pearson), is used as a test of a candidate's proficiency in spoken English in a variety of contexts, from job interviews to academic exams. Apart from the fact that the test is administered over the phone, Chun's (2006: 301) critique mainly targets two points. First, shorter target sentences may be repeated by parroting and do not necessarily test processing for meaning; quoting Buck (2001: 79), it is argued the task "might test no more than the ability to recognize and repeat sounds, and this may not require processing of the meaning at all. … [This] clearly fails Anderson's (1972) criteria for proof of comprehension". Second, target sentences are completely unrelated to any discourse or setting, so that the task does not reproduce a realistic communicative situation: "my interpretation of the speech production needed in the real-life domain of school and work necessitates the ability to create and interpret discourse by relating utterances to their meanings and intentions as well as the setting. Even a parrot can be taught to repeat short

sentences devoid of any meaning or context". This critique does not seem to take into account the "comprehensive body of psycholinguistic research that shows that this task does engage linguistic processing resources, and breakdowns in repetition performance in language learners occur in predictable patterns (e.g., misplaced grammatical morphemes, lexical substitutions, etc.; Ellis et al. 2006; Radloff 1991)" as indeed was pointed out by the test developers in their response (Downey et al. 2008: 163–164). Van Moere (2012: 330) also claims that the EIT is "more communicatively authentic than many people realize", citing in support a variety of arguments. First, speakers often tend to make their own speech similar to that of the interlocutor's in terms of both vocabulary and grammar (Levinson 1983: 313; Brown & Yule 1983: 89), which also plays an important role in the management of the interaction (Tannen 2007: 52). Further, it has been suggested that paying attention to the form of an interlocutor's speech may be advantageous in psycholinguistic terms, as the speaker is able to recycle that form and focus on the intended meaning (Swain 1985; Bygate 2001).

Similar heated exchanges of opinions indicate that the EIT often produces output which is interpretable in radically different ways. To minimise this risk, methodological rigour is essential.

Within the VILLA project, the issues summarised above were addressed as follows. First, all target sentences were of the same length (9 syllables). Second, each stimulus sentence was followed by a short distractor task, albeit of a non-verbal nature. Finally, learner WM store was controlled using Meara's (2005: 8–10) Llama D test, in which learners heard a target sentence followed by a shorter string of sounds and were asked to decide whether or not the shorter string was comprised in the target sentence. The words used in the target sentences are based on the names of flowers and natural objects in a British Columbia Indian language, synthesised using AT&T Natural Voices (French). It is thus highly unlikely that any VILLA learner should be able to process them for meaning.

The test is inspired by research by Service (1992; Service & Kohonen 1995) and Speciale et al. 2004, who argue that the ability to recognise repeated sound patterns may be beneficial for both word learning and the noticing of morphological variability.

In the present work the Llama D test will be mainly considered as a linguistically motivated test of phonological memory. Its output will be used to search for a positive correlation between WM store and success in the EIT, whose theoretical premises will be considered as validated only in the absence of said correlation. Indeed, if repetition accuracy were found to depend on WM store, it would be illegitimate to consider the EIT as a measure of morpho-syntactic skills, rather than mere phonological memory.

2.4.4 The comprehension test

In the VILLA Comprehension test, learners heard short Polish transitive sentences and subsequently saw two pictures in which the same two referents (a man and a woman) play different thematic roles: the same referent has the role of agent in one picture and of patient in the other one (Figure 2.5).
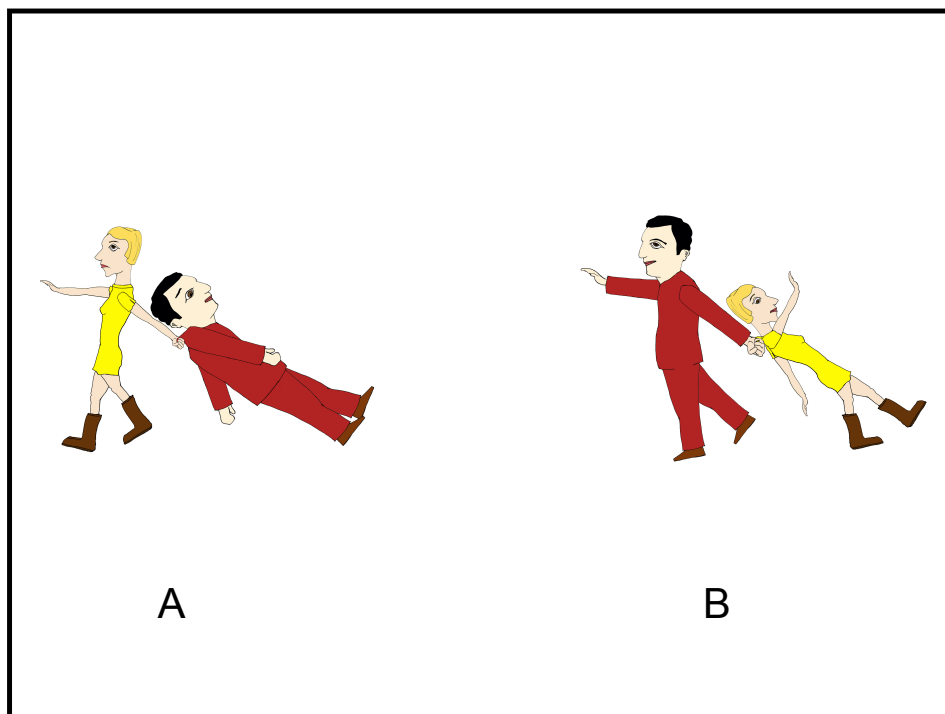
Figure 2.5: Comprehension test: alternative descriptions of the target utterance

The learners' task was to select the picture which in their opinion best depicted the stimulus sentence. Responses were marked in pen on an answer sheet. The data thus obtained were digitalised manually in spreadsheet format and then further manipulated and analysed with R (R Core Team 2017).

2.4.4.0.1 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 while pictures were projected on screen. Learners took the test after 9 hours (T1) and 13:30

hours (T2) of exposure to the input, consistently with the timing of the EI 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 had the structure NP - Verb - NP. Only two nouns were used for this test, namely *brat*, 'brother', and *siostra*, 'sister'. The verbs were the same employed in the EIT, namely *ciągnie*, 'pulls', *pcha*, 'pushes', *pozdrawia*, 'greets', and *woła*, 'calls'. Each noun appeared in both its NOM and ACC form; further constituent order varied (SVO, OVS, OSV), each occurring in eight target sentences.

Table 2.16: Comprehension test, example of target sentences with the verb woła

siostr-a	*woła*	*brat-a*	*brat*	*woła*	*siostr-ę*
sister-NOM	calls	brother-ACC	brother.NOM	calls	sister-ACC
siostr-a	*brat-a*	*woła*	*siostr-ę*	*brat*	*woła*
sister-NOM	brother-ACC	calls	sister-ACC	brother.NOM	calls
brat-a	*woła*	*siostr-a*	*siostr-ę*	*woła*	*brat*
brother-ACC	calls	sister-NOM	sister-ACC	calls	brother.NOM

Table 2.17 presents the relevant forms of the paradigm of the two target nouns employed in the test. *Brat* follows the declension of masculine animate nouns, *siostra* that of feminine nouns in -a.

Table 2.17: Comprehension test, paradigm of the target nouns

	brat 'brother'	*siostra* 'sister'
NOM	*brat*	*siostr-a*
ACC	*brat-a*	*siostr-ę*

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.

2.4.5 Semi-spontaneous production

An essential part of the VILLA experimental protocol consists in the monitoring of learner output during classes. To this end, participants were seated in front of

directional microphones which recorded everything they said during the whole course, as illustrated in Figure 2.6. The entire output of each participant was recorded on a separate track.

Figure 2.6: VILLA classroom set-up

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 using grammatical structures or vocabulary which had been previously practiced collectively with the teacher. The data presented in this book were collected during one such occasion, which took place during lesson 7.2, after roughly 10:30 hours of exposure to the input. The development of the interlanguage at that stage should be roughly comparable to that probed though the structured tests (the EIT and the Comprehension test) at T1 (lesson 6.2, 9 hours of input exposure). Given the amount of work required to prepare the raw production data for analysis, only a subset of the VILLA data-set could be analysed, i.e. the Italian MB input group.

Participants were divided into 7 groups of 2 and a group of 3. Each group was given a set of cards containing information about several referents that learners were asked to describe to each other. Each card in a learner's set only contained part of the 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 his or her card, the partner would try to

identify the character, asking questions to complete the missing data. 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.

For this study, the fragments relative to the interactive episode were extracted from the track of each participant and merged according to the 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. The data were then transcribed along the lines described in §2.4.1 and further divided into one-verb utterances. Synchronised video recordings proved of great help to identify participants, providing the transcriber with an additional clue, in addition to the sound of learners' voices. The resulting corpus comprises 60 utterances produced by 17 learners.

3 The Elicited Imitation Task

3.1 Research questions and hypotheses

While a more detailed discussion of the theoretical premises of the Elicited Imitation test (EIT) was given in Chapter 1, it is worthwhile to repeat here that according to its rationale, the task does not require learners to simply repeat a string of sounds, but rather to decode its meaning and re-produce it "in their own words", i.e. based on the present state of their interlanguage grammar. In this respect, the task is used in this work as an approximation of a productive task. Unlike in free production tasks, however, EI makes it possible to maintain full control over the target structure, which seems particularly urgent in the case of rare, non-obligatory targets such as the OS structures investigated in this work.

The present chapter describes the structure of the VILLA EI task and presents the results obtained by the various L1 groups of the project. The analysis aims to identify the impact that independent variable: word order in target sentence, input, and L1 exert on dependent variable: accurate repetition of inflectional markers. Based on the information presented in the introduction, the following hypotheses may be formulated regarding the effect of the variables taken into consideration.

- Word order: SO targets are expected to be processed with greater ease than the corresponding OS targets. This is due to the greater frequency of SO structures in the VILLA input as well as to their general unmarkedness from a pragmatic and typological point of view.

- L1: source languages in which case is expressed morphologically (such as German) should prove an advantage in the processing of a highly inflected L2 like Polish. Conversely, the rigid word order and lack of inflectional morphology of English and (spoken) French may slow down the acquisition process.

- Exposure to the input: greater length of exposure to the input (measured in hours) is expected to be beneficial to the development of the learners' morphosyntactic skills.

Finally, an attempt will be made to correlate the results of the EI task with the Llama D test of implicit influence of pattern recognition ability on phonological memory, in order to verify whether or not the learners' repetition performance is influence by this variable, whose involvement is typically excluded in the description of the rationale of the EI task.

3.2 Results

3.2.1 Overview of learner output: overall repetition accuracy

Overall repetition accuracy refers to the number of target segments which are correctly reproduced in a learner's response. For the time being, the analysis will exclude inflectional endings, which will be treated in greater detail in the next sections. From a strictly phonological point of view, case endings such as [a] and [e] are segments like any other in the target string, and could therefore contribute to a measure of phonological accuracy. On the other hand, in the target language they are also inflectional morphemes conveying grammatical information. According to the assumptions of the EI task, the learner *might* use them to derive and express syntactic functions, if the interlanguage has developed a morphosyntactic principle of utterance organisation: however, there are no *a priori* means to establish whether that is the case or not. In short, it is hard to tell if the distribution of case endings is due to reasons pertaining to phonology (because the learner heard them that way) or morphosyntax (because the learner wanted to express a given syntactic function using the corresponding case ending).

Turning to an overview of phonological accuracy, most of the times lexical items appear to be fairly recognisable. Linguistic material may be omitted (1) or substituted (2): compare the target in a. with learner output in b.

- (1) a. /ʤefʹʧink ʹʧoŋnie portuʹgalk/
  b. [ʤefʹʧink ʧon portuʹgalk]

- (2) a. /ʤefʹʧink ʹʧoŋnie portuʹgalk/
  b. [dyʹʧink ʹbardzo portuʹgalk]

In extreme cases, words may be mispronounced so badly that it is no longer possible to map them to existing word forms of the input, as [ʧyʒank] in (3). At times, learner output appears to combine bits of two or more words from the

input, as in (4)[1], in which the item [untʃeˈtʃelk] shows traces of the input words *dziewczyna* [dʑefʧina] 'girl' and *nauczycielka* [nauʧiˈtɕelka] 'teacher'.

(3) /studentke pxa nauʧitʃelka/
     [ʧyˈʒank ˈpxa untʃeˈtʃelk]

(4) /portuˈgalk ˈʧongnie dʑefʧink/
     [portuˈgalk ˈʧoni ʧefˈtʃelk]

3.2.2 Overview of learner output: repetition of case endings

The following transcripts present examples of correct (6) and incorrect (7) repetition of the case ending -[e] in identical target sentences, reported in (5). Each sentence is followed by the speaker's code and the time it was uttered. In incorrect repetitions both nouns are marked as -[a], on the model of the nominative case in the input, whereas in target-like output the two nouns are marked with different endings. Typically, each sentence presents either one or no errors.

(5) a. /artɨstk-e pozdravja twumatʃk-a/
         artist-ACC cheers   interpreter-NOM
         'The interpreter cheers the artist.'

     b. /dʑevtʃiŋk-a   tɕɔŋgnje portugalk-e/
         little.girl-NOM pulls   Portuguese.woman-ACC
         'The little girl pulls the Portuguese woman.'

     c. /brazɨlijk-a   vɔwa kuxark-e/
         Brazilian.woman-NOM calls cook-ACC
         'The Brazilian woman calls the cook.'

(6) a. [artist-e   pɔzdravja tumatʃk-a]   (5104, T2)
         artist-ACC cheers   interpreter-NOM

     b. [dʑevtʃjenk-a dʑɔnkje portugalk-e]   (5118, T1)
         little.girl-NOM pulls   Portuguese.woman-ACC

     c. [brazilik-a   voˈa kurark-e]   (1119, T2)
         Brazilian.woman-NOM calls cook-ACC

(7) a. [artistk-a posdravja tumatʃk-a]   (2117, T2)
         artist-NOM cheers   interpreter-NOM

---

[1]See **Saturno2015** for a discussion of similar examples and of the implications of transcription for all subsequent stages of data analysis.

b. [dʒevtʃinkn-a hm   hm   portugalk-a]   (4122, T1)
   little.girl-NOM omission Portuguese woman-NOM

c. [brazilik-a wave   kukark-a]   (3118, T1)
   Brazilian   woman-NOM calls   cook-NOM

The same may happen with the repetition of -[a] NOM, although 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] ((8a); target in (8b)).

(8) a. [brazilijk-ɛ   vowa vowa kuxark-ɛ]   (4113, T1)
        Brazilian.woman-ACC calls   calls cook-ACC

     b. /brazɨlijk-e   vɔwa   kuxark-a/
        Brazilian.woman-ACC calls cook-NOM
        'The cooks calls the Brazilian woman.'

In rare instances, sentence with two errors may occur, in which case endings are swapped ((9a); target in (9b)).

(9) a. [brazilijk-a   vowa kuxark-e]   (2104, T1)
        Brazilian.woman-NOM calls   cook-ACC

     b. /brazɨlijk-e   vɔwa kuxark-a/
        Brazilian.woman-ACC calls   cook-NOM
        'The cook calls the Brazilian woman.'

3.2.3 Repetition of –[e]

3.2.3.1 Descriptive statistics

The EI task was scored as follows: for each combination of time, word order and target ending, the scores of one or zero were assigned depending on whether or not the learner-produced output matched the expected target. The sum of these score was then divided by the number of targets produced (typically eight, but possibly less as a result of omissions or unrecognisable output, which were excluded from the analysis).

SO targets are generally processed with greater accuracy than OS ones, although scores remain rather low - below 50% in most cases. Mean scores vary greatly across L1s, suggesting that there may be an important influence of the native language. Finally, variance is very high, which indicates that learners perform very differently from each other.

Information as to the performance of individual learners at T1 is presented graphically in Figure 3.1. In addition to standard boxplots, the individual data points are presented; their size is directly proportional to the number of learners achieving each score (also specified by the digits in white). Descriptive group statistics are provided in Table 3.1.
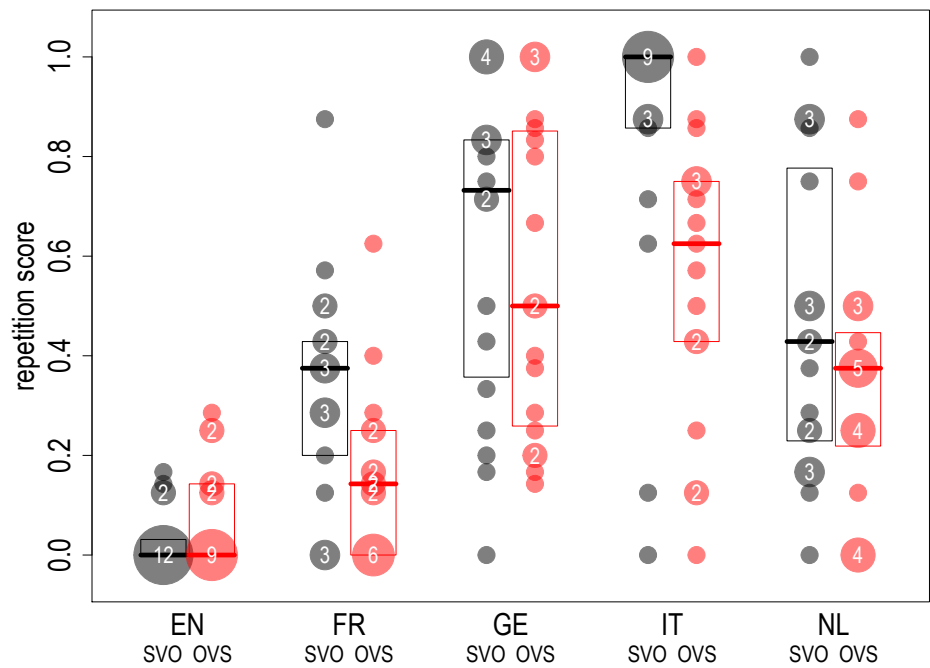
Figure 3.1: EI task, [e] targets, T1, scores by L1

Table 3.1: EI task L1 descriptive statistics, -[e] ending, SO targets, T1

L1	SO			OS		
	mean	sd	n	mean	sd	n
FR	0.34	0.48	123	0.16	0.37	119
GE	0.62	0.49	82	0.57	0.50	110
IT	0.83	0.37	126	0.55	0.50	129
NL	0.48	0.50	147	0.33	0.47	155
EN	0.04	0.19	108	0.09	0.29	112

3 The Elicited Imitation Task

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 least one learner managed to reach 100% accuracy.

Individual variability within the L1 groups is extremely high, with learners' scores ranging from 0% to 100%. The only exception, again, is the English group, in which scores are consistently close to floor level.

The picture looks fairly similar at T2 (Figure 3.2 and Table 3.2), although a general improvement can be observed in the data.
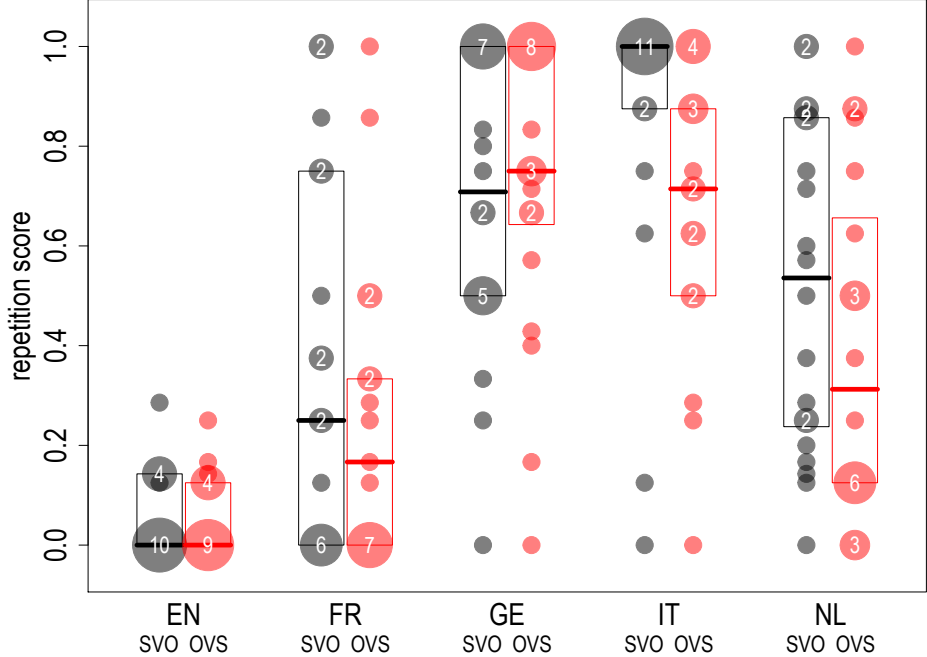
Figure 3.2: EI task, [e] targets, T2, scores by L1

3.2.3.2 Inferential statistics

A generalised linear mixed model (Baayen 2008) with binomial error structure and logit link function (Likelihood Type 3-test) was fitted to the data. Fixed effects comprised the L1 (five levels, reference level = EN), word order (binary, reference level = OS) and time (binary, reference level = T1) and the Llama test

Table 3.2: EI task descriptive statistics, -[e] ending, OS targets

L1	SO			OS		
	mean	sd	n	mean	sd	n
FR	0.39	0.49	126	0.26	0.44	121
GE	0.69	0.46	72	0.74	0.44	111
IT	0.87	0.34	127	0.69	0.47	131
NL	0.53	0.50	146	0.39	0.49	157
EN	0.06	0.24	118	0.07	0.25	118

score (continuous, 0 to 1) as linear predictors, as well as the following two-way interactions: L1:word order, L1:time, and time:word order. Random effects included random intercepts for target sentence and participants as well as within-subject uncorrelated random slopes for time and word order.

The rationale for including the interactions was as follows. The learners' ability to correctly repeat -[e] may be influenced by the word order of the target sentence, with SO targets generally facilitating correct repetition, and OS targets hindering it. In turn, the extent of this word order effect may be variably influenced by the learner's L1. Further exposure is thought to be generally beneficial to repetition, but the extent to which results improve between T1 and T2 may be also determined by the learners' L1: speakers of certain languages may improve more markedly than speakers of other languages. The effect of time is also likely to be constrained by word order.

The summary of the model is presented in Table 3.3.

The Llama score does not appear to be a significant predictor, which suggests that sensitivity to phonological patterns is not involved in determining success at the EIT. The three hypothesised interactions were explored by comparing the full model described above to three null models, each lacking the single interaction of interest. Statistical significance was assessed based on likelihood ratio tests: P values were corrected for multiple comparison using the Holm correction. The results are presented in Table 3.4.

Only the interaction involving L1 and word order proved to be statistically significant, which indicates that the impact of word order (SO vs. OS) varies based on the learner's L1.

The significant interaction was subsequently explored through pairwise comparison. The results relative to the role of the L1 are presented in Figure 3.3, in which blue bars represent confidence intervals for least square means. Pairwise

Table 3.3: Model summary

Predictors	response		
	Odds Ratios	*CI*	*p*
(Intercept)	0.02	0.01 − 0.10	**<0.001**
time [2]	0.76	0.22 − 2.62	0.661
L1 [FR]	2.79	0.85 − 9.19	0.091
L1 [GE]	38.27	11.65 − 125.66	**<0.001**
L1 [IT]	15.27	3.96 − 58.84	**<0.001**
L1 [NL]	9.70	3.01 − 31.24	**<0.001**
WO2 [SO]	0.45	0.12 − 1.64	0.225
llama	4.57	0.71 − 29.39	0.110
time [2] * L1 [FR]	1.71	0.37 − 7.88	0.491
time [2] * L1 [GE]	2.97	0.65 − 13.68	0.161
time [2] * L1 [IT]	3.31	0.70 − 15.75	0.132
time [2] * L1 [NL]	1.96	0.46 − 8.38	0.362
time [2] * WO2 [SO]	0.93	0.57 − 1.53	0.775
L1 [FR] * WO2 [SO]	6.17	2.22 − 17.19	**<0.001**
L1 [GE] * WO2 [SO]	3.13	1.06 − 9.30	**0.040**
L1 [IT] * WO2 [SO]	19.50	6.26 − 60.74	**<0.001**
L1 [NL] * WO2 [SO]	5.17	1.95 − 13.74	**0.001**
Random Effects			
σ^2		3.29	
τ_{00} participant		1.41	
τ_{00} target_no		0.87	
τ_{11} participant.WO2SO		2.19	
τ_{11} participant.time2		2.37	
ρ_{01} participant.WO2SO		1.00	
ρ_{01} participant.time2		-0.14	
N participant		90	
N target_no		16	
Observations		2438	
Marginal R^2 / Conditional R^2		0.550 / NA	

Table 3.4: Full/null model comparisons

predictor	Chisq	Chi Df	p
time :L1	3.023	4	> 0.05
time : word order	0.812	1	> 0.05
L1 : word order	32.607	4	< 0.01

comparisons are statistically significant if the red arrows do not overlap. Statistically significant contrasts are presented in Table 3.5.

Table 3.5: Pairwise comparisons, L1 : word order interaction (only significant contrasts shown)

Condition	pairwise comparison
T1 OS	EN < GE (p < 0.01), IT (p <.01), NL (p < 0.01); FR < GE (p <0 .01), IT (p < 0.01) NL < GE (p = 0.04)
T1 SO	EN < FR (p < 0.01), GE (p < 0.01), IT (p <.01), NL (p < 0.01) FR < GE (p = 0 .01), IT (p < 0.01)
T2 OS	EN < GE (p < 0.01), IT (p <.01), NL (p < 0.01); FR < GE (p <0 .01), IT (p < 0.04)
T2 SO	EN < FR (p < 0.01), GE (p < 0.01), IT (p <.01), NL (p < 0.01) FR < GE (p < 0 .01), IT (p < 0.01)

Turning to the effect of word order (Figure 3.4), it appears that although SO targets generally produce higher scores (except for the L1 English group), the difference is only significant for the L1 Italian group (p < 0.01 at both test times).

3.2.4 A different perspective

The analysis presented thus far has shown that certain L1s seem to be associated to higher repetition accuracy when compared to other languages: for instance, French speakers scored on average 0.16 at T1 on the repetition of -[e], whereas the Italians scored 0.55. From this one might conclude that L1 Italian has a positive effect on processing accuracy. These, however, are but mean values, collapsing the
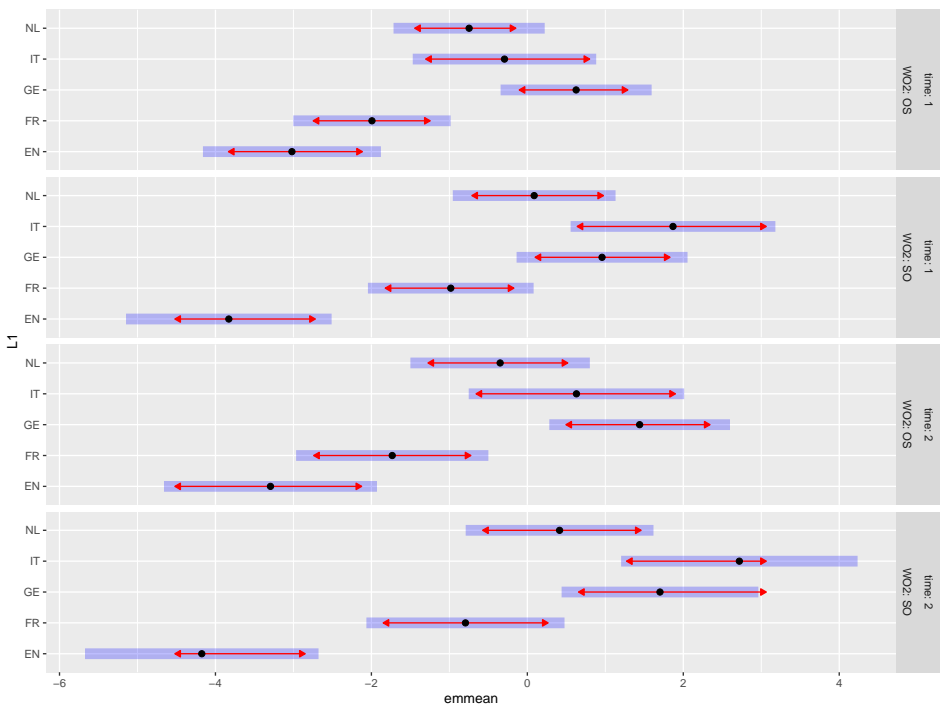
Figure 3.3: Pairwise comparisons, L1 : word order interaction

results of an entire group. But individual learner performance may vary greatly even within the same L1 group, making the idea of a "group interlanguage" quite problematic. Thus, alongside group averages, which may be informative as to the role of a given L1 on the average groups scores, it seems worthwhile to describe learners in terms of their individual processing strategies, operationalised as a set of scenarios. Such outcome would be particularly welcome for an analysis rooted on the Learner Variety theoretical paradigm. Therefore, individual profiling will be used throughout the book to present an alternative view to inferential statistics: it is argued that combined, the two methods may contribute to a better description and interpretation of the data. The present section describes the rationale of this approach.

Within this study, processing profiles may be seen as belonging to three scenarios:

a) learners rely on a single uninflected word-form of each lexical item, with no morphological variation;
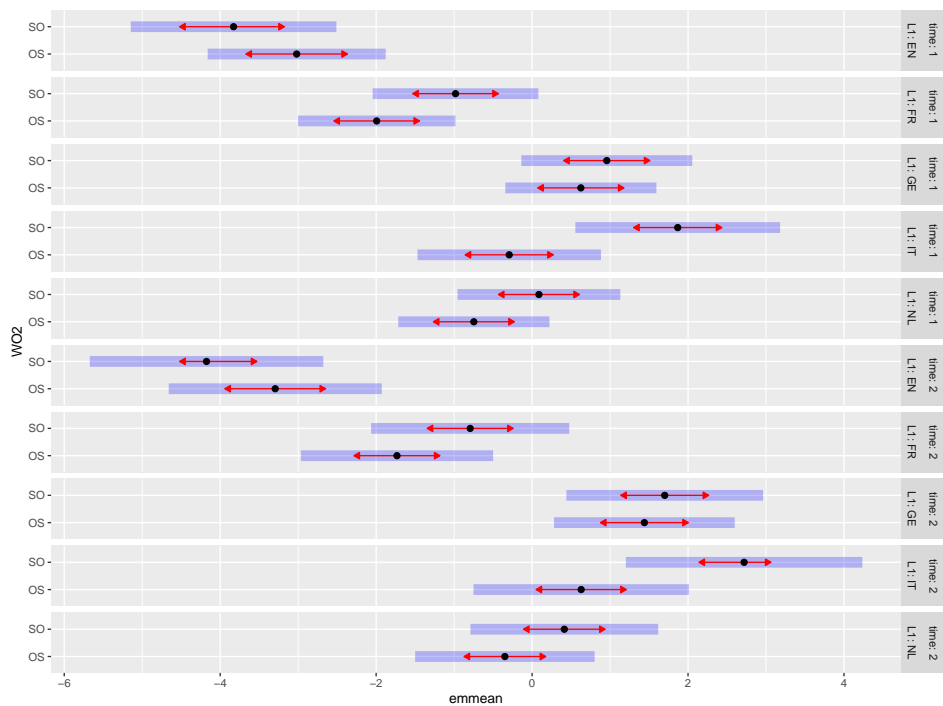
Figure 3.4: Pairwise comparisons, word order: L1 interaction

b) learners may have *noticed* some morphological variation in the input, but they cannot make sense of it based on a systematic 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) might be called chance performance, roughly equivalent to guessing. With only two values to choose from (-[a] and -[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, are characterised by word-final -[a]. Syntactic functions are expressed 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 apparently coherent with the regularities of the target

language, although in an EI test the possibility cannot be ruled out that target-like performance in fact derives from particularly developed phonological memory.

In order to assign learners to the corresponding scenarios, one needs to statistically compute the probability of observing a given result on the basis of a statistical distribution which appropriately models the task at hand. The binominal distribution 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 make it possible to answer such questions as "what is the probability of obtaining head six times if one throws a coin eight times?". If the probability is too small, conventionally below 5%, one may conclude that the coin is not fair, i.e. that it is biased towards a particular result. In the present 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, one should conclude that performance is not random, i.e. that the learner is applying a morphosyntactic principle.

The modelling of the EI task as a coin throwing experiment may seem questionable on several grounds. Indeed, such an approximation is fairly intuitive in the context of a forced-choice response task, such as the Comprehension task described in Chapter 5, in which learners are simply asked to select the correct alternative out of two possible responses. If they pay no attention at all to the target sentence, and only chose pictures through guessing, then the probability that either picture is selected should be 50%. This is not the case in the EI task, in which participants are required to actively produce output, the model for which (i.e., the expected response) is provided in the stimulus sentence.

Moreover, the two possible answers [a] and [e] may not be equally probable or available to the learner. In fact, as it will be shown in Chapter 8, [a] seems to be the unmarked, basic word-form of lexical items, so that if either ending tends to overextend onto the other, most probably it will be [a] overextending onto [e]. More generally, it is common for initial learner varieties to overextend any given word-form onto all others: clearly, the overextended value should be seen as more probable. In the present context, repeating [e] should require a conscious effort on the side of the learner, thus mirroring an intentional strategy.

Finally, the guessing of a binary value relies on the assumption that the trial may only result in two values. This is indeed the rationale of the VILLA EI test, in which the target structure only opposes [a] to [e]. However, it is impossible to tell whether or not learners were aware that the task only targeted two inflectional endings, especially if one considers that it also included a variety of

other structures as distractors: learner output thus may be potentially more varied than that, as even in the limited VILLA input lexical items occur in more than just two word-forms. In sum, the possibility that learners performed the task by guessing alone might seem rather remote.

While this all is true, in principle, the reality is slightly different. The unmarkedness of [a] certainly contributes to explaining why [e] repetition scores tend to zero in some learners, who consistently produced the alternative ending in all contexts. Intermediate scores, however, fit into this picture less well and suggest that target items do imply a choice between [a] and [e], at least in some learners. Regarding the argument that if learners were guessing, then they would produce more than just two endings, it appears that the cases in which learners produce an ending other than -[a] or -[e] (with the exclusion of centralised -/ə/) are extremely rare. After all, the EI task does include a stimulus question in which the expected response is provided. If participants listen carefully to these sentences (which is obviously a prerequisite for the successful completion of the task), they may notice that a) target nouns exhibit some variation across target sentences, and b) that variation only contrasts [a] to [e]. Thus, it does not seem unlikely that the set of possible endings in the learner's mind only comprises [a] and [e], even though other forms occur in the input. Nevertheless, this does not imply that the learner has already identified the regularity which governs their distribution in the input. If that is the case, then the learner might know that either [a] or [e] is required in the task, but will not be able to tell which should be supplied in the individual target sentences: under these circumstances, randomly supplying either ending, i.e. guessing, may indeed sound like a realistic strategy.

To summarise, one should first ask whether or not the individual learner noticed that target nouns vary in their inflectional ending, the possible options being [a] and [e]. If not, the learner will consistently apply a positional principle, so that the statistical test described above becomes superfluous.

If, in contrast, the learner has noticed that there is some variation, two scenarios are again possible: a) it may be that the regularity underlying the distribution of the two endings has already been identified, and that morphosyntactic marking in the output is conscious and systematic; or b) if the regularity is still unclear, endings may be supplied randomly or at least unsystematically. The statistical test described above should be used to distinguish learners who at a given test time behave according to scenario a) or b).

To exemplify, Table 3.4 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 and the resulting mean accuracy. Finally, the column "EI_p" indicates the probability of obtaining a value equal

to or *greater* than that observed in the data if the learner performed the task by guessing. This last value is computed based on the upper tail of a binomial distribution defined by the number of correct responses ("EI_correct"), the total number of trials ("EI_trials") and a probability value set at 0.5. The lower the value, the less likely it is that the learner could obtain such a score or a greater one by mere guessing: in other words, this is the probability of rejecting the null hypothesis that "the learner's repetition of [e] was *not* systematic and intentional" when this is in fact true. Clearly, the output of the test makes little sense in the extreme case in which the learner provided no instances of [e]. The opposite extreme case in which the learner only provided correct repetitions of [e] is also hard to interpret, as the test indicates that the probability of obtaining a score greater than that observed (which is not possible, given the limited number of trials in the task) is 0. For all intermediate cases, the test verifies how likely it is that the outcome was not the product of a systematic strategy. In the case of 7 correct responses out of 8, this probability is close to 0; the fewer the correct repetitions, the more likely it is that no systematic strategy was applied.

Table 3.6: Determining above-chance performance

Subject	WO	Time	EI_correct	EI_trials	EI_mean	EI_p
2102	OS	2	0	8	0.00	1.00
2101	SO	2	1	7	0.14	0.94
2104	OS	1	2	8	0.25	0.86
2118	OS	1	3	6	0.50	0.34
3106	OS	2	6	8	0.50	0.03
5105	SO	1	7	8	0.88	<.00
2108	SO	1	8	8	1.00	0

A word of caution is needed on the possibility of type 1 errors. In the traditional approach, the 0.05 threshold represents the risk which one is willing to accept that what looks like an identifiable tendency in the data (e.g. group A performs better than group B) is in fact due to chance and does not apply to the entire population, but only to the specific sample under examination. Since the present analysis is also based on a statistical test, the same risk is applies here. However, in the present case the 0.05 risk concerns not the entire group (which is not set a priori), but the individual learner: there is a 0.05 possibility that a learner whose performance was classified as "above chance accuracy" in fact did not apply any systematic principle, and only had some lack while performing the

task randomly. Theoretically, the reverse risk also exists, whereby learners did attempt to apply a systematic principle, but failed to do so, but in the context of the present experiment, this situation seems hardly explainable.

In the present analysis, learners are not grouped a priori (as in a treatment vs. non treatment experiment), but based on their performance. Since there is a 0.05 probability that each learner was assigned to the wrong group because of a statistical error, the exact number of learners comprised in each group should be treated with some care.

3.2.5 Repetition of -[e]: a comprehensive picture

The analysis presented so far has failed to provide a comprehensive picture of the behaviour of individual learners across time and word order. This information is provided in Figure 3.5, where each learner is synchronically described in terms of performance at T1 *and* T2, or on SO *and* OS targets. The graph was created using the statistical software R (R Core Team 2017) and the packages *wordcloud* (Fellows 2014) and *extrafont* (Chang 2014) and should be read as follows.

The area is divided into four large squares, representing learner behaviour at T1 in terms of performance on OS (horizontal axis, black) and SO (vertical axis, red) targets. Learners are assigned to the corresponding square depending on whether their performance at T1 differed significantly from chance (upper half of the graph) or not (lower half).

Each large square is further divided into four smaller ones, which describe learner performance at T2 based on the same rationale. The combination of the square in which the learners lies at T1 (large square) and T2 (small square), in this order, determines the *scenario* in which they fall. Scenario 1:3, for instance, identifies the large square no. 1 and the small square no. 3.

In determining learner processing strategies and their evolution over time, one should proceed as follows. First, identify in which main square the learner is found. If, for example, a learner is in the large square no. 1, that means that the output at T1 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 no. 3, then at T2 the output of that learner was still different from chance on OS targets, but no longer so on SO ones.

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: square 1 contains learners who process both types of targets with above chance level accuracy; square 2 those who perform at or below chance level. Among the latter, 9 improved on both types of targets at T2 (scenario 2:1), while 6 showed an improvement on

3 The Elicited Imitation Task
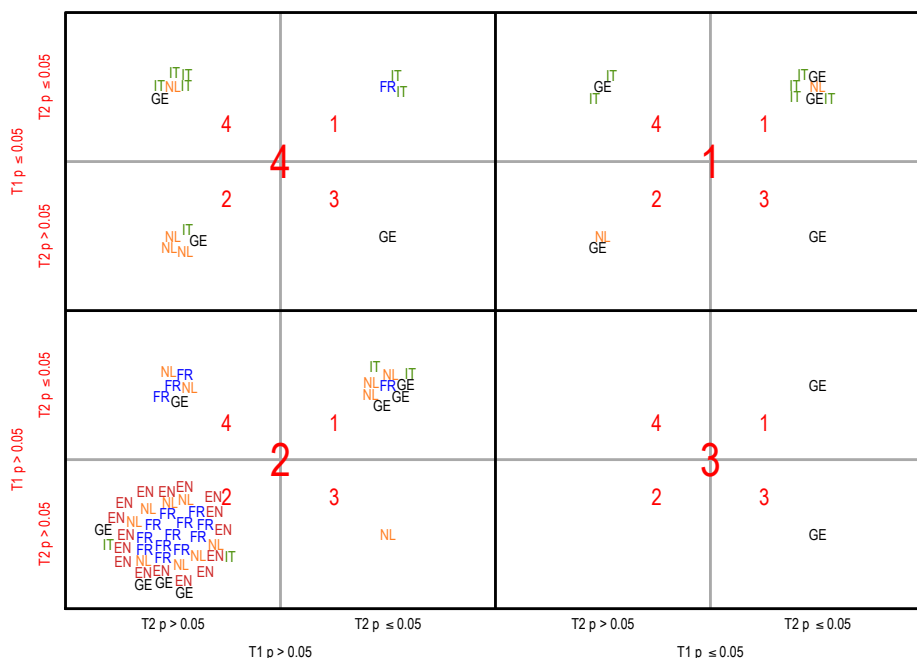
Figure 3.5: EI task, [e] targets, individual processing strategies

SO targets alone (scenario 2:4). A single participant improved on OS, but not SO targets (scenario 2:3). Scenario 1:1 indicates that all targets were processed with above-chance accuracy at both T1 and T2; scenarios 1:2, 1:3 and 1:4 indicates that performance was above chance at T1, but not so at T2, in which either SO (1:3), OS (1:4) or both types of targets (1:2) did not satisfy the criterion.

Squares 3 and 4 indicate a difference in the processing of word order. 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, comprises 15 learners at T1.

Overall, it seems 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. Nevertheless, the vast majority of participants at T1 is found in square 2, indicating chance-level behaviour on both target types.

The graph can also be used to study the evolution of processing strategies in the repetition task over time, depending on the word order of the target sentence. The first obvious observation is that for most learners, there is no evolution whatsoever. The bulk of the data set (42 learners out of 88) can be found in scenario

2:2, which indicates chance behaviour under all conditions (OS and SO targets, at both T1 and T2). This group includes all English L1 learners, most French, about a half of the Dutch, and only a 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 T2 alike.

A few learners show an improvement 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 at T1 failed to systematically repeat [e] under any circumstances, but at T2 improved on SO, OS, or both target types, respectively.

A few participants 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, though rare cases can be found in scenarios 1:3, 1:4 and 1:2: these learners were able to process all targets at T1, but at T2 failed to systematically repeat [e] in SO, OS and all targets, respectively. There might be various explanations for this rare and apparently illogical behaviour. In addition to variables beyond experimental control, such as motivation, tiredness, distractedness, equipment malfunction, such behaviour may be due to border-line scores at T1: even a single additional error thus could have determined their being on either side of the threshold.

3.2.6 Repetition of -[a]

The data set concerning the repetition of [a] is characterised by an evident ceiling effects for all language groups (Figure 3.6 and Table 3.7).

Table 3.7: L1 group scores for the repetition of [a], T1

L1	SO			OS		
	mean	sd	n	mean	sd	n
FR	0.98	0.15	127	0.99	0.09	130
GE	0.98	0.14	145	0.99	0.08	141
IT	0.96	0.19	133	0.95	0.23	128
NL	0.92	0.28	154	0.88	0.32	145
EN	0.96	0.20	101	0.99	0.09	116

As detailed in Chapter 3, in the VILLA input nominative [a] is indeed the most frequent and widespread ending in the paradigm of feminine nouns, in addition
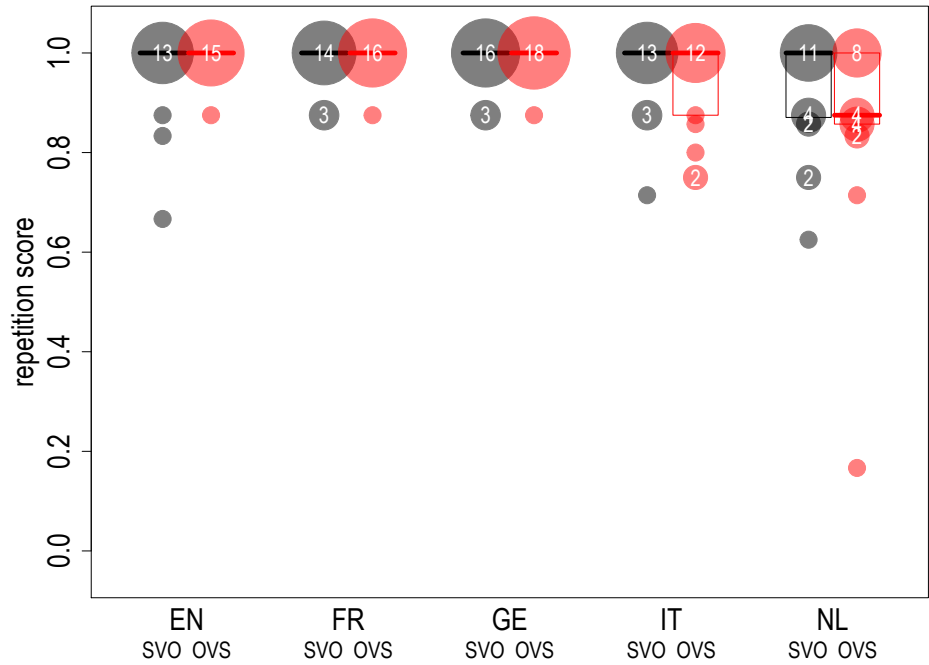
Figure 3.6: EI task, [a] targets, T1, scores by L1

to instantiating both the citation form of lexical item as well as the form in which they were first introduced in the input. It is thus hardly surprising that it may overextend onto the much rarer and specialised accusative ending [e]. However, it is interesting to observe that some learners failed to repeat -[a] in all of the cases in which it was required; this tendency also seems to slightly vary across L1s. Since any output different from either [a] or [e] was excluded from the analysis, a failure to repeat [a] necessarily means that the marked ending [e] was produced. It is important to point out that this observation is not equivalent to saying that the accuracy of the repetition of [e] increases: the obligatory occurrences of [a] and [e] constitute different data-sets and are fully independent of each other. An error in the repetition of [a] may result in the two target sentence nouns being marked as [e], or alternatively to the swapping of the expected case endings, if an error is made in the repetition of target [e], too.

Curiously enough, the errors in the repetition of [a] seem to be maintained and even increase in number at T2 (Figure 3.7 and Table 3.8).

Table 3.9 lists the learners whose probability of correctly repeating -[a] does not differ from chance, again computed for each test time based on a binomial distribution described by the number of correct responses, the number of trials, and
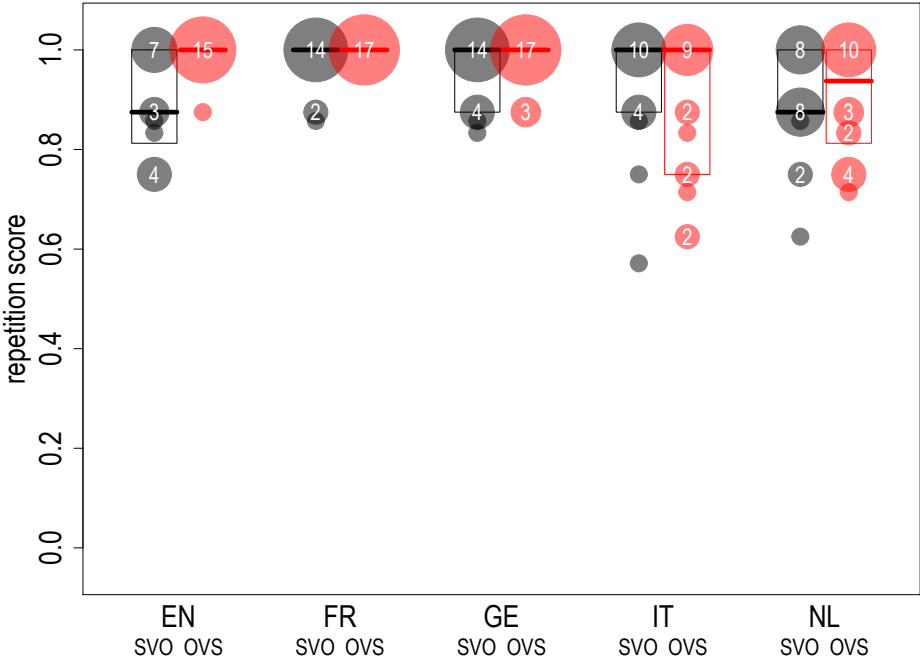
Figure 3.7: EI task, [a] targets, T2, scores by L1

Table 3.8: L1 group scores for the repetition of [a], T2

| | | SVO | | | OVS | |
L1	Mean	Sd	n	Mean	Sd	n
FR	0.98	0.15	132	1.00	0.00	133
GE	0.96	0.19	155	0.98	0.14	155
IT	0.93	0.26	134	0.88	0.32	128
NL	0.90	0.30	158	0.90	0.30	150
EN	0.89	0.31	119	0.99	0.09	119

3 *The Elicited Imitation Task*

a 0.5 chance threshold. These learners do not appear to correctly repeat target [a] with a frequency which could be called 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 it provides the number of correct responses, the number of trials and the resulting mean accuracy. Finally, the column "EI_p" indicates the probability that the learner performed the task by guessing. As discussed in the previous section, this information should be discarded in the extreme cases in which all responses are either correct or incorrect, a scenario whose linguistic interpretation is quite clear anyway. These values are provided for both endings, with the following rationale: if the learner was truly guessing, then one should observe a chance result for both [a] and [e] targets, as these are the only two alternative answers between which one can chose.

Table 3.9: EI task, repetition of –[a] at chance level

| Participant | L1 | WO | Time | EI_correct | | EI_trials | | EI_mean | | EI_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
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	1
5105	IT	SO	1	5	7	7	8	0.71	0.88	0.06	<.00
5105	IT	SO	2	4	7	7	7	0.57	1.00	0.23	0
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, these learners belong to only three L1 groups, the vast majority being speakers of Dutch or Italian. 5105 appears in the table twice because data were collected at both at T1 and T2. Word order and test time, in contrast, are fairly varied.

Some learners (2108, 2118, 5105 at T2, 5115) seem to perform better on the repetition of [a] than of [e]. All other learners conform to the expected pattern, in which repeating [a] appears somewhat easier than repeating [e]. As far as the ending [a] is concerned, the repetition score of some of the participants (2102, 2118, 5105 at T1, 5106) only slightly fails to reach statistical significance: typically,

their p value is 0.06, their mean 0.71, and the correct/total ratio is 5/7, which means that they made two errors out of seven trials. All appear to have missed a trial, which in turn may mean that they supplied an ending other than -[a] or -[e], or alternatively that they failed to repeat an entire target stimulus. In the former case, this behaviour may witness to a certain creativity on their side, which is an indication of system restructuring. The latter case may be speculatively linked to the fact that some participants spent a long time on the distracting phase of the exercise (copying a geometric figure on the answer sheet), which may have somewhat confused their memory of the target. In any case, the 0.05 threshold was set arbitrarily with the purpose of indicating a reasonably small probability, and one could argue that 0.06, although undoubtedly greater, is not so much greater.

These findings may be seen against the bigger picture of -[a] processing. Based on the rationale introduced in the preceding section, Figure 3.8 plots learners according to their performance on the repetition of [a] in OS (horizontal axis, black) and SO (vertical axis, red) targets at T1 and T2.
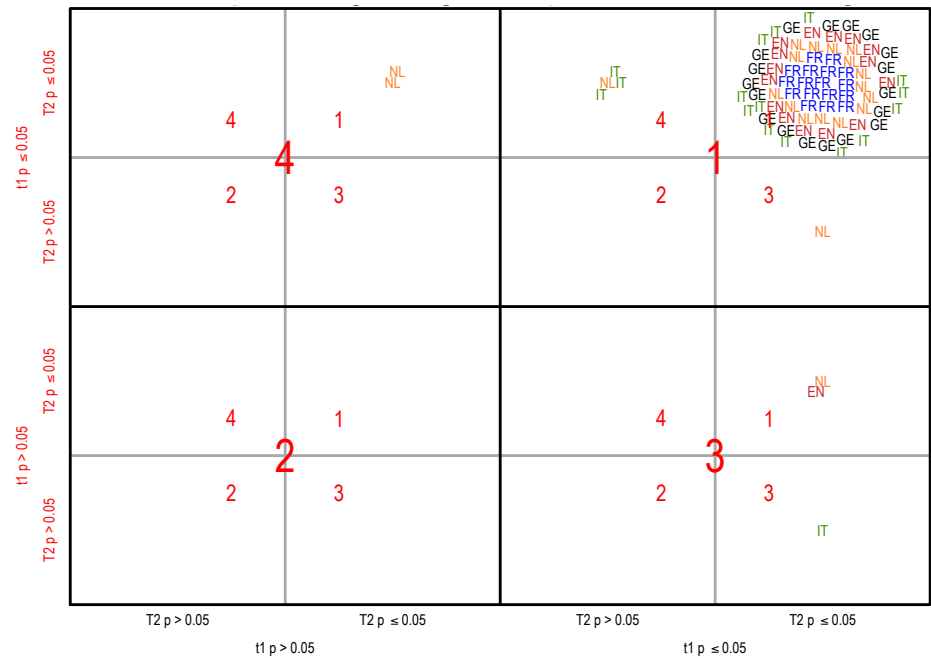
Figure 3.8: EI task, [a] targets, individual processing strategies

Virtually all learners lie in scenario 1:1, which corresponds 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 just discussed.

3.3 Summary

The VILLA EI test highlighted several tendencies, which may be summarised as follows:

- Morphological marking, i.e. the presence of the non-basic case ending [e] is apparently more widespread in SO than OS targets.

- There is a clear effect of cross-linguistic influence in the case of the groups with the highest (L1 German) and lowest (L1 English) scores. The performance of the Italian group is also in line with that of the German participants, although this result was not expected based on contrastive analysis.

- Further exposure to the input is clearly beneficial to the development of morpho-syntactic skills.

It should be borne in mind that in the absence of a comprehension or a translation task it is impossible to verify what learners really meant to say (if anything) through their output. This in turn raises doubts as to the layer of language effectively targeted by the task: in the absence of this information, it is quite possible that learners did not reproduce the content of the stimulus sentence based on their provisional interlanguage grammar, as assumed by the rationale of the task, but simply repeated it as a string of sounds. Both hypotheses have supporting evidence. The rote repetition hypotheses seems realistic in light of output in which target lexical items are hardly recognisable, which suggests that the learner was not striving to reproduce them based on a mental representation, however approximate, but simply tried to retrieve them as sounds from working memory.

On the other hand, the notable difference between the repetition accuracy of the [e] ending in SO as opposed to OS targets suggest that there may be an effect of syntactic structure, which of course can only be hypothesised if the learner processes targets for meaning and attempts to identify their grammatical structure. Even in this case, however, an alternative perception-based explanation may be proposed: in SO targets, the non-basic [e] ending is found in the maximally salient word-final position, which may facilitate its being noticed and reproduced by learners even in the absence of processing for meaning.

In sum, it seems that while a few clear tendencies may be identified, based on the EI data alone it is impossible to definitively establish whether learners' output is based on morphosyntactic processing or perceptual prominence. In order to better describe the behaviour of the VILLA learners the following two chapters will make use of a comprehension test, whose results will prove useful to interpret the output of the EI task.

4 The comprehension test

4.1 Research questions and hypotheses

In line with the overall approach of this research, the comprehension test probes the learner's use of case endings by manipulating word order, based on the assumption that while the meaning of SO targets can be derived based on both a positional and a morphosyntactic principle, on OS targets only the morphosyntactic principle is adequate, as the subject of the utterance no longer occurs in its canonical initial position. Two values of OS word order are considered, i.e. OVS and OSV.

The learners' performance on OS targets thus makes it possible to quantify the extent to which inflectional morphology plays a role in identifying syntactic functions in comprehension.

4.2 Results

4.2.1 Descriptive statistics

This section first presents descriptive statistics relative to learner data, then attempts to identify any statistically significant tendencies using a statistical model. The following section will interpret the same data from the viewpoint of the approach described in §3.2.4, with the aim to detail the individual set of skills of each learner.

Figure 4.1 graphically displays learner scores on SVO, OSV and OVS targets at T1.

Descriptive statistics are provided in Table 4.1.

Figure 4.2 graphically displays learner scores on SVO, OSV and OVS targets at T1.

Descriptive statistics are provided in Table 4.2.

A few preliminary remarks can be made based on these descriptive statistics. First, as expected, SO scores are much higher than their OS equivalents in all cases. Curiously, though, the mean scores below 100% as well as the rather high

4 *The comprehension test*
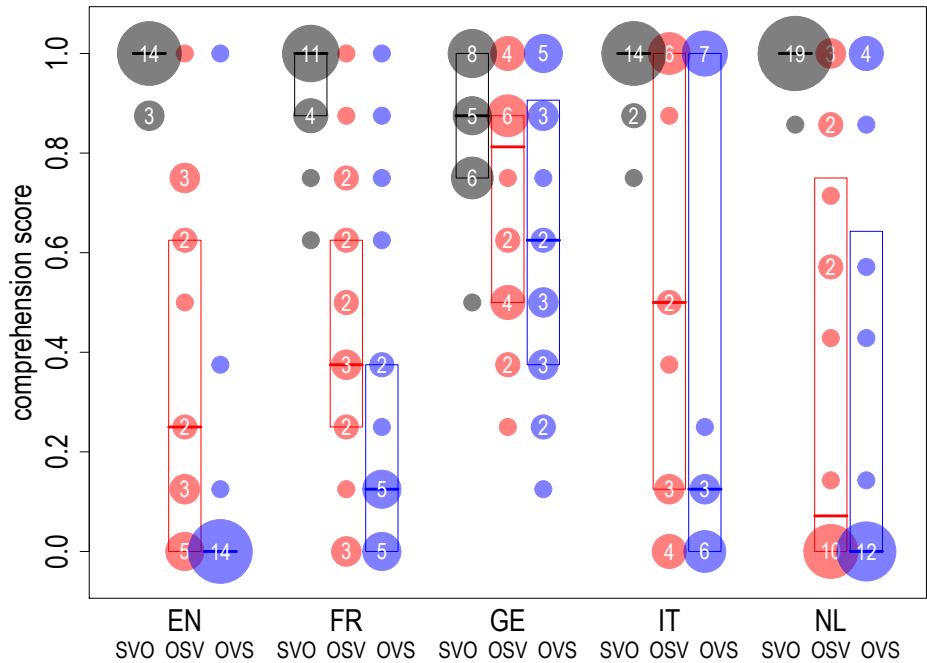
Figure 4.1: Comprehension test, scores by L1 and word order, T1

Table 4.1: Comprehension task, descriptive statistics, T1

L1		SVO		OSV		OVS	
	n.	mean	sd	mean	sd	mean	sd
EN	17	0.98	0.15	0.35	0.48	0.09	0.28
FR	17	0.93	0.25	0.43	0.50	0.29	0.45
GE	20	0.87	0.34	0.71	0.45	0.64	0.48
IT	17	0.97	0.17	0.51	0.50	0.45	0.50
NL	20	0.99	0.08	0.36	0.48	0.30	0.46
mean	-	0.95	0.20	0.47	0.48	0.35	0.43
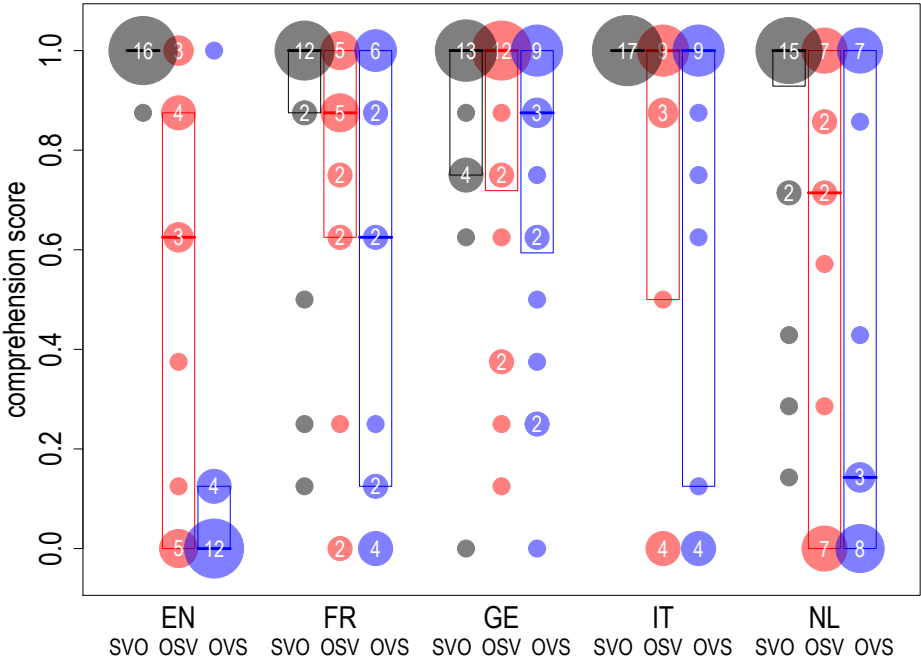
Figure 4.2: Comprehension test, scores by L1 and word order, T1

Table 4.2: Comprehension task, descriptive statistics, T2

L1		SVO		OSV		OVS	
	n.	mean	sd	mean	sd	mean	sd
EN	17	0.99	0.08	0.52	0.50	0.08	0.28
FR	17	0.86	0.34	0.72	0.44	0.55	0.49
GE	20	0.87	0.33	0.80	0.39	0.75	0.43
IT	17	1.00	0.00	0.71	0.45	0.66	0.47
NL	20	0.86	0.34	0.55	0.49	0.43	0.49
mean	-	0.92	0.22	0.66	0.45	0.49	0.43

standard deviations point to the fact that some learners actually made several errors on SVO targets too, which runs contrary to the initial hypotheses.

Regularities are also observed in the difference between the two OS constituent orders. Accuracy on OSV targets is higher in all cases, the only exception being the German group at T2. The English group stands out in this respect, too, in that the difference between the two values of 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 in this group do so in a rather uniform manner.

4.2.2 Inferential statistics

A generalised linear mixed model with binomial error structure and logit link function (Likelihood Type 3-test) was fitted to the data using the R package *lme4* (Bates et al. 2015): fixed effects comprise the L1 (factor, five levels: EN, FR, GE, IT, NL, reference level=EN), word order (factor, binary, reference level=OS) and time (factor, binary, reference level=1) as linear predictors, as well as their two-way interactions: L1:word order, L1:time, and time:word order.

The rationale for including the interactions is as follows. The learners' ability to identify the syntactic structure of the target is hypothesised to be influenced by target sentence word order, SO generally having a positive effect, OS having a negative effect. In turn, the impact of word order may be modulated by the learner's L1. Further exposure is thought to be generally beneficial, but the extent to which results improve between T1 and T2 may be also determined by the learners' L1 (speakers of certain languages improving more markedly than speakers of other languages) and word order (within the same L1 group, either word order may show greater improvement over time).

To simulate individual variability, finally, the model includes random intercepts for participants and test items as well as interacting random slopes for word order and time. The model output is presented in Table 4.3.

The three hypothesised interactions were probed by comparing this full model to three null models, each lacking the single interaction of interest. Statistical significance was assessed based on likelihood ratio tests (Table 4.4). Multiple comparison was addressed using the Holm correction.

The interaction between time and L1 does not reach statistical significance, but both terms engage in other statistically significant interactions. The latter were explored through pairwise comparisons. In Figure 4.3 and 4.4, blue bars depict

Table 4.3: Output model

	response		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.08	0.02 − 0.39	**0.002**
time [2]	3.26	0.87 − 12.21	0.080
L1 [FR]	4.97	0.63 − 38.99	0.127
L1 [GE]	53.17	7.05 − 400.88	**<0.001**
L1 [IT]	11.26	1.33 − 95.56	**0.026**
L1 [NL]	3.00	0.39 − 23.02	0.290
WO2 [SO]	2367.21	332.20 − 16868.10	**<0.001**
time [2] * L1 [FR]	2.86	0.45 − 18.22	0.265
time [2] * L1 [GE]	1.87	0.30 − 11.64	0.504
time [2] * L1 [IT]	4.30	0.57 − 32.21	0.156
time [2] * L1 [NL]	1.51	0.24 − 9.62	0.665
time [2] * WO2 [SO]	0.36	0.18 − 0.75	**0.006**
L1 [FR] * WO2 [SO]	0.02	0.00 − 0.18	**0.001**
L1 [GE] * WO2 [SO]	0.00	0.00 − 0.01	**<0.001**
L1 [IT] * WO2 [SO]	0.06	0.00 − 0.70	**0.025**
L1 [NL] * WO2 [SO]	0.09	0.01 − 0.86	**0.036**
Random Effects			
σ^2	3.29		
τ_{00} participant	8.33		
τ_{00} target_no	0.56		
τ_{11} participant.time2	5.63		
τ_{11} participant.WO2SO	6.98		
ρ_{01} participant.time2	-0.11		
ρ_{01} participant.WO2SO	-0.93		
ICC	0.73		
N participant	91		
N target_no	24		
Observations	4248		
Marginal R^2 / Conditional R^2	0.334 / 0.818		

4 *The comprehension test*

Table 4.4: Comprehension results, single term deletion

predictor	Chisq	Df	Pr(>Chisq)
time : L1	2.484	4	>0.05
time : word order	7.599	1	0.01
L1 : word order	35.840	4	<0.01

confidence intervals: for any pairwise comparison, two terms differ significantly if the red arrows do not overlap.

Figure 4.3 depicts the interaction between time and word order. No statistically significant difference between T1 and T2 can be observed for SO targets, which indicates no significant improvement in time. The reverse is true for OS targets, in which a statistically significant improvement can be observed between T1 and T2 for all L1 groups except the L1 English group, whose performance does improve in time, but not to a significant extent.

Figure 4.4 depicts the interaction between L1 and word order. Notable facts are reported in Table 4.5. The symbols ">" and "<" indicate significantly better and significantly worse performance, respectively. Only significant contrasts are reported.

Table 4.5: Pairwise comparisons, L1: word order (WO) interaction, significant contrasts

condition	significant contrasts
T1, OS	GE > NL (p = 0.03), EN (p < 0.01)
T1, SO	GE < IT (p = 0.01), EN (p < 0.01);FR < EN (p = 0.02)
T2, OS	EN < GE (p < 0.01), IT (p = 0.02)
T2, SO	GE < IT (p = 0.03)

4.2.3 Individual processing strategies

Based on the approach discussed in §3.2.4, this section attempts to compute the likelihood that learners might have performed the Comprehension test with
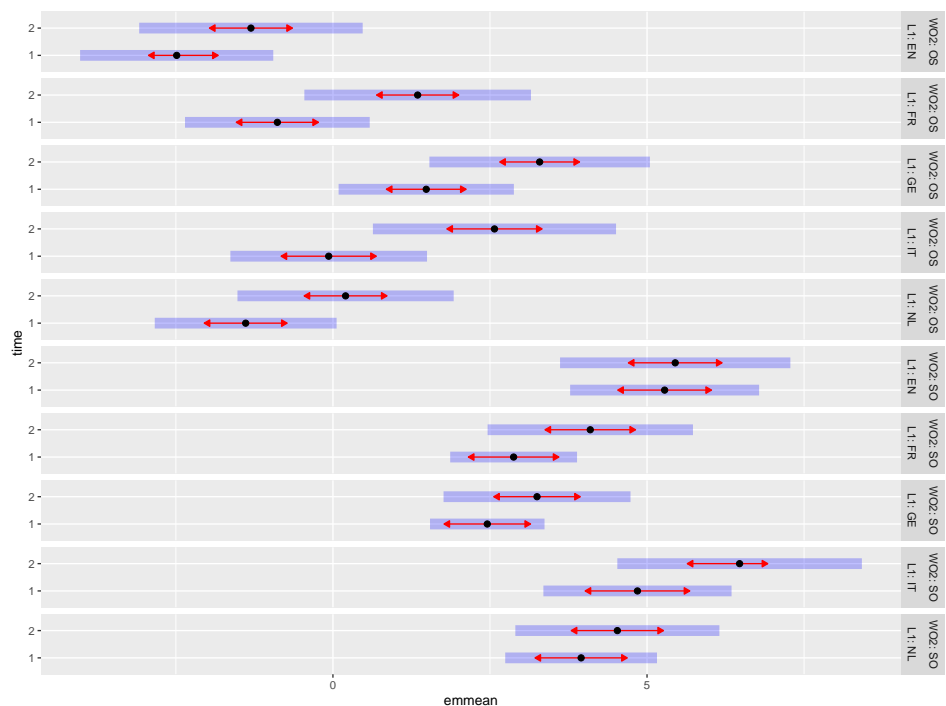
Figure 4.3: Pairwise comparisons, effect of time across L1 and word order

above chance level accuracy, that is, that they responded correctly in such a consistent and systematic way that the existence of a morpho-syntactic principle of utterance organisation may be hypothesised.

Figure 4.5 shows the number of participants who can be said to have applied a target-like morphosyntactic principle in their responses to the comprehension task at T1 (target-like behaviour was defined as systematic, above-chance performance). The digits indicate the actual number of participants for each category.

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. It is curious that two learners (one German, one French) did not achieve above chance accuracy on this type of targets.

Nevertheless, quite a few participants seemed able to correctly process OS targets: this regards 31 learners on OSV and 25 on OVS targets. All L1s are represented, although values for the Italian and German groups are higher than others.

4 The comprehension test
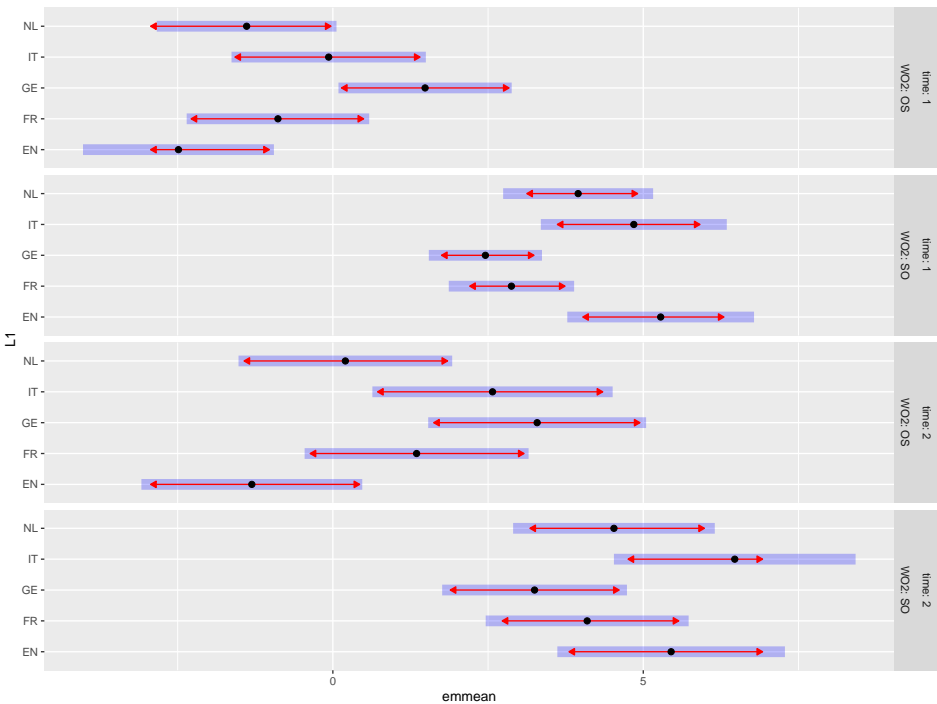
Figure 4.4: Pairwise comparisons, effect of L1 across time and word order

The two values of OS (OSV and OVS) appear to be very similar. If a difference exists, it is very slight and in favour of OSV.

Figure 4.6 presents the data at T2.

Two main tendencies can be observed. The first concerns the marked increase in the number of learners with above chance level accuracy in the processing of OS targets, which grows in all L1 groups. With the only exception of the L1 English group, the improvement concerns both values of OS, although the advantage of OSV over OVS remains. Although less evidently than at T1, the L1 Italian and L1 German groups still achieve better performance than the other groups.

Second, the number of learners failing to reach above chance accuracy on SO also increases.

4.2.4 A comprehensive picture

Data can also be displayed so that the scores of each individual learner may be synoptically seen as a function of L1, word order, and time. The objective is to
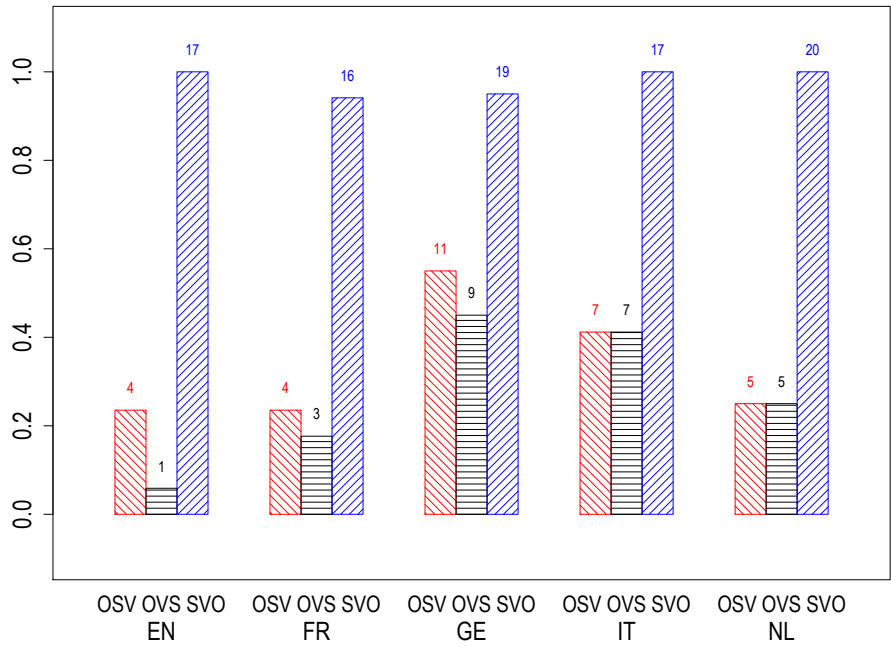
Figure 4.5: Comprehension task, learner above chance

perform a simple cluster analysis to verify whether learners can be grouped based on these factors.

In Figure 4.3, T1 and T2 scores are presented on the horizontal (black) and on the vertical (red) axis, respectively. On both axes, scores 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.

Learners are thus identified by a combination of scores at T1 (horizontal axis, black) and T2 (vertical axis, red), that is, by their position in one of the 64 squares in which the graph area is divided. Each data point represents an individual learner, whose L1 is also specified.

Out of the 64 theoretically possible scenarios, only a few are realised in practice, and fewer still include the bulk of the subjects. As each square corresponds to a varying degree of success in the test, i.e. on various types of targets, one 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. For a synchronic analysis, one needs to con-

4 The comprehension test
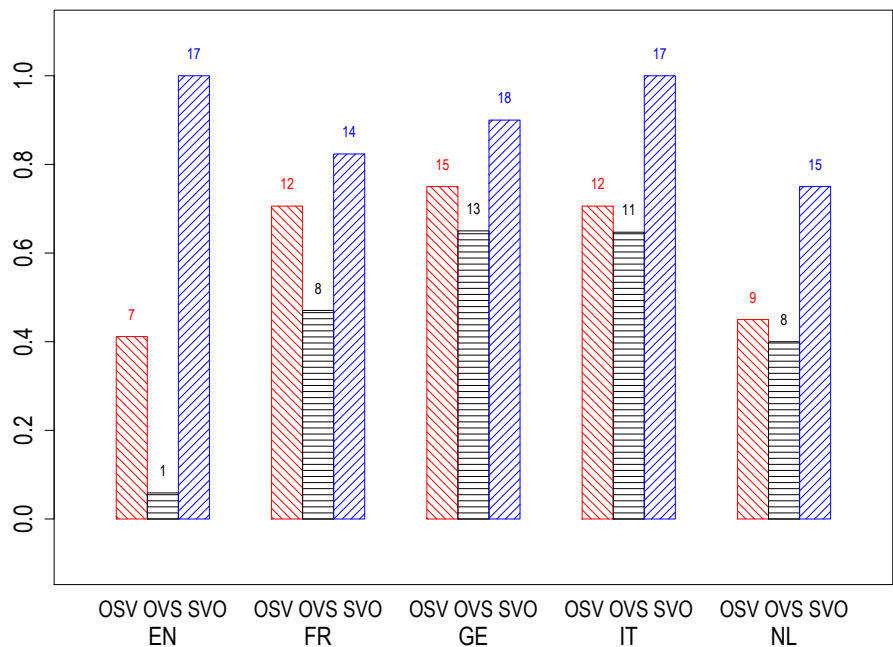
Figure 4.6: Learners significantly above chance by L1, comprehension task

sider the column sum, for T1, or the row sum, for T2, in order to compute the number of participants performing in a specific manner at either test time.

It is evident that at T1 most learners (56) show the following score: SVO 1, OVS 0, OSV 0, which corresponds to a clear 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 pure positional principle (fourth line in the graph) is reduced to 26, whereas those always using a morphosyntactic strategy (first line) now number 36. 15 participants, 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.

One can also study the evolution of learners' processing strategies over time. It seems most relevant to describe the potential evolution of the 56 participants who at T1 were found to adopt a pure positional strategy (SVO 1, OVS 0, OSV 0). 23 of them did not change their processing strategy, consistently applying the
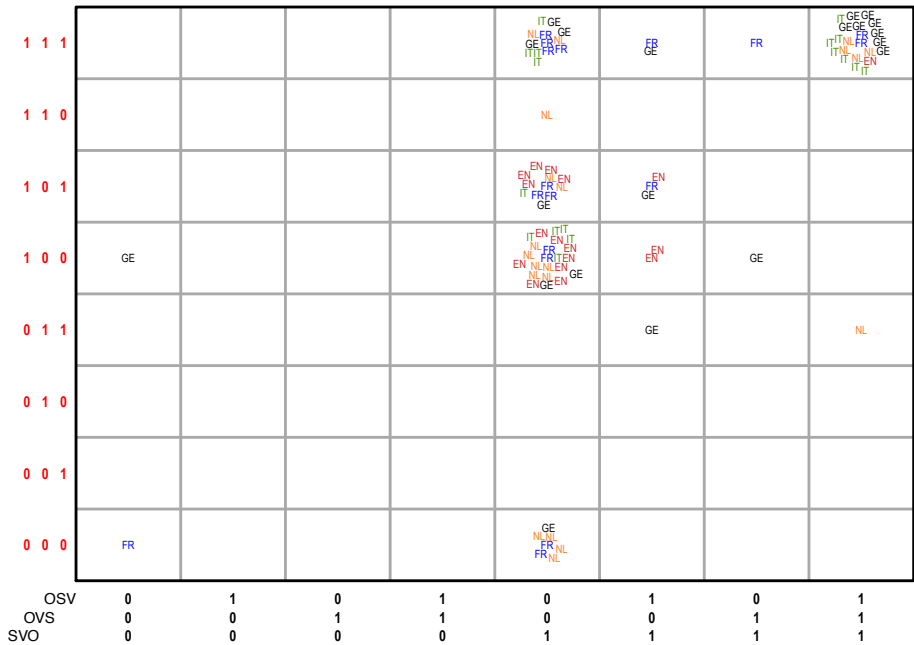
Figure 4.7: Comprehension task, individual processing strategies

same strategy at T2 as well. In contrast, 13 participants moved all the way towards a morphosyntactic strategy, so that at T2 they proved able to consistently derive meaning from both SO and OS targets. 12 learnt to process OSV structures, but still not OVS; only 1 learner exhibits the reverse evolution. This information fits in with the synchronic data, which showed that at both 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 which case OSV develops first.

Finally, 7 subjects (last line, fifth column) move to a stage in which no word order is processed above chance. Beside the fact that additional exposure to the input is apparently detrimental to these learners, this last scenario is particularly surprising in light of the fact that the meaning of SO targets can be correctly identified based on a positional principle.

4.2.5 Differential processing of OS word orders

The analysis so far has revealed obvious gaps in scores between SO targets, on the one hand, and OS targets, on the other hand. These differences are not prob-

lematic 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 that OSV ones. Evidence for such claim comes from various sources: alongside marked differences in mean scores, the processing of OVS targets was often found not to reach above chance performance; further, scenarios in which, at a given time, OVS structures are processed more accurately than OSV are rare; diachronically, OSV almost always develops before OVS. This section first describes the phenomenon in detail and then reports on a statistical test to verify whether the observed differences are statistically significant and require a specific explanation.

Figure 4.4 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 2 and 4). Taken together, the two scores provide an overall picture of learners' behaviour on both OVS and OSV targets at the same time. The main area is divided into four main squares, each representing a processing scenario at T1 according to the conventions summarised in Table 4.6.

Table 4.6: Scenarios, rationale

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

Scenario 1 indicates that the learner processes both OSV (horizontal axis, black) and OVS (vertical axis, red) targets based on a morphosyntactic principle, whereby the first NP is always interpreted as the subject. Scenario 2 is the reverse, that is, both types of target are processed positionally, which leads to an incorrect interpretation of the sentence. In scenario 3, OSV targets are processed in a target-like manner, whereas OVS targets are not; the opposite happens in scenario 4. The last two scenarios are particularly relevant for the present research question, as they 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 manner, an indication of the evolution in time of learner processing strategies is included in the graph. Overall scenarios are identified by the two digits corresponding to learner performance at T1 and T2, in that order.
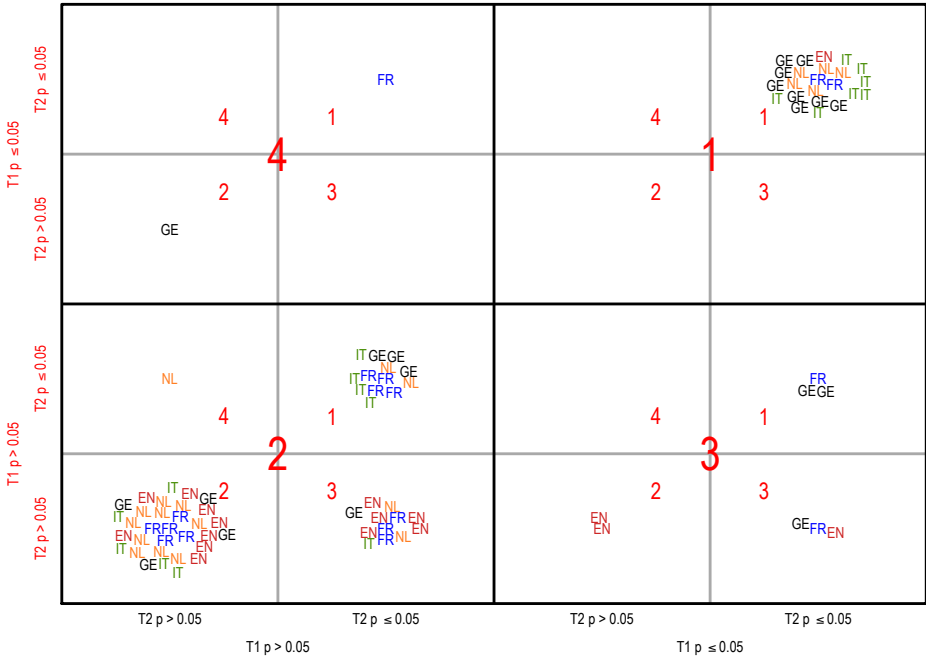
Figure 4.8: Comprehension task, individual processing strategies, OSV-OVS targets

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. Their situation is stable between T1 and T2.

Conversely, learners in scenario 1;1 apply a target-like morphosyntactic strategy at both test times.

Two further major clusters originate from scenario 2 at T1, which indicates a positional principle on both types of targets at T1. However, these learners evolve differently with time: those in scenario 2;1 move on to scenario 1 at T2, which means that over time they learnt to generalise a morphosyntactic strategy to all OS targets. Those in scenario 2;3 managed to do so only on OSV targets, and

4 The comprehension test

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 tend to be acquired first.

Synchronically, more learners appear in scenario 3 than in scenario 4 at both T1 and T2, as shown in Table 4.7 and 4.8 respectively.

Table 4.7: Comprehension test, OS targets, learner distribution across scenarios, T1

		OSV	
		p ≤ 0.05	p > 0.05
OVS	p ≤ 0.05	23	2
	p > 0.05	8	58

Table 4.8: Comprehension test, OS targets, learner distribution across scenarios, T2

		OSV	
		p ≤ 0.05	p > 0.05
OVS	p ≤ 0.05	40	1
	p > 0.05	15	35

It thus appears that OSV targets are indeed easier to process that OVS; in the following lines a few reasons for this will be explored. In SVO targets (1), 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.

(1) siostr-a   woła brat-a
     sister-NOM calls brother-ACC
     '(The) sister calls (her) brother.'

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.

(2) brat-a   woła siostr-a
    brother-ACC calls sister-NOM
    '(The) sister calls (her) brother.'

Morphosyntactic processing requires learners to be aware of the gender and inflectional class of target nouns: as (2) makes it clear, both nouns may be marked by the same ending, whose meaning (NOM.SG.F vs. ACC.SG.M) depends on the paradigm to which the noun belongs. Combined with the modest prominence of case endings and the pressure exerted by the test, this may confuse learners, leading them to mistake these targets for instances of SVO utterances. In other words, it may be the case that whenever learners encounter 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 with OSV targets (3), in which the structure of the utterance is quite different: the two noun phrases come first, followed by the verb. This order 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.

(3) brat-a   siostr-a   woła
    brother-ACC sister-NOM calls
    '(The) sister calls (her) brother.'

4.2.5.1 Differential processing of OS word orders: Inferential statistics

To test the effect of OS word order statistically, the generalised linear mixed model described in §4.2.1 was compared to an identical model, in which however the predictor WO2 (word order with two values, i.e. SO and OS) was substituted with the predictor WO3 (word order with three values, i.e. SVO, OVS and OSV). The model output is presented in Table 4.9.

A likelihood-ratio test reveals a statistically significant difference between the two models (Chisq = 131.41, Df = 10, $p < 0.01$), which suggests that accounting for the differential processing of OS word order configuration is beneficial to the interpretation of the data. However, pairwise comparisons (Figure 4.9) reveal that the difference in score between the processing of OVS and OSV is significant only for L1 English learners ($p < 0.01$ at both test times).

Table 4.9: Model output

Predictors	response		
	Odds Ratios	*CI*	*p*
(Intercept)	0.28	0.07 − 1.09	0.067
time [2]	4.15	0.99 − 17.38	0.051
L1 [FR]	2.33	0.38 − 14.30	0.360
L1 [GE]	17.70	2.96 − 105.90	**0.002**
L1 [IT]	4.08	0.61 − 27.20	0.146
L1 [NL]	1.43	0.24 − 8.67	0.696
WO3 [OVS]	0.01	0.00 − 0.04	**<0.001**
WO3 [SVO]	681.07	111.54 − 4158.79	**<0.001**
time [2] * L1 [FR]	2.52	0.34 − 18.69	0.367
time [2] * L1 [GE]	1.31	0.18 − 9.44	0.787
time [2] * L1 [IT]	3.28	0.38 − 28.23	0.280
time [2] * L1 [NL]	1.18	0.16 − 8.65	0.872
time [2] * WO3 [OVS]	1.37	0.73 − 2.58	0.324
time [2] * WO3 [SVO]	0.41	0.19 − 0.91	**0.028**
L1 [FR] * WO3 [OVS]	31.41	4.97 − 198.62	**<0.001**
L1 [GE] * WO3 [OVS]	117.22	17.14 − 801.53	**<0.001**
L1 [IT] * WO3 [OVS]	94.50	11.59 − 770.64	**<0.001**
L1 [NL] * WO3 [OVS]	37.62	5.60 − 252.89	**<0.001**
L1 [FR] * WO3 [SVO]	0.04	0.00 − 0.31	**0.002**
L1 [GE] * WO3 [SVO]	0.00	0.00 − 0.03	**<0.001**
L1 [IT] * WO3 [SVO]	0.16	0.02 − 1.66	0.124
L1 [NL] * WO3 [SVO]	0.19	0.02 − 1.61	0.129
Random Effects			
σ^2	3.29		
τ_{00} participant	6.15		
τ_{00} target_no	0.24		
τ_{11} participant.time2	6.63		
τ_{11} participant.WO3OVS	3.83		
τ_{11} participant.WO3SVO	5.19		
ρ_{01} participant.time2	-0.14		
ρ_{01} participant.WO3OVS	0.70		
ρ_{01} participant.WO3SVO	-0.91		
ICC	0.76		
$N_{participant}$	91		
N_{target_no}	24		
Observations	4248		
Marginal R^2 / Conditional R^2	0.375 / 0.852		
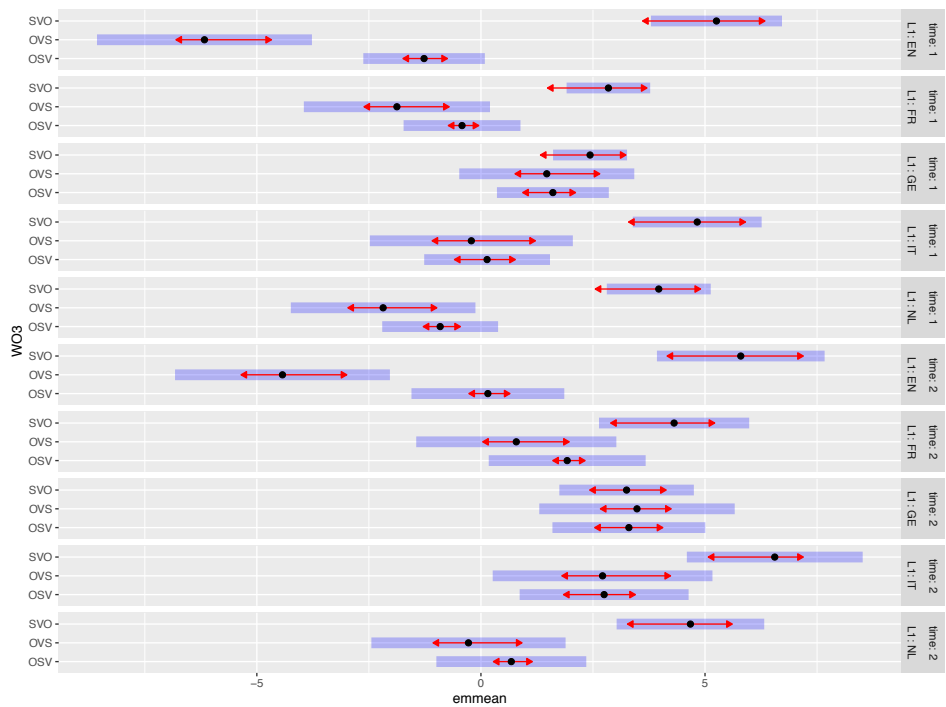
Figure 4.9: Pairwise comparisons, WO3 across L1 and time

4.3 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. The main findings can be summarised as follows. 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 fair 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.

5 A comprehensive view of morphosyntactic skills

5.1 Research questions and hypotheses

The purpose of this chapter is to verify to what extent the VILLA learners acquired a morphosyntactic principle of utterance organisation, whereby grammatical meaning is encoded by inflectional morphology independently of word order. To this purpose, the results of the EI task (Chapter 3) and of the comprehension test (Chapter 4) are correlated, arguing that "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 [...]" (Pallotti 2007: 326). The present analysis aims to verify whether the target structure is simultaneously mastered in both comprehension and repetition, or if it develops in either of them first, in either SO or OS word order. The effect of other predictors such as input exposure and L1 is also investigated.

5.1.1 Task type

The EI task requires learners to phonologically decode, and possibly comprehend, the target sentence and then reproduce it based on their interlanguage grammar. In a sense, it should in principle, encompass the same skills needed for the comprehension test, although there are some important differences. The first is that repetition may take place without comprehension, although the task was designed so as to make this unlikely. Secondly, repetition involves a further skill, that is, language production, and it may therefore be argued that it represents a more complex task comprising several skills at the same time. For this reason it is expected that it will produce poorer results, i.e. some learners may be able to process a given target in comprehension, but not in repetition.

5.1.2 Word order

As has been argued in chapters 4 and 5, the effects of word order differ in the two tasks at hand. In the comprehension test, word order directly correlates with

markedness, as SO targets conform to the first-noun principle while OS targets violate it. The picture is more complex in the case of the EI task (see section Chapter 4), but in short it can be said that OS targets should prove harder for the following reasons: a) OS violates the first-noun principle; b) when producing OS structures, the non-nominative case ending must be supplied outside its canonical position, which, in initial interlanguages, is typically post-verbal; c) the non-nominative case ending occurs in the perceptually non-salient utterance-medial position.

It is thus expected that results on OS targets will be poorer in both tests.

5.1.3 Correlation between task type and word order

The overall purpose of this chapter is to identify the contexts in which learners can be hypothesised to adopt a morphosyntactic principle, "context" referring to a combination of task type and target sentence word order (e.g. repetition of OS targets). Further, it may be the case that, in order to master one particular context, learners must be able to develop others first (e.g. the comprehension of OS targets and the repetition of SO targets). The analysis thus aims to identify possible implications between contexts (e.g. the repetition of OS targets implicating their comprehension), in order to identify a difficulty scale.

5.1.4 Cross-linguistic influence

It is expected that speakers of languages with relatively free word order and case marking should be favoured. Within the VILLA project only German possesses such characteristics, albeit to a more limited extent that Polish. Rigid word order as observed in English and French is hypothesised to impose a positional principle on the learner, thus slowing down the acquisition of the target structure. L1 biases may also prompt learners who are not familiar with the category of case to rely on information such as animacy and word order to identify the agent of the sentence. These cues are admittedly relevant in the processing of many Polish real-life utterances, but were purposefully excluded in the present experimental paradigm.

5.1.5 Exposure to the input

Intuitively, additional exposure to the input can only be beneficial for the acquisition of the target structure. In addition, two more specific questions may be formulated.

- Implicational sequence of development. Improving in one context may require learners to master others first. This requires that the implicated abilities should be already developed at T1, or develop between the two test times. It is predicted that learners will improve in comprehension before repetition, and on SO targets before OS ones.

- Cross-linguistic differences. This question regards the presence of interactions between a possible implicational scale of development and the learners' L1. It may be the case, for instance, that speakers of a certain L1 improve on particular combination of task type and word order. Participants advantaged by their L1 from a cross-sectional perspective should also exhibit faster and more significant gains.

5.2 The comprehension test as a disambiguator to the EI task

The comprehension test can be helpful to shed light on some of the questions which emerged from the analysis of the EI task. Given a repeated utterance like (1), it is impossible to establish *a priori* whether or not the learner truly attempted to encode any specific meaning, and, if so, what this may be.

(1) [artɨstk-a pozdravja twumaʧk-a]
     artist-NOM cheers   interpreter-NOM

The question is particularly relevant in the case of OS targets. Learners may fail not only on the repetition of the non-basic ACC ending, but also on the comprehension of the target sentence, which, according to the first-noun principle, could be processed as a default subject-initial utterance. With regard to a target utterance like (2):

(2) /arˈtɨstk-e pozˈdravja twuˈmaʧk-a/
     artist-ACC cheers   interpreter-NOM
     'The interpreter cheers the artist.'

An output like (1) may instantiate at least three underlying structures (3):

(3) a. artist-OBJ cheers interpreter-SUBJ
     b. artist-SUBJ cheers interpreter-OBJ
     c. artist cheer interpreter

(3a) corresponds to target-like comprehension of the OS target, the deviant output in (1) owing to a failure to *produce* a non-basic word-form, which nonetheless was correctly identified in comprehension. In (3b), the utterance is interpreted as subject initial based on a positional principle and in spite of inflectional morphology. In (3c), finally, the learner only identifies lexical items with no attached grammatical meaning. In the latter case, the learner is not reproducing a sentence, but rather a list of words with no meaningful connection.

In the absence of a translation test, output like (1) is bound to remain ambiguous. The comprehension test makes it possible to reduce the degree of uncertainty concerning the learner's underlying structures as exemplified in (3). Should the comprehension test show that a particular learner is incapable of processing OS targets, then it would be highly unlikely that the same learner could have processed the same target correctly in the EI task. At most, the learner might have attempted to encode a SO utterance, so that the source of incorrect output lies in comprehension: one could thus exclude option (3a). If a learner performs above chance on the comprehension test, in contrast, the possibility exists that he might have tried to produce an OS utterance, though failing to encode grammatical meaning through case endings. The difficulty in this case should be localised at the level of repetition, rather than comprehension.

To summarise, correlating the two tests cannot provide final answers as to the learner's strategies of utterance organisation, but makes it possible to exclude unlikely explanations of the observed output.

5.2.1 Correlating the repetition and comprehension tests: methodology

In order to correlate the results of the repetition and the comprehension tests, a first intuitive approach might be plotting the learners' scores in various test conditions (task, word order, time) side by side, as in histograms or boxplots like Figure 5.1. However, with this approach, it is impossible to trace the behaviour of individual participants across various conditions, as is indeed the purpose in the present work, because individual learners are not univocally identified. To exemplify, it is impossible to tell how the single participant scoring just below 0.8 in the comprehension of SO targets performed in the other conditions.

Moreover, differences in aggregated scores, even if proved to be statistically significant, do not necessarily have a linguistic interpretation. Finally, each graph only describes the performance of a single L1 group at either test time. In the case of a complex project such as VILLA, it would take ten such graphs to fully describe tendencies in the data-set.

5.2 The comprehension test as a disambiguator to the EI task
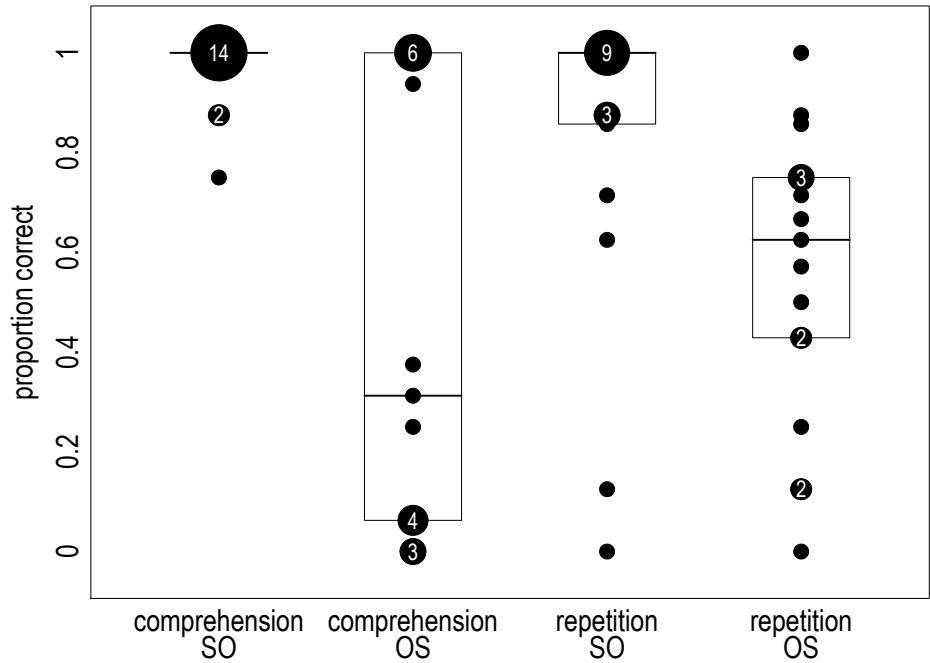
Figure 5.1: Italian learners' performance at T1, comprehension test and EIT

To address this first limitation, Table 5.1 presents the proportion of learners scoring statistically above chance in each test condition, based on the methodology described in §3.2.4.

Table 5.1: Learners scoring above chance

	OS								SO							
	T1				T2				T1				T2			
	Comp		EI		Comp		EI		Comp		EI		Comp		EI	
	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
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, that is, whether or not the learners can be thought to have applied a morphosyntactic strategy. Again, however, no information is provided as to the score of individual learners in various conditions.

5.2.2 Scenarios

In order to portray a comprehensive picture of learners' morphosyntactic skills, "scenarios" (see §?? are introduced as a methodological tool. Scenarios represent a single, global score of the learners' processing skills in both comprehension and repetition. For each test, scores are coded as "positive" or "negative" based on the rationale described in §4.2.4. Four scenarios are possible (Table 5.2):

Table 5.2: Scenarios, rationale

		comprehension	
		+	−
repetition	+	**1**	**4**
	−	**3**	**2**

In scenarios 1 and 2, both tests are performed above chance and at or below chance level, respectively. Scenarios 2 and 4, in contrast, point to a situation in which learners perform well in one test and poorly in the other one. It is thus possible to investigate whether the two skills are correlated in the learners' competence, the alternative hypothesis being that either might develop earlier in time.

The use of this tool is exemplified on the basis of the results obtained by all learners at T1 on OS targets. The area of Figure 5.2 is divided into four squares, each corresponding to a scenario, indicated by a large number in red. Learners are identified by a coloured digraph according to their L1 (Table 5.3).

Table 5.3: Figure 5.2, identifiers

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

Depending on whether or not the performance of the learner in question varies from T1 to T2 or not, the digraph is printed in lowercase or uppercase letters, respectively. Learners identified by capital letters will no longer be in the same

position in the graph depicting the situation at T2 (Figure 5.3), while those identified by small letters remain in the same square at both T1 and T2.

Crucially, participants are identified univocally by their position in the graph, at the intersection of their comprehension and repetition scores. The position of each learner in the square does not reflect actual scores in the two tests: rather, for the reasons previously discussed, the graph only indicates the participants' performance in terms of scenarios. Their position within each square is simply meant to improve readability.

Figure 5.2: Repetition and comprehension scores, OS targets, Time 1

An obvious cluster comprising more than a half of the dataset at T1 is located in scenario 2, indicating that neither test was performed with above chance accuracy. The second largest cluster corresponds to scenario 3, indicating above chance scores in comprehension, but not in repetition. Finally, a smaller group of learners can be found in scenario 1, indicating that already at T1 some learners managed to process OS targets morpho-syntactically in both comprehension and repetition. Coherently with the assumptions of the EI task, very few learners are found in scenario 4, which corresponds to above chance performance in repetition, but not in comprehension.

5 A comprehensive view of morphosyntactic skills

The picture presented so far is still incomplete, as it only depicts learner performance on OS targets. In order to provide a comprehensive picture of learners' processing, though, it would be desirable to have a synoptic representation of performance on SO targets as well.

This may be exemplified by focussing on the learners located in scenario 2 in Figure 5.2 (scoring at or below chance level for both comprehension and production). Figure 5.3 depicts their performance on SO targets following the same rationale.
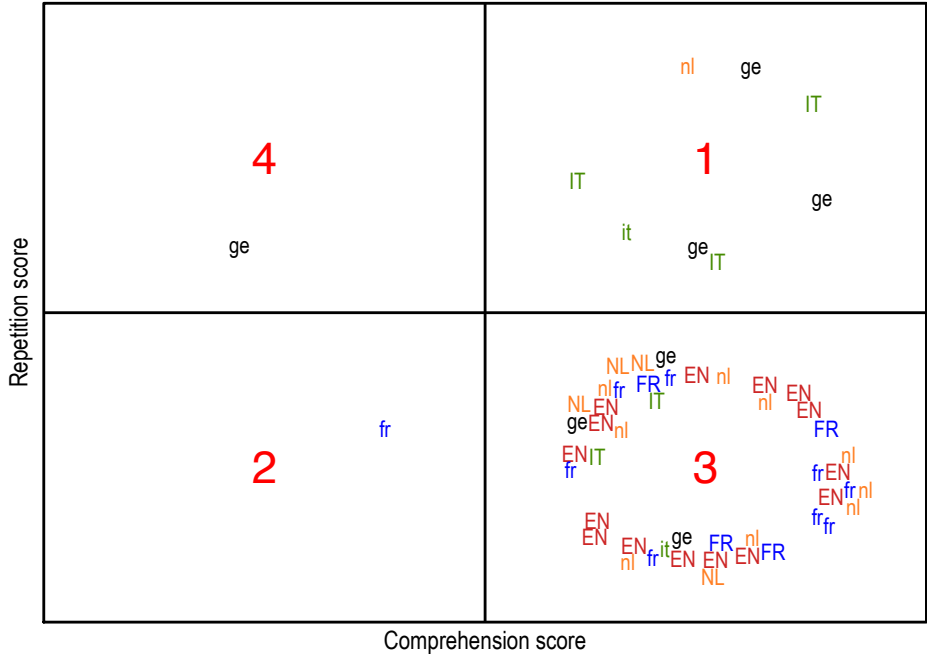
Figure 5.3: Repetition and comprehension scores on SO targets for learners in sc. 2 on OS targets, Time 1

Again, an obvious cluster can be identified, this time in scenario 3. Such good performance on SO targets is hardly surprising, as target-like interpretation may be achieved based on either a positional or a morphosyntactic principle. The exiguity of data points in scenario 1, in contrast, witnesses to the greater difficulty of the EI task, although 8 learners do exhibit above chance accuracy. Finally, Scenarios 2 and 4 violate the assumptions of both test rationale and word order manipulation and are coherently empty.

The next step consists in merging the information presented in Figure 5.2 and

5.2 The comprehension test as a disambiguator to the EI task

Figure 5.3 into a single, comprehensive representation. This is achieved as shown in Figure 5.4.
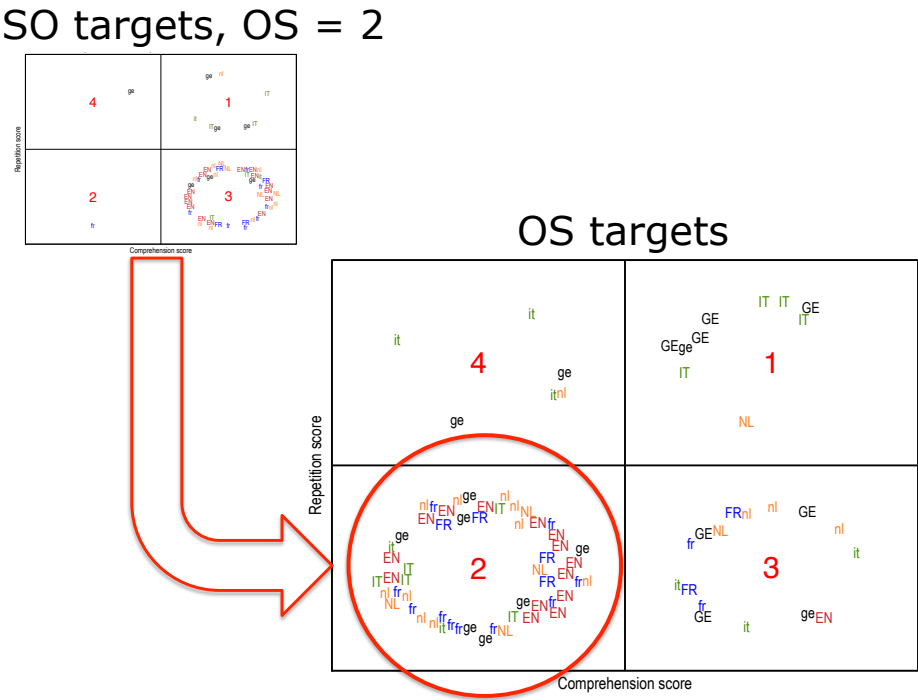
Figure 5.4: Merging the information presented in Figure 5.2 and Figure 5.3

The main squares of the graph representing the processing of OS targets are further divided into 4 minor squares, depicting the processing of SO targets by the learners comprised in the main square. Both representations rely on scenarios, arranged clockwise (4,1,3,2) for both targets, where 1 represents above-chance performance in both tests, 2 under-chance performance in both tests, and 3 and 4 depicting the situation of learners who perform above chance in one test and below in the other. To exemplify, scenario 2 on OS targets comprises 57 learners (circled in red). Based on their performance on SO targets, these learners may be grouped as follows: sc. 4: 1; sc. 1: 8; sc. 4: 1, sc. 3: 47. This distribution is graphically represented by the small square on the top left. But it would also be useful to show the performance of each learner on both SO and OS targets at the same time: to this purpose, the representation in the red circle, which only indicates sc. 2 performance on OS targets, is substituted with the square on the top left,

5 A comprehensive view of morphosyntactic skills

which adds information as to the same learners' performance on SO targets. The final result is shown in Figure 5.5.
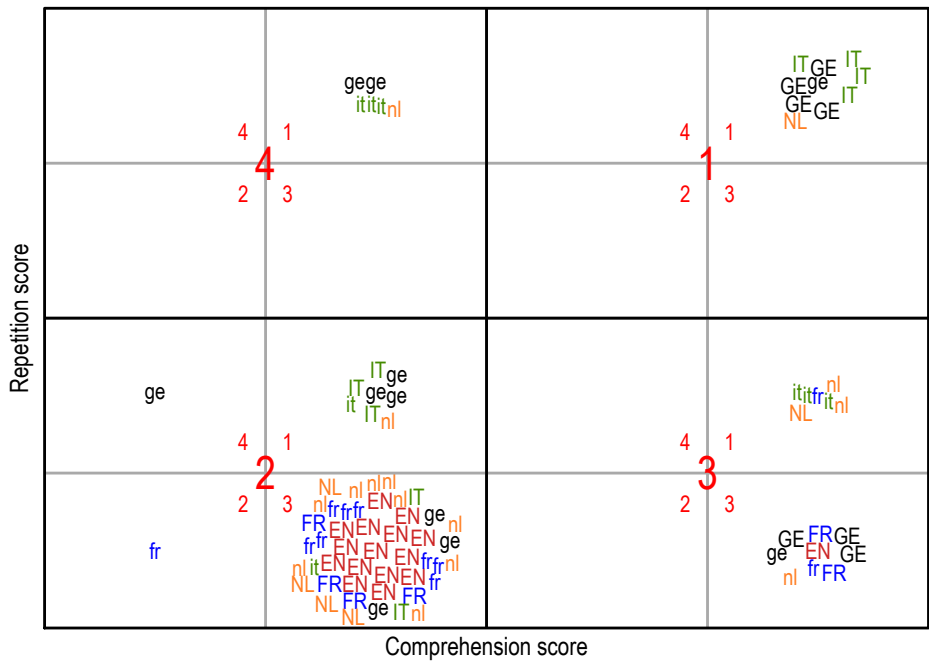
Figure 5.5: Scenarios, T1

5.3 An overall picture of learner morphosyntactic skills: results

Figure 5.5 is divided into 16 squares, which correspond to unique combinations of OS and SO processing scores. Each square is identified by two coordinates, corresponding to the OS scenario (large numbers) followed by the SO scenario (smaller numbers). To exemplify, scenario 2;3 corresponds to the largest cluster observed (second from left, bottom row). Again, some of the theoretically possible scenarios are linguistically unmotivated, and are accordingly empty. A rationale of linguistically motivated scenarios is provided below.

1;1 Full morphosyntactic principle. Both tests are performed with above chance accuracy on both types of targets.

3;1 On SO targets, both tests are performed above chance; with OS targets, only the comprehension test is. The repetition of OS targets proves the hardest task.

2;1 Both tests are performed with above chance accuracy on SO targets; on OS targets, neither is. Independently of the test, OS targets are harder than SO ones.

2;3 Positional principle. only the comprehension test on SO targets is performed with above chance accuracy.

The scenarios identified in Figure 5.5 are represented analytically in Table 5.4. Each is broken down into its test and word order components. The last row computes the number of learners who perform above chance in each combination of test and word order.

Table 5.4: Implicational hierarchy at T1

OS repetition	SO repetition	OS comprehension	SO comprehension	scenario	n.
+	+	+	+	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		

Following Aldai & Wichmann (2018; see also Nyqvist 2018; Wichmann 2015; 2016; Hatch & Lazaraton 1991: 210-212), the matrix was submitted to a significance test of the degree of scalarity applying matrix randomization statistical testing (Janssen et al. 2006) based on Guttmann scaling. To this purpose, the R script made available by Aldai & Wichmann (2018) as well as the *vegan* R package (Oksanen et al. 2019) were used. A solid hierarchy emerges (GC 95,74, p < 0.01):

OS repetition ⊃ OS comprehension ⊃ SO repetition ⊃ SO comprehension.

A few observations can be made. First, either task is harder on OS targets than on SO ones. Secondly, within a given constituent order, the EI task is harder than the comprehension test. Finally, all scenarios which are coherent with the hypotheses concerning constituent order and the EI task are indeed part of the

hierarchy. There is an exception to this rule, however: scenario 3;3, comprising 9 learners, which is not part of the hierarchy and yet does not violate any assumption:

3;3 comprehension scores are above chance on both SO and OS targets, whereas repetition scores are at chance level.

This scenario suggests that, independently of the target structure, repetition is harder for the learners than comprehension. Truly, it was not predicted that if learners can process OS targets in comprehension, they should be able to process SO targets in repetition, too. However, the vast majority appears to follow this pattern. 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 situations seem equally possible.

Another 5 learners are found in scenario 4;1, which contradicts the assumptions of the EI task:

4;1 on SO targets, both tests are performed above chance; on OS targets repetition is above chance and comprehension is at chance level.

The two remaining scenarios comprising a single learner each (2;4, 2;2) make little sense linguistically, and may be due to the participants' lack of commitment or to random variation in their non-systematic responses.

5.3.1 Effects of additional exposure to the input

The picture presented above describes the situation at T1 (9 hours). This section presents the results obtained after an additional 4h30 of instruction (T2). Figure 5.6 presents the data in terms of scenarios.

The main patterns observed at T1 seem to hold at T2 as well. The largest cluster still corresponds to scenario 2;3, although it now comprises fewer learners, whereas the cluster corresponding to a full morphosyntactic principle (1;1) nearly doubled. Finally, greater dispersion across scenarios is observed than at T1. The same tendencies are represented analytically in Table 5.5.

Scalability analysis (GC = 92.9, p < 0,01) indicates a slightly different hierarchy than observed at T1:

OS repetition ⊃ SO repetition ⊃ OS comprehension ⊃ SO comprehension.

The effect of task at T2 thus appears to be slightly more relevant than that of word order, whereas the opposite was true at T1. Nevertheless, it should be

Figure 5.6: Scenarios, T2

Table 5.5: Implicational hierarchy at T2

OS repetition	SO repetition	OS comprehension	SO comprehension	scenario	n.
+	+	+	+	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	37	42	80		88

pointed out that the difference between the second and the third step of the scale is minimal (5 learners at both test times).

Moving further, it is worthwhile to investigate whether any regularities may be detected in the evolution of learner processing strategies over time. Evolution patterns will be represented as a combination of the scenarios in which a learner is found at T1 and T2.

The first column of Table 5.6 ("pattern") lists all observed combinations of scenarios evolving from T1 to T2. This set comprises 30 items, a small fraction of the full set of possible combinations, amounting to 256 patterns, which shows that evolutionary patterns are not random

The second column shows the proportion of learners adopting each pattern. In the following columns, proportions are computed on the basis of each L1.

The first striking observation regards the lack of clear cross-linguistic patterns in the data. Second, among the five most common patterns, three - comprising 28%, 10% and 7% of the data respectively - indicate no change between T1 and T2. Next, the English L1 group is the most homogeneous - all learners but one are found in scenario 2;3, corresponding to a clear positional principle, and all learners show now change at all from T1 to T2. All other language groups exhibit more dispersion, with most clusters comprising just a single learner, and some representing a few learners.

5.4 Inferential statistics

To statistically verify the tendencies identified so far, a generalised linear mixed model with binomial error structure and logit link function (Likelihood Type 3-test) was fitted to the data. The dependent variable is given by a matrix reporting each learner's successes and failures for a given combination of predictors. Control predictors include task type (binary factor, reference level=EIT), word order (binary factor, reference level=OS), L1 (factor, EN, FR, GE, IT, NL, reference level=ENG) and test time (binary factor, reference level=T1). The interactions which proved significant in the previous analyses (i.e. L1:word order and time:word order, see sections 3.2.3 and 4.2.1) were also added. The model is designed to test the two-way interactions concerning the predictor "task", i.e. task:time, task:word order, task:L1, whose underlying hypothesis is that the effect of task type varies depending on target sentence word order, test time and learner L1, respectively.

Random effects include random intercepts for participants and target items as well as correlated random slopes for time, test time and test type.

Table 5.6: Patterns of morphosyntactic processing over time

pattern	tot (n=88)	EN (n=16)	FR (n=17)	GE (n=18)	IT (n=17)	NL (n=20)
2_3 2_3	28%	94%	24%	0%	12%	20%
1_1 1_1	10%	0%	0%	22%	24%	5%
2_3 3_3	8%	0%	24%	6%	0%	10%
3_3 3_3	7%	6%	12%	17%	0%	0%
3_1 1_1	5%	0%	6%	0%	18%	0%
2_1 2_1	3%	0%	0%	0%	18%	0%
2_3 1_1	3%	0%	0%	6%	6%	5%
2_3 2_1	3%	0%	6%	0%	0%	10%
2_3 2_2	3%	0%	12%	0%	0%	5%
2_1 2_3	2%	0%	0%	0%	6%	5%
2_3 4_4	2%	0%	0%	6%	0%	5%
3_1 3_3	2%	0%	0%	0%	0%	10%
4_1 1_1	2%	0%	0%	0%	12%	0%
1_1 1_2	1%	0%	0%	6%	0%	0%
2_1 1_1	1%	0%	0%	6%	0%	0%
2_1 1_3	1%	0%	0%	6%	0%	0%
2_1 4_1	1%	0%	0%	6%	0%	0%
2_2 2_4	1%	0%	6%	0%	0%	0%
2_3 2_4	1%	0%	0%	0%	0%	5%
2_3 3_1	1%	0%	6%	0%	0%	0%
2_3 4_1	1%	0%	0%	0%	0%	5%
2_4 2_3	1%	0%	0%	6%	0%	0%
3_1 3_1	1%	0%	0%	0%	0%	5%
3_3 1_1	1%	0%	6%	0%	0%	0%
3_3 1_3	1%	0%	0%	0%	0%	5%
3_3 3_1	1%	0%	0%	6%	0%	0%
4_1 2_1	1%	0%	0%	6%	0%	0%
4_1 2_3	1%	0%	0%	0%	0%	5%
4_1 3_1	1%	0%	0%	0%	6%	0%
4_1 4_3	1%	0%	0%	6%	0%	0%

Convergence issues unfortunately made it impossible to include a more complex structure. The summary of the model is presented in Table 5.7.

The interactions involving the predictor "test" were tested by comparing the full model described above to three reduced models, each lacking the single interaction of interest (Table 5.8).

Pairwise comparisons show that the predictors interact in a complex way, producing numerous statistically significant contrasts. Performance in the two tasks was usually statistically significant, which confirms the initial hypothesis that the EIT is indeed more demanding than the comprehension task. The only contrasts which proved *not* significant involved the OS word order and the German, Italian and Dutch L1 groups at both test times.

5.4.1 Repetition in the absence of comprehension

The rationale of the analysis presented so far is that learner comprehension and repetition scores combined should provide a comprehensive picture of the principles of utterance organisation observable in the learner variety. The validity of this approach relies on the assumptions of the EI task, namely that target repetition is impossible without its comprehension. Phonological memory should not play any significant role in this test.

Nevertheless, a few learners appear to violate this assumption. For each relevant combination of test time and word order, Table 5.9 provides comprehension and repetition scores of learners who at least at one test time appeared in scenario 4, along with the probability of observing such a 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 repetition and comprehension scores are consistently very high or very low, which excludes the possibility that the subject were assigned to scenario 4 only because they slightly exceeded score thresholds.

Participants in scenarios 2;4 and 4;4 exhibit higher scores in repetition than comprehension. This behaviour is hardly explicable in that they fail to score above chance in the comprehension of SO targets, which can be indifferently processed based on word order or inflectional morphology. A single participant is located in scenario 4;3, which surprisingly indicates above chance repetition of OS, but not SO targets, and just the opposite situation in the comprehension test. Such behaviour seems rather erratic and does not lend itself to a specific explanation. It must be mentioned, nevertheless, that results may slightly inflated because of repeated statistical testing.

Table 5.7: Model output

Predictors	mat		
	Odds Ratios	*CI*	*p*
(Intercept)	0.06	0.02 − 0.17	**<0.001**
time	3.13	2.27 − 4.33	**<0.001**
L1 [FR]	3.27	0.77 − 13.89	0.108
L1 [GE]	15.64	3.86 − 63.46	**<0.001**
L1 [IT]	8.37	1.89 − 37.05	**0.005**
L1 [NL]	2.52	0.63 − 10.20	0.193
WO2 [SO]	315.01	135.99 − 729.67	**<0.001**
test [EIT]	0.06	0.02 − 0.21	**<0.001**
WO2 [SO] * test [EIT]	0.07	0.04 − 0.10	**<0.001**
L1 [FR] * test [EIT]	4.35	0.93 − 20.43	0.062
L1 [GE] * test [EIT]	18.74	4.12 − 85.13	**<0.001**
L1 [IT] * test [EIT]	16.60	3.40 − 81.05	**0.001**
L1 [NL] * test [EIT]	12.92	2.91 − 57.40	**0.001**
time * test [EIT]	0.77	0.54 − 1.08	0.126
time * WO2 [SO]	0.46	0.32 − 0.64	**<0.001**
L1 [FR] * WO2 [SO]	0.34	0.15 − 0.80	**0.012**
L1 [GE] * WO2 [SO]	0.06	0.03 − 0.14	**<0.001**
L1 [IT] * WO2 [SO]	1.01	0.40 − 2.55	0.979
L1 [NL] * WO2 [SO]	0.44	0.20 − 0.97	**0.042**
Random Effects			
σ^2	3.29		
τ_{00} participant	3.30		
τ_{00} participant.1	1.00		
τ_{00} participant.2	2.19		
τ_{11} participant.testEIT	4.03		
τ_{11} participant.1.WO2SO	0.52		
τ_{11} participant.2.time	1.30		
ρ_{01} participant	-0.82		
ρ_{01} participant.1	-1.00		
ρ_{01} participant.2	-1.00		
ICC	0.41		
N $_{participant}$	91		
Observations	720		
Marginal R^2 / Conditional R^2	0.494 / 0.704		

5 A comprehensive view of morphosyntactic skills

Table 5.8: Single-term deletion

predictor	Chisq	Df	Pr(>Chisq)
task : word order	191.892	1	< 0.01
task : L1	17.557	4	< 0.01
task : time	2.339	1	> 0.05

Table 5.9: 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

The facts reported above should induce a little caution as to the assumptions of the EI task. This section therefore aims to verify whether or not it is really possible to perform the EI task in the absence of comprehension. To this end, the VILLA EI task was administered to new groups of Italian, French and German participants selected on the basis of the VILLA guidelines (see §??)[1]. These new test-takers were not exposed to any Polish input, so that it was impossible for them to process target sentences for meaning: the only skills they brought to the task was their phonological memory. Their performance therefore should be comparable to that of a VILLA subject who did not process targets for meaning, but only repeated them as a string of sounds. Will these participants with no comprehension skills be able to repeat ACC case endings? If that were the case,

---

[1]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 Wątorek at the CNRS SFL, Paris, and by Christine Dimroth and Johanna Hinz at Münster University, respectively. Sincere thank to all of them for their helpful effort.

we should conclude that the VILLA EI task does not fulfil the assumptions of this kind of task.

A few examples of repetitions of the target sentence in (4a) are presented in (4b-d).

(4) a. /ʤev'ʧink-e   'tɕɔngnje portu'galk-a/
         little.girl-ACC pulls   Portuguese.woman-NOM
     b. ['ʧefnie ne'tswo na portu'gala]   (German group, subject 2)
     c. [dʒi'kinʧi 'kɔnʧe portu'garʧe]   (French group, subject 1)
     d. [tsip'tirne ʧo portu'gal kta]   (Italian team, subject 3)

The two non-transparent words /ʤevʧinke/ and /tɕɔngnje/ are hardly recognisable, whereas the transparent word in final position sounds decidedly closer to the target, although accurate repetitions only concern that part of the word which is recognisable in both the target language and the subject's L1, namely the stem /portugal/. Suffixes and inflectional endings are mostly omitted or substituted with random linguistic material. At the same time, 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, as [e] in [tsiptirne] (4d). Since processing for meaning is to be excluded, one has 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 [e] in [portugarʧe] in (4c) for target [a] in /portu'-galka/.

Nevertheless, comparing the output of the VILLA learners to that of first-exposure participants is not necessarily a legitimate operation. The examples in (5b-l) report the repetition of the target sentence in (5a) as performed by learners who perform above chance in repetition, but not comprehension. Leaving inflectional endings aside for the moment, the output produced by these learners is quite different from that of the informants in (4), as lexical items are clearly recognisable and produced with considerable accuracy. The overall picture will be discussed and interpreted in the following Chapter 7 on semi-spontaneous production.

(5) a. /ʤev'ʧink-e   'tɕɔngnje portu'galk-a/
         little.girl-ACC pulls   Portuguese.woman-NOM
     b. [dʒew'ʧinke 'ʧɔngnie portu'galka]
     c. [dʒew'ʧenkə portu'galska]

    d. [dʒewˈʧɨnknɛ ˈportu ˈporta ˈbazu port portuˈgalka]

    e. [dʒjefˈʧinka dʒ ˈʧɔɲɲe portuˈgalka]

    f. [dʒjefˈʧinka kn eh ˈʧɔɲe portuˈgalkon]

    g. [portuˈgalka]

    h. [dʒewˈʧinke ˈʧɔngnie portuˈgalka]

    i. [dʒewˈʧɨnə ˈʧorgo portuˈgalka]

    j. [dsziewˈʧɨnkɛ ˈʧgnie portuˈgalka]

    k. [portuˈgalka ˈʧɔɲe dʒefˈkinkje]

    l. [portuˈgalke ˈʧɔngnie dʒewˈʧinke]

5.5 Conclusion

Clear tendencies emerge from the analysis of morphosyntactic skills in the structured tests, pointing to the relatively greater difficulty of the EI task and of OS targets. Even though the majority of learners consistently apply a positional principle of utterance organisation, it is an impressive result that at least a fraction of them seems to be able to apply a morphosyntactic principle after only 9 hours. Their number 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.

6 Semi-spontaneous production

6.1 Research questions and rationale

Following the analysis of learner performance in the structured tests, the present chapter aims to observe learner morphosyntactic skills in a more realistic communicative situation, arguably closer to real language use. To this end, it presents and discusses the output elicited through a semi-spontaneous production task in which learners took part in pairs or small groups. Because of the amount of work required to transcribe and analyse such interactional data, only a subset of the database (the Italian Meaning Based edition of the project) will be considered.

After a qualitative analysis of learner-produced utterances, the study will apply the same statistical tool employed in the previous chapters in order to determine whether or not learners may thought to apply a morphosyntactic principle in their output. The results are then compared to the scenarios identified in the previous chapters in order to appropriately collocate semi-spontaneous production along an implicational scale of task difficulty.

6.2 An overview of learner output

In learner's utterances, new referents 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 instantiated by the name of the character (1).

(1) a. [ɔna jɛst ʤovann-a].
        she is   Giovanna-NOM
        'She is Giovanna.'

    b. [ʤovann-a   ɛst nautʃitʃɛlk-on].
        Giovanna-NOM is teacher-INS
        'Giovanna is a teacher.'

6 Semi-spontaneous production

When a referent has been introduced, it is typically referred to using personal pronouns (2), although person names may be repeated in consecutive utterances (3):

(2) a. [to   jest marj-a].
         this is   Maria-NOM
         'This is Maria.'

    b. [ɔna lubi herbat-a].
         she likes tea-NOM
         'She likes tea.'

(3) a. [to   jest marj-a].
         this is   Maria-NOM
         'this is Maria'

    b. [marj-a   jest njemk-on   i   tumatʃk-on].
         Maria-NOM is   German-INS and interpreter-INS
         'Maria is German and an interpreter.'

Zero anaphora can be encountered (4b) following utterances in which the subject is expressed by either a name or a pronoun (4a).

(4) a. [ɔna   jɛst portugalk-on   (...)].
         she.NOM is   Portuguese-INS
         'She is Portuguese.'

    b. [i   zna   jɛski   angjelsk-i].
         and knows language.NOM/ACC(?) English-NOM/ACC
         'And she speaks English.'

No examples can be found in which common nouns express the subject function.

Nouns in the object function may be correctly marked as accusative, both in a sequence of feminine nouns only (5a) and in sequences containing both feminine and masculine nouns (5b, where *kot* 'cat' and *pies* 'dog' are masculine):

(5) a. [ɔn lubi tʃokolad-e   kav-e   i   erbat-e].   (5106)
         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].   (5111)
   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, occurrences can be found in which all feminine nouns in the object function are marked as nominative (6):

(6) a. [ɔna lubi herbat-a].   (5119)
        she likes tea-NOM

        'She likes tea.'

     b. [ɔna lubi tʃekolad-a   kav-a   i erbat-a].   (5107)
        she likes chocolate-NOM coffee-NOM and tea-NOM

        'She likes chocolate coffee and tea.'

In other cases still, feminine nouns with the object function seem to randomly appear with a nominative or accusative endings, with no apparent regularity (7):

(7) ɔna lubi erbat-e i   kav-a.   (5107)
     she likes tea-ACC and coffee-NOM

     'She likes tea and coffee.'

Errors in the case marking of the object most commonly involve an overextension of the NOM ending [a]. Marginal non-target-like endings include bare consonants and [ən], with 6 and 1 instances respectively (Table 6.1). In some cases the influence of other known languages can be hypothesised, as in [matema'tik] as opposed to German *Mathematik* /matema'ti:k/. In other cases, the ending may be modelled on other word forms present in the input: in [kerbatən], for instance, the ending [ən] may be a trace of the instrumental masculine ending *-em* [em].

As such non-target endings were produced by only four learners, it seems that this phenomenon should be a matter of individual variability whose causes are beyond experimental control.

No examples of OS word order were found in the data.

6.2.1 Morphological variability and relation to the input

The present section aims to describe morphological variability among the lexemes which occur in the OBJ function. The purpose of this step is to verify whether or not the semantics of specific lexical items associates them more closely to either syntactic function (as indeed is the case in the input, see Chapter 3) and consequently to a specific inflectional ending.

Table 6.1: Semi-spontaneous production task, 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'matɨke/
5118	[paolo lubi matema'tik i mu'zik].	[mu'zik]	/'muzɨke/
5115	[on lubi literatura i matematik].	[matematik]	/mate'matɨke/
5117	[ɔn lubi matematik i psa].	[matematik]	/mate'matɨke/
5117	[marta lubi kav i kerbatən].	[kav]	/'kave/
5117	[marta lubi kav i kerbatən].	[kerbatən]	/her'bate/

To this purpose, Table 6.2 lists the lexemes produced by learners along with their English translation. For each entry, the table indicates first the citation form and its English translation, then the overall accuracy with which the word received accusative marking ("mean"). This value is computed as the ratio between the number of accusative forms and the total number of occurrences ("freq") produced by all participants in contexts expressing the Object function. The following column ("participants") indicates the number of participants who produced the lexeme (regardless of how it was inflected). 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. Figures are presented relatively to the NOM and ACC cases as well as and cumulatively for all other cases combined ("other"). Given the nature of the task, not many occurrences were elicited for each lexeme: the most common item (*literatura*, 'literature') occurred 17 times in the whole data-set, while 6 words (e.g. *rodzina*, 'family') only occurred once.

The following analysis is limited to common nouns, which leads to the exclusion of the names *Anna, Ewa* and *Chorwacja*, 'Croatia'. The rationale for this decision is that proper names may not be treated as common nouns, despite the fact that, in Polish, they are inflected along the same inflectional paradigm.

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

Regarding morphosyntactic accuracy, the whole continuum from 100% to 0% is represented. As made clear by example (7) above, different nouns may receive either marking even within the same utterance.

It may be hypothesised that this variability in morpho-syntactic accuracy may result from a biased distribution in the input: if a given lexeme mostly occurs in

Table 6.2: Lexemes produced by learners in interaction

		output			input			
lexeme	translation	Mean	Freq.	participants	NOM	ACC	other	Tot
Chorwacja	Croatia	1.00	2	1	1	0	2	3
Ewa	Ewa	1.00	2	1	123	4	34	161
rodzina	family	1.00	1	1	32	0	12	44
tata	dad	1.00	1	1	54	19	0	73
żona	wife	1.00	2	1	44	4	0	48
herbata	tea	0.78	9	9	16	52	0	68
kuchnia	cuisine/kitchen	0.75	4	3	20	22	0	42
literatura	literature	0.71	17	11	1	53	0	54
matematyka	math	0.57	7	5	0	9	0	9
kola	coke	0.50	2	2	8	30	0	38
mama	mum	0.50	2	1	54	8	0	62
żaba	frog	0.50	2	2	61	48	0	109
piłka	ball	0.33	3	3	30	23	0	53
czekolada	chocolate	0.30	10	6	13	58	0	71
kawa	coffee	0.30	10	6	36	48	0	84
Anna	Anna	0.00	1	1	243	0	2	245
córka	daughter	0.00	6	5	27	4	0	31
lalka	doll	0.00	1	1	22	7	9	48
muzyka	music	0.00	1	1	1	19	0	20
pizza	pizza	0.00	1	1	12	31	0	43
siostra	daughter	0.00	3	2	19	13	3	45

a specific word-form, then learners may associate it with the corresponding syntactic function or, at least, with the corresponding case marking. To exemplify, if a word only occurs in the accusative case, like *matematyka,* 'mathematics', learners may note and remember it in its accusative form only. If this is the case, then accuracy for accusative case marking should be very high, in principle 100%. In addition, this word-form should overextend to all others, including the nominative case: in other words, the basic word-form of this lexical item should be modelled on the accusative case.

One way to explain the observed variability in the accuracy of ACC marking is to hypothesise that this might be influenced by the proportion of instances in which a given lexeme appears in that word-form in the input. The more commonly the word appears in the input as ACC as opposed to NOM, the more accurately it should be marked as ACC in learner output as well.

If this is the case, plotting the relative frequency of accusative forms against case-marking accuracy should result in a straight line with a positive slope. Fig-

ure 6.1, however, shows no apparent pattern, suggesting that the expectations are not borne out in the data. A few words which in the input hardly ever occur in the accusative case show a mean accuracy of 100% (e.g. *rodzina* 'family'), while others, whose ACC word form is much more common, exhibit much lower accuracy scores (e.g. *muzyka* 'music').
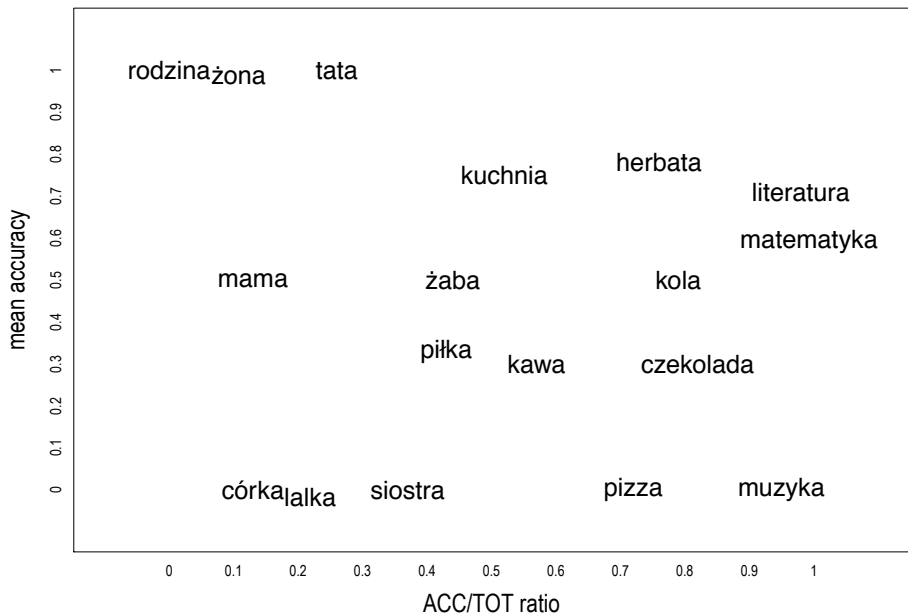
Figure 6.1: Semi-spontaneous production, mean accuracy and ACC/TOT ratio

One may now consider the cumulative frequency of a given lexical item, in order to verify the claim that if a word is very frequent in the input, then it should be more available to the learners, and therefore more easily retrievable. In turn, if a word is easily retrievable, then perhaps the learner could devote more resources to inflectional morphology.

If there were a correlation between overall lexical frequency and grammatical accuracy, lexemes in Figure 6.2 should distribute along a positive slope, with accuracy increasing together with input frequency. Quite clearly, this is not the case. Differences in learner morphosyntactic skills thus do not seem due to a biased distribution of word-forms in the input.

To conclude this section, it is worthwhile to point out that many of the words discussed in the analysis are fairly infrequent in the output data, as they were
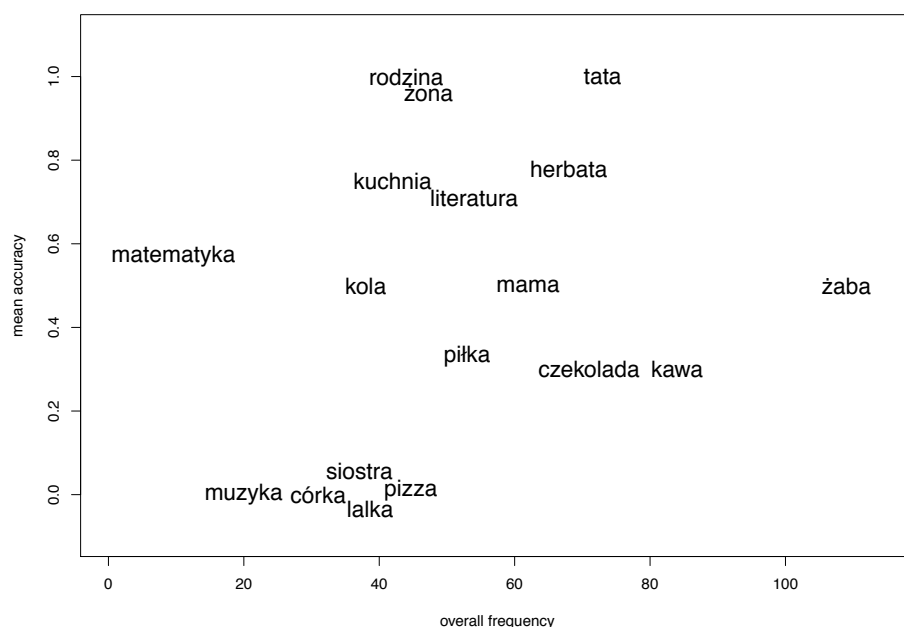
Figure 6.2: Semi-spontaneous production task: mean accuracy / overall frequency

only produced by as few as a single learner. Tendencies regarding the properties of lexical items thus interact with the performance of individual learners, to which the next section is devoted. In any case, the analysis just concluded highlighted no obvious relation between input and output, in spite of the strong tendencies identified in the input (Chapter 3).

6.2.2 Same-word utterances

This section discusses case-marking variability within the same lexeme in the output of individual learners. The question may be pursued by looking at the output of participants with a mean accuracy different from 0 or 1, and in which the same lexical item occurs more than once (Table 6.3). If the rule governing case marking is simply unstable, then repeated lexical items should appear sometimes in their nominative, sometimes in their accusative form, with no apparent logic. If, on the other hand, case marking obeys a systematic principle, then each item should always appear in the same word-form in the same syntactic context.

A few cases (e.g. the single utterance of learner 5115) are evident instances of disfluencies and self-corrections. Other learners exhibit a more variable picture

6 *Semi-spontaneous production*

Table 6.3: Same-word utterances

participant	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	[on ma kɔta i sən i tsurka].	[tsurka]
5104	[ɔn ma sən i tsurka i fortɛpjan].	[tsurka]
5109	[ɔna lubi kava ərbate leteratura].	[kava; ərbate; leteratura]
5109	[on lubi literatura i matematike].	[literatura; matematike]
5109	[ɔn lu ɔn lubi literature matematike].	[literatura; matematike]
5109	[ɔna lubi kava i korvate ʧekolada].	[kava; korvate; ʧekolada]
5109	[ɔn lubi lɔde ʧekolada i korvatje].	[ʧekolada; korvatj]e
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]

of case-marking with the same lexical items. 5109 produces three instances of *literatur-a*, 'literature-NOM', and one of *literatur-ę*, 'literature-ACC'; 5113 produces one instance of *kaw-a*, 'coffee-NOM', and one of *kaw-ę*, 'coffe-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.

6.3 Evaluating a syntactic principle of utterance organisation

In the following section, the statistical tool introduced in Chapter 3 will be applied to the production data discussed in this chapter in order to verify whether or not learners use inflectional morphology in a target-like and systematic manner, that is, following a morphosyntactic principle.

It is worthwhile to start with an overview of the data-set. Table 6.4 reports the lexemes produced by each participant in the ACC form. Not all learners produced all lexemes: the number of participants producing each lexeme ranges from a minimum of 1 (e.g. *żona*, 'wife') to a maximum of 11 for *literatura*, 'literature'.

One can now proceed to verify what principle of utterance organisation each learner may be thought to have adopted. The analysis will only focus on those

Table 6.4: Lexemes produced by learners

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

participants who produced at least three occurrences of a feminine noun in the object function. Table 6.5 indicates the following information: mean score ("mean"); number of correctly case-marked feminine nouns ("correct"); overall number of feminine nouns produced ("contexts"); number of lexical types produced ("lexemes"); ratio between number of utterances and number of lexical types, in which a value of 1 indicates that each lexeme occurs in only one utterance, while higher values indicate that at least some occur more than once. The last three parameters are useful to obtain a more complete picture of the interlanguage: while high mean scores might suggest that the learner has mastered the L2 morphosyntactic system, a reduced number of utterances might lead to questioning this claim. By the same token, few lexemes might suggest that the learner is not applying a rule, but only replicating chunks extracted from the input and not necessarily analysed in terms of morphosyntax. Following the approach described in detail in Chapter 4, the last column indicates the probability that learners achieved the observed scores or higher if they were not applying a systematic morpho-syntactic principle.

A few learners have a p value close to 0, which seems to suggest a systematic use of the target morphosyntactic principle. Learner 5106 made no errors at all; regarding 5102, the passage in which her two errors occur is reported in (8):

6 *Semi-spontaneous production*

Table 6.5: Semi-spontaneous production task, morphosyntactic principle probability by learner

participant	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
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
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
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

(8)   a.   [ɔna ɔn ma brat-a].   *STU
         she he has brother-ACC

      'She has (a) brother.'

      b.   [ma siostr-a].   *STU
         has sister-NOM

      '(She) has (a) sister.'

      c.   [i   ma dzurk].   *STU
         and has daughter.?

      'And (she) has a daughter.'

No convincing explanation could be found for the item [dzurk]. Apart from the examples just discussed, learner 5102 proved fairly accurate over a wide range of utterances and lexical items.

At the other end of the spectrum, two learners (5118 and 5119) did not produce any accusative marking (three and eight obligatory contexts, respectively). 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 p values are presented in Table 6.5 instead of a binary

classification, as was done, on practical grounds, in the chapters devoted to the structured tests. For reasons of consistency with the previous analysis, though, 10% can be taken as a working threshold. P values below this figure indicate that the probability of rejecting the null hypothesis (learners achieved the observed results without systematically applying a morphosyntactic principle) when this is in fact true is lower than 10%.

Regarding the participants whose p values are above 0.1, it cannot be firmly asserted 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 must have noticed some morphological variation in the input, identifying the word forms in which lexical items may appear. 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.

6.4 Correlating the structured tests with semi-spontaneous output

A further question may be whether or not the learners' morphosyntactic skills differ depending on the task through which they are elicited. The previous chapters described learner performance as observed in a structured test, while the present analysis focuses on semi-spontaneous production. The two contexts are quite different from each other in at least two respects. First, the structured tests present an ideal, yet artificial environment for the use of the target structure, while the production task recreates a realistic communicative situation in which the L2 is used not as part of an exercise, but in order to achieve some goal. Secondly, the production task may appear more complex from a cognitive point of view, which in some models, like Skehan & Foster's (2001) Limited Attentional Capacity Model, should produce poorer performance because of the dispersion of attentional resources it brings about. Therefore, the present research question may be summarised as "what can learners do in a realistic communicative situation given their results in the structured tests, which should elicit their very best theoretically possible performance?".

For each participant who produced at least three obligatory contexts, Table 6.6 presents the p-value computed in the preceding section, representing the probability of observing this proportion of correct case marking or higher if the learner is *not* applying a morphosyntactic principle. The last column reports the learners' global score in the structured tests after comparable input exposure (T1),

expressed in terms of scenarios (see Chapter 6).

Table 6.6: Correlation between spontaneous interaction and structured tests

participant	semi-spontaneous production p	structured task scenario
5102	< 0.01	1;1
5101	0.11	1;1
5112	0.13	1;1
5113	0.50	3;1
5118	0.88	3;1
5117	0.31	2;1
5115	0.66	2;1
5119	1.00	2;1
5107	0.50	2;3
5114	0.81	2;3
5106	< 0.01	4;1
5109	0.50	4;1
5104	0.69	4;1

All participants with a p value below 10% belong to scenario 1;1. It thus seems that in order to be able to systematically produce case inflection in spontaneous production, a learner must be able to process SO and OS targets in both the EI and the comprehension test, although as mentioned no OS utterance was observed in the production task. Even learners who successfully repeated the ACC ending in the SO, but not OS targets of the EI test failed to do so in their spontaneous output.

6.5 Interlanguage principles of utterance formation and interpretation

The analysis so far has shown that only a minority of participants consistently uses morphology to express meaning in their semi-spontaneous output: yet all of them managed to successfully complete their task. On what principles did they rely then to express and decode meaning? The following qualitative analysis aims to identify the linguistic means which allow learners to successfully identify or express the intended meaning.

Most often, the referents involved in the utterance differ in their animacy, whereby the animate referent has the greatest probability of being the experiencer (9a). When referents do not differ in their animacy, default SO word order can be relied on (9b). 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.

(9) a. [ʤulj-a   lubi herbat-e kol-a   i   tʃokolad-a]   (5114)
         Giulia-NOM likes tea-ACC   coke-NOM and chocolate-NOM

     'Giulia likes tea, coke and chocolate.'

    b. [luk-a   xoxa hann-a].   (5114)
        Luca-NOM loves Anna-NOM

     'Luca loves Anna.'

Animacy contrasts and default word order structure the output of all learners, even of those who appear to be able to use inflectional morphology productively (10).

(10) a. [ɔna lubi literatur-e].   (5102)
          she likes literature-ACC

     'She likes literature.'

     b. [i   ɔn ma ʒɔn-e   ev-e]   (5102)
         and he has wife-ACC Ewa-ACC

     'And he has a wife, Ewa.'

The same principles operate in native speech as well, as witnessed by the input analysis presented in §3. The vast majority of transitive utterances involves both a contrast in animacy and default SO word order (11a); if the utterance has an OS structure, (11b), animacy still ensures that meaning can be easily decoded. In case there is more than one animate referent (11c), correct decoding may rely on SO word order alone. Only in a minority of utterances is morpho-syntactic analysis indispensable to decode meaning correctly, as both referents share the same value of animacy in the presence of marked word order (11d).

(11) 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.'

6.6 Summary

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

The analysis shows that 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. 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 §6), in which, based on their semantics, specific lexical items are more likely to perform the subject or object syntactic function.

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. The output of those learners who repeat the same lexemes more then once showed that alongside limited variability, the same word tends to occur in the same word-form when it is repeated. Although this observation seems to suggest that learners memorised the lexical item in a specific word form, the limited amount of data does not allow for any generalisations.

The statistical analysis of case-marking shows that only a few learners inflect nouns with above chance accuracy. 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 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 is sufficient to express the meaning required in the task. Indeed, the vast majority of input utterances can be interpreted on their basis as well.

7 Discussion

The present chapter summarises and discusses the results obtained in the preceding chapters.

7.1 Input

This chapter aimed to scrutinise the controlled VILLA input in order to appropriately contextualise the learner task in the linguistic tests discussed later on in the book.

The analysis of the association between case endings and the corresponding syntactic functions showed that, indeed, the ending -[a] is more strongly linked to the SUBJ function than the ending -[e] is linked to the OBJ 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.

Further, while the object function was characterised by relatively high type variety, with numerous inanimate nouns performing it, 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. The high type variety 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. However, since the VILLA project was not designed to investigate this particular research question, it is impossible to pursue it any further based on the present data.

The input was then scanned for all possible sentence models, described in terms of the combination 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;

7 Discussion

- animacy (animate vs. inanimate).

Not surprisingly, only a fraction of the 96 theoretically possible patterns were attested in the input. The trends highlighted by the analysis of type frequency were confirmed: the subject tends to be instantiated by personal pronouns or person names, while the object shows a privileged association with inanimate nouns. The target structures of the two structured tests are rare or absent altogether from the input when all parameters are considered: however, figures markedly rise when a morphological perspective is adopted, whereby word class and animacy are ignored (when legitimate) and patterns are considered as mere sequences of inflectional endings occurring in a given order.

The following sections aim to add a few details which may not emerge sufficiently from an exclusively quantitative analysis, such as the effect of information structure on the frequency of selected morphosyntactic structures. It further discusses the implications of the input distribution just reviewed for the learner task.

7.1.1 information structure

The VILLA input was designed to allow for rigorous experimental control over a large set of variables, but at the same time it was delivered in the form of a communicative, interactive language course. In order not to sound unnatural, the teacher would inevitably produce the structures which she judged pragmatically most appropriate, even if these were *not* the structures targeted in the language tasks. To exemplify, in a context in which the same known characters are mentioned over and over again, it is pragmatically appropriate to refer to them using personal pronouns or their names, like *ona* 'she' or *Julia*, rather than a common noun indicating their nationality of profession, like *kucharka* 'cook'. Together with the structure and contents of the course, this pragmatic constraint led to a very low frequency of constructions targeted in the VILLA structured tests, which exclusively include common nouns.

Across all models of transitive structures, 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 trend is not surprising if one considers the topics covered throughout the course, which for the most part described a handful of human characters along with their likes and dislikes and their relation with each other. Human referents clearly have the greatest chances of being the subject of transitive sentences for obvious extralinguistic reasons. The distribution of pronouns and person names to refer to the

same human referent, in contrast, is regulated by information structure. In a typical VILLA input sequence, pronouns are commonly used to refer to entities which have been previously introduced using a person name (1), although such topic maintenance is also frequently achieved by repeating the character's name as well.

(1) a. Juli-a   lubi żółwi-a.   TEA*
         Julia-NOM likes turtle-ACC
         'Julia likes the turtle.'

     b. ona lubi lizaki   lubi lizaki.   TEA*
         She likes lollipops.ACC likes lollipops.ACC
         'She likes lollipops.'

     c. ona lubi żółwi-a i   lubi czekolad-ę.   TEA*
         She likes turtle-ACC and likes chocolate-ACC
         'She likes turtle and chocolate.'

Entities are mainly introduced using person names rather than common nouns for reasons related to discourse. This claim can be best instantiated on the grounds of the context in which the class is working on the PowerPoint slide in Figure 7.1, trying to decide what course character likes or owns each of the objects depicted therein.

Based on information previously provided during the lesson, the learners can decide between Julia and Filip (top right), both well known course characters. In this particular communicative context the objects represent the discourse topic, while the two children are in the focus position. The sequence opens with the utterances in (2):

(2) a. kto lubi czekolad-ę?   TEA*
         who likes chocolate-ACC?
         'Who likes chocolate?'

     b. eh Gaston[1]?   TEA*

     c. tak Juli-a   lubi czekolad-ę.   TEA*
         Yes Julia-NOM likes chocolate-ACC
         'Yes, Julia likes chocolate.'

     d. czekolad-ę   lubi Juli-a.   TEA*
         chocolate-ACC likes Julia-NOM
         'Julia likes chocolate.'

[1]Gaston is the pseudonym of one of the learners.

Figure 7.1: PowerPoint slide from the VILLA input

The teacher first asks Gaston who likes chocolate, whether Julia or Filip. Input transcription at this stage is only available for the teacher's speech and does not comprise the learner's response, but only teacher feedback. Judging on the teacher's third turn, however, Gaston's answer must have been correct, at least in terms of content; in any case, the teacher repeats (or recasts) the learner's response. Even though the 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 utterance-initial *Julia*, which highlights her as the sentence focus. Only later will the teacher produce the equivalent OS utterance.

Judging on the apparent interchangeability of the two word orders, one may wonder if learners even deemed it necessary to pay attention to such syntactic devices, since different syntactic structures correspond to identical meaning. Speakers of languages which also allow the functional 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. On the other hand, the school-like context in which the project was carried out might have prompted students to pay attention to these details of the target grammar even if it seemed difficult to associate competing forms to the corresponding meaning.

This example is precious to understand two important points. First, not only are OS sentences more marked than their SO equivalents, 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, the example clarifies why person names are so much more frequent than common nouns in transitive structures. Teacher speech is mainly based on PowerPoint slides which depict the same characters over and over again. Thus, even if each course character is identified by a particular nationality and profession, expressed in turn by common nouns, the course characters become so familiar that it would seem somewhat 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 EI task required learners to process common nouns in the absence of any context, something which they could arguably be ill-equipped to do at such an early stage of acquisition and on the sole basis of the input just described.

7.1.2 Form-function association

The analysis of form-function associations has shown that, based on a statistical analysis of the input, it is simpler to associate [a] to the subject function than [e] to the object functions. Before moving on to a more detailed discussion of the mechanisms of such an analysis, it seems worthwhile to point out a few important details which may prove helpful to provide a more comprehensive picture of the learner's task in the VILLA project.

The first is that while subconscious input analysis and associative learning certainly play a role in SLA, there are many other factors which may concur to explain learner behaviour. As far as the nominative case in [a] is concerned, for instance, it most often coincides with the citation form of lexical items, i.e. the form which was usually introduced first throughout the course and which was used out of context. 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 example (3), the teacher first uses the noun in the accusative case, then asks the class to repeat it aloud in the nominative.

7 Discussion

(3) a. ona lubi kuchni-ę   włosk-ą.   TEA*
        she likes cuisine-ACC Italian-ACC
        'She likes Italian cuisine.'

    b. proszę mówić kuchni-a   włosk-a.   TEA*
        please say   cuisine-NOM Italian-NOM
        'Please say Italian cuisine.'

Similar factors, while not quantitative in nature (a word initially introduced in the nominative case may be then used predominantly in the accusative case, e.g. *herbata* 'tea') certainly contribute to the prominence (here understood as the possibility of remembering it) of one or another form.

Further, widespread morphological syncretism may hinder the univocal identification of form-function associations. In perfectly legitimate sentences like (4a) and (4b), nouns performing different syntactic functions are marked by the same inflectional ending [a] because they belong to different inflectional paradigms. Curiously, in this respect such utterances resemble those produced in the EI test by learners who cannot yet manipulate inflectional morphology (4c, here transcribed in standard orthography).

(4) a. siostr-a   woła brat-a
        sister-NOM calls brother-ACC

    b. ona ma brat-a.
        she has brother-ACC

    c. * dziewczynk-a ciągnie portugalk-a
          little.girl-NOM pulls   Portuguese.woman-NOM

The two models can only be distinguished based on the grammatical gender and animacy of the two nouns involved, because the endings they exhibit are formally identical. This may easily confuse learners, as it adds a further factor to take into consideration when computing the form-function association between case endings and syntactic functions: not only is the syntactic function relevant, but grammatical gender also needs to be accounted for. This in turn is not predictable, although in the case of human nouns it almost always coincides with biological sex. There are exceptions to this rule, though: in (5), both the subject and the object are realised by nouns inflected according to the feminine paradigm, namely *córka*, 'daughter' and *tata*, 'Dad'. The latter, however, is semantically masculine. Nevertheless, it should be pointed out that although fairly frequent, the word *tata* is the only lexical item characterised by such properties.

(5) córk-a   Juli-a   kocha tat-ę.
     daughter-NOM Julia-NOM loves dad-ACC
     '(The) daughter Julia loves (her) father.'

A further point of complexity in the VILLA input is represented by the fact that in Polish the default case of direct objects under the scope of negation is not the accusative, as one would expect, but the genitive. In the paradigm of masculine animate nouns, the [a] ending characterises both the accusative case and the genitive (6a). In the feminine paradigm, on the contrary, the two endings are clearly distinct, so that direct objects are marked by different case endings depending on whether or not their verb is negated (6b). Finally, the genitive is also the case in which the subject appears when an existential verb is negated (6c).

(6) a. nie ona nie kocha dziadk-a   Karol-a.
         no she not loves grandpa-GEN Karol-GEN

         'No, she doesn't love grandpa Karol.'

     b. babcia   nie lubi muzyk-i.
         grandmother not likes music-GEN

         'Grandmother doesn't like music.'

     c. nie ma Karol-a
         not has Karol-GEN

         'Karol is not here.'

The last example is crucial in that it confuses the relation between syntactic function and inflectional ending. These regularities are quite systematic and easily described if basic meta-linguistic concepts and rules are introduced, but the VILLA input included no such explanations.

Another important point concerns the selection of the meaning engaged in the form-function association. In the present analysis "subject" and "object" were chosen because they correspond most accurately to the meaning expressed in Polish by the morphemes [a] and [e], respectively. There is no guarantee that the learner identified the same relation, however: in fact, it may be argued that such top-down expectations resemble Bley-Vroman's (1983) comparative fallacy, whereby the interlanguage is analysed not in terms of its internal organisation, but of the target it is supposed to imitate. In fact, different learners may assign different meaning to the same morpheme. Bernini (2018b) and Dimroth (2018: 28-33) both discuss two forms of the word *strażak* 'fireman' as can be encountered

in narrations produced within the Italian and German VILLA editions, respectively, whereby a form in *-k* (e.g. [ˈstraʒak]) modelled on the nominative case opposes a form in *-em* (e.g. [straˈʒakjem]), modelled on the instrumental case *strażakiem*. In the Italian data, the opposition seems to vehiculate the functions "subject/controller" vs. "oblique", while in the German data a "singular" vs. "plural" seems more probable.

The learner's task is further complicated by the differential object marking (DOM) encountered in the masculine paradigm. Nouns referring to things appear in a form identical to the nominative case, characterised by a zero morph attached to the consonantal stem (7a and 7c), while the accusative case of masculine animate nouns (7b) present an [a] ending (7d), which is also found in the genitive case (7e). This last observation highlights the fact that DOM complicates the association between form and function on metalinguistic, rather than statistical grounds: on hearing the two forms in (7d) and (7e), the learner can be hardly expected to identify a comprehensive morphosyntactic rule, especially in the absence of an understanding of the category of case and detailed information as to Polish inflectional morphology. It must be said, however, that all VILLA L1s except German do not inflect nouns for case, so that encountering the same word form in different syntactic functions should not be particularly problematic for speakers of these languages. Nonetheless, the typological difference between the VILLA L1s and Polish is quite evident and a provisional hypothesis should be formulated to account for it.

(7) a. to   jest balonik-∅
         this.NOM is   balloon.M-NOM

     b. to   jest strażak-∅
         this.NOM is   fireman-NOM

     c. Jan-∅   ma balonik-∅
         Jan-NOM has balloon.M-ACC

     d. Jan-∅   zna   strażak-a
         Jan-NOM knows fireman.M-ACC

     e. to   jest samochód-∅ strażaka
         this.NOM is   car-NOM   fireman-GEN

A final point concerns the form component. Discussing the form-function association between [e] and "object", it was stated above that i) most feminine nouns in the accusative case are characterised by word-final -[e], and ii) only a small proportion of words in -[e] are indeed instances of ACC.SG.F. While i) seems

unproblematic, ii) may not seem entirely adequate to model the learner's task during input processing. This point begs the researcher to take a stance, depending upon the answer to the following question: when establishing form-function associations through contingency learning, can learners distinguish words sharing a given form, but belonging to obviously different word classes, and treat them in a different manner? Since in the tasks discussed in this work the VILLA learners were required to process nouns referring to human beings, one of their goals was to identify the forms (i.e. the inflectional endings) in which such words may appear, possibly attempting to discern any regularity governing their distribution. Thus, in the case of ACC.SG.F [e] one may wonder whether it is relevant to know how many input *words* in -[e] really encode human nouns in the accusative case, or whether it is only relevant to know how many *human nouns* in the input end in -[e].

In the former case, the learner will need to analyse all words in -[e] which comprise feminine nouns, indeed (e.g. *portugalk-ę* 'Polish.woman-ACC.SG'), but also verbs (*idzie* 'go.PRES.3SG') adverbs (*dobrze* 'well'), adjectives (*jakie* 'which.NOM/ACC.SG.N') and conjunctions (*ale* 'but'). Upon encountering a word in –[e], learners will (subconsciously) note whether or not it encodes the target meaning, in answer to the implicit question "how often do words in –[e] represent feminine human nouns in the object function?".

In an alternative scenario, learners will separate nouns from all other categories and simply compute a list of the possible endings of human nouns along with their relative frequency (how many nouns with human referents are characterised by word-final –[e]? How many by –[a]? and so on).

The consequences of this decision for the estimation of form-function association are important.

If learners are assumed to be able to distinguish word classes, then only a subset of input words sharing the form under investigation should be considered in the computation of form frequency, which will result in a higher form-function association index, all other things being equal. To exemplify, the range of words in -[e] relevant for the meaning "accusative case of feminine nouns" would comprise e.g. *portugalk-ę* 'Polish.woman-ACC.SG', but not *idzie* 'go.PRES.3SG', *dobrze* 'well', *jakie* 'which.NOM/ACC.SG.N' or *ale* 'but', despite the fact that all share the form of interest -[e]. If learners cannot distinguish word classes, in contrast, then the same count should comprise any word in –[e].

In order to accurately acknowledge the fact that input processing may be selective, it seems appropriate to compute function > form associations based on language exemplars in which the meaning in question is present. From this per-

spective, the learner's task is to identify how many feminine nouns in the object function are characterised by word-final [e], and how many are not.

When this rationale is applied to the VILLA data, the surprising results presented in Table 7.1 and Table 7.2 are obtained. Note that these tables are an elaboration of Table 3.1 and Table 3.2, respectively, in which the columns no longer relevant have been shaded: in fact, following the approach adopted in this section, form-function association is simply the ration between the number of words including both form and function and the number of words encoding function regardless of form. Both association indexes are close to 1, which indeed reflects the fact that in the peculiar VILLA input most NOM.SG.F are characterised by word-final [a], and most ACC.SG.F are characterised by word-final [e].

Table 7.1: Form > function index for syntactically relevant contexts only, ACC.SG.F [e]

a	b	c	d	e	f	g
L1	-[e]	ACC FEM	ACC FEM -[e]	d/b	d/c	e*f
EN	6541	441	415	6.34%	94.10%	5.97%
FR	3789	365	345	9.11%	94.52%	8.61%
GE	6424	414	379	5.90%	91.55%	5.40%
IT	7675	624	598	7.79%	95.83%	7.47%
NL	6570	519	494	7.52%	95.18%	7.16%

Table 7.2: Form > function index for syntactically relevant contexts only, NOM.SG.F [a]

a	b	c	d	e	f	g
L1	-[a]	NOM FEM	NOM FEM -[a]	d/b	d/c	e*f
EN	7091	2471	2437	34.37%	98.62%	33.89%
FR	4239	1798	1730	40.81%	96.22%	39.27%
GE	7100	2338	2265	31.90%	96.88%	30.91%
IT	9345	3198	3161	33.83%	98.84%	33.43%
NL	7483	2787	2721	36.36%	97.63%	35.50%

This observation however is in stark contrast with the results of the tasks discussed in this work, which clearly show that [a] tends to be overextended onto

[e] by a vast number of learners. It seems, therefore, that form-function association may not be the most influential factor to determine which input word-form will be selected as the basic word-form of the learner variety. Based on the tables presented above, raw token frequency appears to be a good candidate, as the instances of [a] NOM are almost six times as frequent as [e] ACC. Moreover, the nominative case appears in a much larger number of contexts than the accusative case, and is the default citation form of nouns. In sum, the approach adopted here in order to model learner selectivity in the computation of form-function association appears heavily biased by the fact that raw frequency is not taken into account.

7.1.3 A learner variety perspective

As Dimroth (2018) points out, it is often difficult to draw a line between the claims and predictions of the learner variety approach and usage-based theories. As far as the input is concerned, particularly, both consider it an essential component for interlanguage development, the raw material which the learner communication faculty will shape in order to reach the set communicative objectives.

From this perspective, inflectional morphology seems by no means indispensable to interpret input sentences, although it certainly is a characteristic and obligatory feature of the target language. It is not surprising then that learners can easily do without it and still communicate effectively, especially when a context is available. The analysis of learner's semi-spontaneous productions has clearly shown that semantics (animacy) and default SO (controller – theme) word order are usually sufficient to express the simple meaning required in the VILLA tasks.

The same trends are also commonly encountered in a vast proportion of input utterances: SO structures are preferred even in a language such as Polish, with its complex nominal morphology and the theoretical possibility to manipulate word order at will. By the same token, agents tend to be animate and patients tend to be inanimate simply because situations usually present this structure. In other words, the principles of utterance organisation in question may be seen as a feature of the basic variety, but are commonly encountered in most instances of verbal communication.

The models corresponding to the target sentences of the structured tests were either absent or rare in the input, precisely because they purposefully eliminate all natural cues to sentence organisation with the exception of inflectional morphology. From a communicative point of view, then, the EIT and the comprehension tasks are little more than exercises targeting meta-linguistic skills. The

communicative principles of the basic variety are hardly applicable, but on the other hand it can be argued that there is hardly any meaning to express.

Since the VILLA participants are all adult, competent speakers of at least one L1, one could argue that their experience in terms of pragmatics and world knowledge may sometimes prevail on the input received. As shown in a study conducted on copular structures (**Saturno2015**), learners often choose to ignore highly frequent input patterns, developing their own interlanguage structures instead. The copular construction with *to* 'this', though extremely common in the input, appeared to be disfavoured both in a structured test and in semi-spontaneous production (Saturno 2018). In the latter context, learners creatively elaborated new, ungrammatical constructions. The structure in (8a) probably has its models in (8b) and (8c).

(8) a. [ɔna jɛst dʒovanna].   (5112)
        She is   Giovanna
        'She is Giovanna.'

     b. to   jest Anna
        this.NOM is   Anna.NOM

     c. Anna jest polką
        Anna is   Polish.woman.INS

In sum, from the perspective of the learner variety approach the results of the linguistic tasks are not particularly unexpected. In the language tasks inflectional morphology is hardly encountered because it is not part of the repertoire of early interlanguages, which prefer to rely on semantics and word order. The same preferences contribute to shaping the input, too, although the latter obviously includes all obligatory traits of the target language, such as inflectional morphology.

7.2 The elicited imitation task

Perhaps the most self-evident result emerging from the analysis of the Elicited Imitation Task (EIT) is that, as expected based on input analysis, the NOM ending [a] shows a marked tendency to overextend onto ACC [e]. For most participants, [a] is indeed the only ending produced, and thus the basic form of nouns, which – if one accepts the theoretical premises of the EI test – should clearly point to a positional principle of utterance organisation, whereby syntactic functions are determined by the relative position of nouns in the utterance.

A much smaller number of learners produce target-like output, in which the endings [a] and [e] alternate depending on the syntactic function of the noun. Such performance – again, based on the theoretical premises of the task – should indicate that the target language morphosyntactic principle of utterance organisation has been correctly identified and can be successfully reproduced in the output. Finally, a set of participants exhibits a variety of complex scenarios.

Word order was found to exert a powerful role, whereby OS targets appear to cause greater difficulties than their SO equivalents. Time of exposure was also shown to be an important factor, whose predictable effect is an increase in performance from T1 from T2, with numerous learners moving closer to the target morphosyntactic principle of utterance organisation. An interaction with word order is observed, too: if partial improvement occurs, it is more likely to be on SO than OS targets. Finally, a weak but significant correlation was found between the LLama test and the score for the repetition of -/e/, averaged for time and word order.

Against this general picture, a few points remain partially unclear. They can be summarised as follows:

- The range of observed case endings;

- Role of word order and processing for meaning;

- better performance on [e] than [a] targets;

- better performance at T1 than T2 targets.

It seems that all these issues ultimately depend on a precise understanding of the mechanism of the EIT, which is itself not completely clear. Therefore, the discussion will start with an attempt to identify the level of analysis into which the EI test may be thought to tap.

7.2.1 Range of case endings

The range of case endings produced by the learners seems to be quite restricted. Such observation is not in accordance with studies on the morphological development in Slavic languages, which suggest that learners first go through a NOM/non-NOM opposition, and only later do they stabilise this generic contrast into a more target-like NOM/ACC distinction (see Chapter 1.

This does not necessarily mean that the VILLA learners acquired case marking better and more quickly than untutored SLA learners. The analysis of the VILLA

7 Discussion

semi-spontaneous production data by Bernini (2016) and Dimroth (2018) shows that utterance structure simultaneously reflects a variety of principles which in spontaneous SLA are typical of different developmental stages, such as the pre-basic variety's 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 learning context, one could propose 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 toler-ance 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 vari-ation (allo-)phonique […], en fondant la possibilité d'oppositions phonémiques". In this respect, he also observes that while several lexical items are relatively sta-ble in their phonological form, others show considerable variability, both in their supposed target and in their phonetic structure, e.g. {ɕpi, ʃpi, spi} for target /ɕpi/, 'sleeps'. Even when the various tokens produced by a learner seem to be map-pable onto specific target forms, and thus to reflect the input to a certain extent, their use is nonetheless functionally differentiated, as it has been shown to be the case with spontaneous SLA (Broeder et al. 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 d) depend on the input. This wealth of data only makes it harder to interpret the output of the EI task, 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 appears to be incontrovertible is that learners must have picked these alternative endings from the appropriate input paradigms, thus showing some sensitivity to it. If one excludes the instances of centralised vowels mainly produced by the German learners, virtually all endings produced are instances of [a] or [e], despite the fact that the relevant paradigm also includes other salient and frequent endings, such as INS [ow̃]. This observation would witness to the fact that learners identified the syntactic context in which the endings [a] and [e] may occur, namely transitive sentences, and conversely that they are aware that in that syntactic context only those endings are allowed.

On the other hand, the fact that only [a] and [e] occur in the EI data is hardly surprising if one accepts that there might be repetition without processing, whereby learners only repeat what they hear without accessing their L2 grammar. In this

situation, only two endings occur in the output simply because only those endings are present in the stimulus sentences.

7.2.2 Sources of error: processing for meaning vs. perception

Compared to previous studies using the EIT, the target structure of the present work introduces additional variables which increase the complexity of the analysis and interpretation of the data. Case marking poses different challenges from other target structures which only affect grammatical correctness, like for instance verb placement as studied by Håkansson (1989) or Schimke (2011). Unlike case marking, the position of a verb in the utterance is not likely to change the overall meaning of the sentence.

In addition, the data produced by the VILLA EIT are limited to the learner's final output. In the absence of a comprehension or translation test, it is impossible to tell what learners meant to say. This is unfortunate, as the output of the EIT is the product of at least three complex processes, namely perception, comprehension and production: errors may lie at any level. By observing the learners' processing of an OS target, for instance, one cannot tell if the underlying grammatical meaning was identified, firstly, and if any effort was made to reproduce it, secondly. It is also impossible to rigorously exclude that learners performed the task without processing targets for meaning. Indeed, this suspicion is further reinforced by the weak but significant correlation between phonological memory capacity as measure by the LLama test and the scores in the repetition of -/e/.

These questions were dealt with in detail in Chapter 6, in which the results of the EI and of the comprehension test were correlated in an attempt to provide a comprehensive picture of learners' processing skills, and will be discussed further in §7.4. For the time being, this section will focus on the clear effect of word order as observed in the EIT alone.

The rationale of word order manipulation in the EIT was that SO targets, being syntactically and pragmatically unmarked, should be easier to process than their marked OS counterpart. This claim is founded on a variety of reasons discussed in Chapter 1, ranging from the typological diffusion of SO as opposed to OS word order, to acquisitional data showing that case marking first develops in SO sentences, to the input analysis presented in Chapter 3, which shows that even within the strictly experimental conditions of the VILLA project, the vast majority of transitive utterances are characterised by an SO structure.

However, understanding the direct impact of these general constraints on the EIT implies a few argumentative steps. For learners to find OS targets harder than

their SO equivalent, it is necessary that they can recognise them as such, which is not obvious. When the OBJ is repeated incorrectly, it receives a basic ending in [a], just like the noun performing the SUBJ function. As both nouns are now marked by an identical ending, what matters to express meaning is their relative position, the first being the SUBJ, the second the OBJ. To express the meaning of the OS target using a positional principle, the learner would need to swap the two nouns: yet this only happens once across the whole corpus. Two alternative accounts may be proposed. The first is that learners understand the OS structure of the target, but since they cannot express the desired meaning using inflectional morphology, they simply renounce to express it at all. This produces an utterance which indeed seems to express a completely different meaning based on a positional principle. The other explanation 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 recognised the lexical items involved without any further specification of their grammatical role, by using an invariable word form ending in [a].

In both cases, it seems that the learners renounced to express the specific meaning of the target sentence, agreeing to repeat a sentence which either did not correspond to the meaning they had identified, or did not corresponded to any meaning at all. Surely this is a powerful argument against the hypothesis that the EIT can be used to approximate spontaneous speech while retaining full control over the target structure. Although not a single learner commented on not being able to express what was really meant, it must be said that because of their lack of context and abstractness, the target structures of the EIT may seem to express a very abstract, generic meaning anyway.

Even if one accepts that learner output may not express any meaning, at least not syntactic, it still remains to be explained why repetition scores are consistently lower on OS than SO targets, although in a situation in which test-takers do not associate case endings to the corresponding syntactic functions, it does not even seem legitimate to speak of word order. Since [a] and [e] do not correspond to SUBJ and OBJ, but are merely two segments, why should there be a difference in scores depending on which one comes first in the stimulus sentence?

The answer may come from perceptual prominence, which in turn is closely related to saliency, however understood: the present work adopts Peters's (1985: 1030) argument 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.

Data on earlier EI task studies show that, indeed, perception may be a relevant factor in explaining the results of this task. Gallimore & 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) and Slobin (1985: 1166) suggest that utterance-initial and utterance-final positions are maximally prominent and accessible for segmentation and storage, whereas utterance-internal positions are harder to access. 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 L2 Polish is affected by word position (utterance initial and final vs. medial) independently of the time of exposure (0, 4 and 8 hrs).

These studies typically considered the perceptual prominence of entire words or free morphemes. But the same rationale can be applied to inflectional morphemes, which for a learner who does not process targets for meaning are indeed mere segments. In terms of perceptual prominence, in SO sentences the ACC ending [e] occurs in utterance-final position, thus gaining maximal prominence (9a). In OS sentences, in contrast, this element always occurs in utterance-medial position, which might make it harder to perceive and consequently reproduce (9b).

(9) a. [nautʃitɕelk-a pxa   studentk-e].
       teacher-NOM pushes student-ACC

   b. [studentk-e pxa   nautʃitɕelk-a].
       student-ACC pushes teacher-NOM

Thus, error distribution could be accounted for by hypothesising that learners are more successful at reproducing target structures if these are more retrievable from a perceptual point of view, as argued in Saturno (2015). In SO sentences, the non-basic ACC ending [e] is in the maximally prominent utterance-final position and stands the best chances of being noticed and processed. The higher error rate in OS sentences, in contrast, may be 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 acoustic 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/.

The varying prominence of the marked [e] ending may be perhaps connected with the bizarre and unexpected instances in which the repetition score is higher in the case of [e] than [a]. This result, setting aside an interpretation based on sheer chance and random variation, probably witnesses to one of the main motors of change in the interlanguage, namely, fear for errors. The learners may have noticed, either from the input or from the test items themselves, that Polish words most of the times present the usual ending [a], but sometimes exhibit the sound [e], whose meaning (if any) might not have been necessarily clear. However, these learners failed to grasp the regularity governing this pattern, while at the same time realizing that they tend to supply [a] in all contexts, which sometimes must be incorrect. For fear of this error, then, they make the opposite one, that is, providing the marked ending [e] slightly more often than required. Finally, one could hypothesise that EIT probes different types of competence depending on the test taker's proficiency level: if indeed targets are filtered through the learner's grammatical system, one should expect more proficient learners do to better at this test because their grammatical system helps them to overcome mnemonic constraints, for instance through "chunking" (Miller 1956), i.e. 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 & Lonsdale (2012), for instance, show that EIT scores significantly correlate with participants' scores on a general English placement test, but not with working memory (WM) scores, and that the lowest-scoring students were unable to repeat anything beyond their WM capacity. In this perspective, even the almost unrecognisable output produced by some participants seems to find a place. If an interlanguage is so undeveloped that the learner cannot recognise lexical items, let alone inflectional endings, then one should expect the EI task to elicit a meaningless string of sounds which vaguely resemble the stimulus sentence.

7.3 The comprehension test

The analysis presented in Chapter 3 attempted to verify whether learners performed an aural comprehension task by relying on a morphosyntactic strategy as opposed to a positional strategy. The results are fairly self-evident, whereby SO targets are processed with far greater accuracy than their OS equivalents. Within the latter group, OSV seem to be more accurately processed than OVS.

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 found, too.

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 OS targets, showing no sign of evolution over time. The cause for this state of things probably lies in the very rigid SVO syntax of English, together with its very limited inflectional morphology, which 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 in different syntagmatic positions. The number of learners correctly processing both OS and SO 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 learners correctly processing SO targets only, in contrast, decreases between the two test times. This result is quite unexpected, as SO 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.

There are, however, a couple of points 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. A 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. However, case endings are not particularly prominent, and depending on the current stage of the interlanguage grammar, they may or may not be attended to by the learners. Finally, because of widespread syncretism across paradigms, they are not necessarily unambiguous if the grammatical gender of individual lexical items is not known. In sum, the positional principle may have a direct impact on the differential processing of OS targets as well. 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 other words, as non-SO. Since only two responses were possible in the VILLA comprehension test, this conclusion appears sufficient.

7.4 A comprehensive view of morphosyntactic skills

Chapter 6 aimed to provide a comprehensive picture of the learners' ability to decode and encode grammatical meaning through inflectional morphology only, without the aid of context or phonology. To this end, the results of the EIT and the comprehension test described in the previous chapters were correlated so as to identify a hierarchy of task difficulty. Learners were grouped together on the basis of scenarios, given by a global score summarising performance in both tests and on targets of either type (SO vs. OS). Four such scenarios, comprising more than 80% of the data set, have a direct, meaningful linguistic interpretation, which provides partial answers to the research question.

Scenarios can be described in terms of a required set of skills, which could be ordered along the following implicational hierarchy: OS repetition ⊃ OS comprehension ⊃ SO repetition ⊃ SO comprehension. 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.

The picture identified at T1 did not change significantly after an additional 4:30 hours of exposure to the input. Although there were slight changes in the size of the clusters, the implicational scale was confirmed at T2 as well. Specifically, the extremes of the continuum appeared to be well confirmed, while some variability occurred in the two intermediate steps, suggesting a similar level of difficulty. The positive effect of additional exposure to the input was made evident by the growing number of learners adopting a morphosyntactic principle, whereby the pure positional principle became less widespread. The data were searched for any preferential patterns of evolution over time, but no clear tendencies could be identified. The most common pattern, in fact, involved no change at all. While the time interval between T1 and T2 was probably too short to produce clear common changes, the great variability of the data witnesses to the development of individual strategies of input processing.

Against this overall picture, a few points require more specific attention.

7.4.1 Relation between comprehension and production

A few scenarios present a coherent, clear-cut situation and can be considered as relatively unproblematic: such is the case of scenario in 1;1, in which all tar-

gets are correctly comprehended and repeated, and of scenario 2;3, which corresponds to a pure positional principle.

Nevertheless, one should consider not only the scores, but also the linguistic operations which scenarios imply. For instance, scenario 1 apparently indicates target-like morphosyntactic processing, but this is not necessarily the case as far as SO structures are concerned. The successful comprehension of this targets may derive from a positional principle, whereas accurate repetition may stem from default post-verbal ACC marking or even rote repetition, if the distractor of the EIT proved insufficient. In sum, it appears that the only reliable context to investigate the learners' use of morphosyntax is OS targets.

Correlating the two tests is essential for the interpretation of the most frequent error encountered in the EIT data, i.e. output in which both nouns are marked with [-a] NOM. Two main cases may be distinguished. If the target is OS, and the learner proved incapable of processing such a structure in the comprehension test (scenario 2 on OS targets), then one can conclude that the participant cannot yet manipulate inflectional morphology to extract and encode grammatical meaning. The same output is more problematic in other situations, namely a) incorrect repetition, but target-like comprehension of OS targets; b) incorrect repetition of SO targets, SO comprehension being achievable through the positional principle if morphology cannot be processed.

The overextension of [-a] NOM onto ACC contexts (requiring [-e] marking) in the EIT, with target-like scores in the comprehension test, may merely signal that the learner cannot yet produce inflectional morphology, although its function in the target grammar has been correctly identified. However, this would imply that the learner correctly interpreted the underlying syntactic structure, but produced output in which both nouns are characterised by identical case endings, which may lead to potential communicative problems in real-life situations. This strategy would indeed cause an incorrect reading if interpreted through the positional principle, which in turn is made inevitable by the identical marking of the two nouns. In other words, the learner might as well understand the grammatical meaning of the stimulus, but is unable to supply the corresponding grammatical markers in the output, perhaps because of the greater cognitive burden exerted by the EIT, or else because the productive use of inflectional morphology is still beyond the current interlanguage stage. However, no learner ever signalled any difficulty in this respect, for instance by stating that what they were saying was not actually what they meant.

An alternative explanation could be suggested. The EI task is inherently quite complex, as learners first have to understand the target, then draw a geometrical figure, and finally repeat the target sentence. It does not seem unrealistic

to think that while learners can fully understand targets when comprehension is exclusively targeted, as in the comprehension test, they may overlook bits of target words when comprehension is part of a more complex task, as in the EI task, which involves, comprehension, memory storage and/or (re)production. It is probable that phonological forms encoding grammatical meaning should be lost first, while phonological forms encoding lexical meaning last longer as a consequence of the limited vocabulary range employed in the test. This claim is supported by research by Ellis & Sagarra (2011), who demonstrated that as task complexity increases, even participants who had proved capable of interpreting inflectional morphology turn their attention exclusively to lexical meaning. In the context of the present work, repetitions in which both nouns are marked as [-a] NOM, even when produced by learners who perform above chance in the comprehension test, may instantiate an underlying structure in which nouns only carry lexical meaning, if any. If this is the case, then at least for some learners the comprehension stage of the EI task does not produce in the learner's mind a complete, "extra-linguistic" picture of the situation described by the target sentence. This claim has crucial consequences for the interpretation of results, which relies on the assumption that the EIT has a reconstructive nature, i.e. asks the participant to describe a given situation in his own words.

7.4.2 Repetition in the absence of comprehension

Similar doubts arise regarding those learners who seem to correctly repeat target sentences in the absence of comprehension of the same type of target, which hints to reliance on phonological memory alone. The distractor must have proved insufficient to saturate the phonological loop, allowing the participant to repeat a string of sounds with no processing for meaning. Since no explanation was provided as to the role of the distractor, some learners would draw the geometrical figure with great care, as though that was an important part of the test. At times it could take them so long that phonological memory could decay spontaneously, so that the distractor could be said to be effective. Others, in contrast, tried to be as quick as possible and focus on repetition, possibly while mentally rehearsing the target during the drawing stage. This approach may have allowed the participant to rely on short-term memory, which could explain scenario 4;1.

To this observation one could add the positive and significant correlation between repetition scores and phonological memory as measured by the Llama test. It appears that the ability to retain strings of sounds in working memory may be of help in the EI task, at least as long as the repetition of grammatical morphemes is concerned.

These facts do not necessarily mean that the VILLA EI task is not effective, though. Indeed, two observations suggest that normally learners do attempt to encode meaning according to the means at their disposal, be it morphosyntax of word order. First, the scores of most learners are consistently lower in repetition than in comprehension. Second, the widespread overextension of the basic word form modelled on the nominative case onto the marked non-nominative ending is in line with present knowledge on early learner varieties.

The problematic output at hand may have a further explanation, lying somewhere in between the extremes of morpho-syntactic processing and rote repetition. In this perspective, the interlanguages in question are sufficiently mature to recognise lexical items, so that these can be stored in working memory not as bare sequence of sounds, but as meaningful chunks. At the same time, form-function associations between case endings and the corresponding grammatical meaning have not yet developed. The learners may well be aware that Polish words can appear in several word-forms: because of the implicit approach of the VILLA course, however, they have not yet identified the rule governing the use of different forms. These learners cannot yet produce meaningful utterances, but they repeat the lexical items in the word-form in which these were identified. Lexical items and word-forms, in other words, are not stored in phonological memory as meaningless strings of sounds, but *recognised* as lexical items in a given word-form, which could be considered as a sort of first step on the way to processing. Most probably, word-forms are not recognised in terms of function, i.e. as "accusative", "object" or "patient", but of form, as in "the form ending in [e]".

7.4.3 Competing scenarios

The analysis of the implicational hierarchy of task difficulty showed that while its extremes are fairly solid, the two medial steps (OS comprehension and SO repetition) received very similar scores and indeed vary in their relative order between T1 and T2. The question may be dealt with in terms of scenarios. Specifically, although scenario 3;3 does not violate any assumptions of the linguistic tasks, yet it is not part of the implicational scale of relative difficulty of task type / word order combinations. This is because learners in this situation systematically fail in the repetition of both SO and OS targets, but process them with above-chance accuracy in comprehension. Learners in the competing scenario 3;1, in contrast, perform accurately in both test on SO targets, but only manage to successfully process OS ones in comprehension. The question is thus whether

the processing of OS targets in comprehension presupposes SO repetition (as in scenario 3;1) or not (as in scenario 3;3).

It seems a viable explanation that this should be a matter of learning style, some learners being more prone to start speaking 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, exhibits dramatic individual variability (Krashen 1985; Granger 2004).

On the statistical side, moreover, one needs to be aware of family-wise errors, which refer to the possibility that because of the size of the data-set, the statistical test may return a small number of false positives, i.e. learners who in reality perform randomly but just happen, by chance, to obtain significant results. While the general tendencies highlighted in the analysis seem rather clear, results as to the exact number of learners achieving a given result should be handled with care. This is particularly true in the case of the implicational scales, which showed that while at T1 OS comprehension implies SO repetition, the opposite is true at T2. The difference in the number of learners achieving a significant result in one condition but not in the other, however, only amounted to five people. Thus, while the two extremes of the scale are very well defined, the two medial steps appear fairly close to each other in terms of difficulty.

7.4.4 Effect of time

The analysis of learner performance at T1 and T2 points to at least three major observations. First, the cluster corresponding to a full morphosyntactic principle (1;1) nearly doubles, suggesting that additional exposure indeed steers the interlanguage towards the target variety. Second, the cluster corresponding to a bare positional principle significantly decreases in size, which indicates that although not all learners fully adopt the target-like morphosyntactic principle, at least they no longer generalise the default first-noun principle in all contexts.

Finally, greater dispersion is observed at T2 than at T1. This could be taken as evidence of various autonomous strategies of input processing being developed by learners as they test hypotheses regarding the structure of the target language.

The analysis of evolutionary patterns did not produce conclusive results, except perhaps for the observation that the bulk of learners tends not to evolve over a 4h30 period. Further, while the L1 proved a meaningful factor in both inferential statistics and cluster analysis, it does not seem helpful to identify preferential patterns of evolution over time. This may be due to the fact that the data set for each L1 group is too limited to identify any significant tendency. Further, the time elapsed between the two test times is probably too short. What one observes is

in fact a picture taken from a collection of individual linguistic systems in fluid development.

7.4.5 Role of the L1

Three main groups may be identified on the basis of the interaction between morphosyntactic skills and L1. The group characterised by superior performance mainly comprises German and Italian learners. The English learners consistently perform above chance level only in the comprehension of SO sentence, the only target which can be successfully processed positionally. The performance of French and Dutch learners lies somewhere in between.

The exclusive reliance of the English subjects on a positional principle hints to the rigid word order of the L1, in which pre-verbal position is the most reliable indicator of subjecthood (MacWhinney et al. 1984). By the same token, the superior performance of the German learners can be explained by the presence of case in their L1, which also licenses flexible word order. The mid-range performance of French and Dutch learners perhaps reflects a similarly intermediate degree of word order flexibility as well as the absence of morphological case on nouns.

The high performance of the Italian learners is somewhat problematic to interpret, as their L1 does not encode case marking on nouns and only allows limited flexibility in word order (Jezek 2003; 2011; 2016). Nonetheless, such limited but systematic variability may suggest that Italian speakers are used to identifying the subject of the verb independently of its position in the sentence: effectively, MacWhinney et al. (1984) show that in this language the subject is identified most univocally not by word order, but by subject-verb agreement, although this observation is not directly relevant here as both nouns in the test target sentences could agree with the verb. Nevertheless, it may be hypothesised that thanks to their L1, Italian speakers are facilitated in the analysis of inflectional morphology, an ability (not a structure!) which they might have transferred from the L1, where some OS constructions are indeed possible.

7.5 Learner semi-spontaneous production

Chapter 7 presented a qualitative and quantitative analysis of learner output in the context of a semi-spontaneous production task, in which the target structure of this book can be examined under radically different conditions compared to the structured tasks considered in the previous chapters.

7 Discussion

The qualitative analysis highlighted that unlike the target sentences of the structured tasks, but in full accordance with the input, subjects are mainly instantiated by pronouns or person names. Object case marking is subject to much variability, ranging from virtually target-like performance to a systematic overextension of the [a] ending. In any case, only SO targets were produced. Particular attention was paid to the output of those participants who repeated the same lexical item more than once, as it was deemed helpful to determine whether a systematic principle applied, or whether on the contrary random variability occurred in the data. The former possibility proved clearly predominant.

The quantitative analysis consisted in the computation of a binary score indicating whether or not the learner considered may be thought to have applied a morphosyntactic principle when producing output. This score was then compared to its equivalent measure computed in Chapter 6, which indicates whether or not the morphosyntactic principle can be thought to have been relied upon in the structured tests. The results show that performance is systematically lower in the semi-spontaneous production task than in the structured tasks, although the target sentences produced were arguably easier from a cognitive point of view, since no OS construction appears in the data.

The ability to manage morphosyntax in an interactional context can thus be seen as the last step of the hierarchy identified for the structured tests:

SO production ⊃ OS repetition ⊃ OS comprehension ⊃ SO repetition ⊃ SO comprehension.

Since according to the "meaning first" principle the successful expression of referential meaning has priority over grammatical accuracy, it seems reasonable that learners should accurately case-mark nouns only once that process has become rather automatized in their interlanguage, so as not to subtract resources to the management of meaningful communication. In untimed structured tests, on the contrary, there is no "real" communicative situation to focus on, so that learners may devote all available resources to the linguistic task at hand.

Such results are fully compatible with the Limited Attentional Capacity Model proposed by Skehan & 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. Even though the structures produced in spontaneous speech are less complex than those encountered in the structured tests (no OS utterances were required or produced), learners engaging in an interactional production task have to face a number of difficulties which are extraneous to structured tests. First, in addition to retrieving lexical items and producing them in the appropriate word-form, they need to keep track of the discourse situation and adapt their output

to the interlocutor in real time. Further, the expression of meaning is functional to achieving a concrete objective, which, although not necessarily important to the learners, still represents the true objective of the task.

It is impossible to tell whether the absence of OS structures reflects a lack of morphosyntactic skills or simply the fact that they were not needed in the communicative situation considered. Indeed, OS structures typically serve the purpose of topicalising the object in order to create contrastive emphasis, an effect which there was no need or opportunity to produce. Therefore, it is impossible to directly test Robinson's (2001; 2005) hypothesis that more complex tasks tend to elicit more complex and accurate language.

Based on the data presented in this chapter, the morphological variability of individual lexemes may be investigated, too. This is a problem concerning the basic word-form in learner varieties, i.e. the word-form which is typically overextended to all contexts by those learners who are not yet ready to functionally manipulate inflectional morphology. Klein & Perdue (1997: 311) observe that in the Basic Variety "lexical items typically occur in one invariant form. It corresponds to the stem, the infinitive or the nominative in the target language; but it can also be a form which would be an inflected form in the target language". Later on, however, these apparently random phonological forms become the basis for the development of systematic morphological contrasts, in which a difference in form corresponds to a difference in meaning. Indeed, the VILLA production data discussed by Bernini (2018b) and Dimroth (2018: 28–33) provide some evidence regarding the beginning of this process. Unfortunately, not much can be contributed based on the production data discussed in this book. Even when they are repeated several times in the speech of the same participant, lexical items in the object function tend to occur in the same word-form. When that is the nominative case, it is quite clear that the lexical item simply occurs in its basic word-form. If the [e] form is consistently produced, on the other hand, one still cannot tell if a morphosyntactic principle is being applied, or if on the contrary that lexical item, for that learner, simply presents a basic word-form in [e]. In other words, the presence of the marked form is not sufficient to postulate the existence of a productive opposition: in order to do that, ideally, the same lexical item should appear in various word-forms, and the same inflectional morpheme should be applied to several lexical items (see Pienemann 1998 and Pallotti 2007 for a discussion of the *emergence criterion*). Otherwise, the possibility cannot be excluded that a specific word simply occurs in its basic word-form, which in principle can be modelled on any form of the paradigm, including the marked term of an opposition. Even though the VILLA project regrettably does not contain enough semi-spontaneous production data to proceed along these lines, a

clear tendency emerges from all the tasks considered in this study, whereby the overextension of the NOM ending [a] is the almost exclusive source of errors.

Finally, the comparison of learner output in the two different contexts of semi-spontaneous production and the structured tests raises the question which output should be taken as representative of the learner's morphosyntactic processing skills. To answer, one should first consider that the production task involves the *production* of a message, the structured tests only require its comprehension and repetition. From a cognitive point of view they thus appear intrinsically less complex, although the target sentences may be more demanding, as indeed was the case.

Secondly, at least some learners proved able to accurately produce inflectional morphology in the structured tests, but only under very specific conditions, in which no time pressure or interactional patterns were present. In this perspective, it seems that the real question is whether or not initial learners can produce accurate morphology in a real (or realistic) communicative situation. Based on the data presented in this chapter, it seems that the answer is that some can, although they will only rely on a selection of the structures which they could master in the structured tests, at least in principle. Moreover, the presence of inflectional morphology does not change the basic utterance structure, which is fully comparable to that produced by learners who do not use inflectional morphology and exclusively rely on a positional principle. In this respect, it appears that in the production task inflectional morphology is something quite accessory to the expression of meaning, but which is required by the grammar of the target language and by the instructional context in which acquisition takes place. Even in those learners who appear to be able to master morphology, the real communicative burden depends on word order and lexical item retrieval.

7.6 Limitations of the study and future directions of research

To conclude this discussion, a few methodological limitations of the present study need to be discussed, which may contribute to further refine future research.

7.6.1 The EI task

The VILLA EI task has two major methodological problems, which increased the complexity of the analysis.

The first concerns the assessment of target comprehension. In a task like the EIT, learner errors may stem from both a comprehension and a production failure. A comprehension test, or, even better, a translation test applied to the test items of the EI test would provide precious information as to whether or not comprehension is target-like. The assessment of comprehension may also be performed through the distractor (e.g. a comprehension question), as indeed was done in previous studies (e.g. Erlam 2006). Timing the test could usefully bring the test closer to the context of time-constrained spontaneous speech. In fact, timed tasks are generally deemed more appropriate for accessing implicit competence (Ellis 2005). Such solutions seem helpful to make sure that learners aim at meaning and avoid focussing on form, in full accordance with the rationale of the test.

The second point regards the distractor, which may not have been be the most appropriate. Drawing a simple geometrical figure does not necessarily inhibit working memory, as required by the EIT rationale. In addition to individual psychometric variation (e.g. working memory span), the ability to repeat a relatively long sentence without processing it for meaning may be attributable to a different approach to the task. Some learners might have taken a longer time to draw the picture, allowing their WM to fade. In contrast, others might have tried to complete the distracting phase as quickly as possible, perhaps while mentally rehearsing the target. Indeed, the case of participants being able to repeat targets in the absence of comprehension clearly points to an insufficiency of the distractor.

A more appropriate distractor should replace the content of their WM with new material, for instance by asking learners to perform simple calculations, read a sentence, answer a question, count from one to ten and backwards, or indeed a comprehension question, as stated above. Only then could one be sure that repetition really involves the re-coding of previously comprehended meaning. At the same time, excessively long and complex distractors may compromise even the recall of semantic information, so that it is not easy to strike a perfect balance and determine the "ideal distractor".

7.6.2 Generalisability of results and communicative situation

The results presented in this chapter should be generalised with caution, as they were obtained through a tightly structured experiment which significantly diverges from realistic language use. Structured tests are designed to investigate specific aspects of a given target structure: in the present case, the learner's ability to identify and express agent and patient on the basis of inflectional morphology alone. This research question required the deliberate exclusion of other

sources of information, such as prosody, semantics and context, on which both L1 and L2 users would normally rely. More realistic interlanguage data can arguably be obtained through spontaneous production, as indeed is recommended in several theoretical approaches (see Krashen 1985; Perdue 1993; Pienemann 1998, to mention but a few). However, such procedure also greatly reduces the chances of encountering relatively rare structures such as the OS sentence under investigation here, which were deemed essential to investigate the role of inflectional morphology in the interlanguage.

When results obtained through spontaneous production and through structures tests are compared, significant performance differences often emerge. For instance, within the VILLA project Watorek et al. (2016) have shown that morphosyntactic accuracy is much poorer in a communicative Route Direction task than in two structured tests. The results of the structured tests thus only represent the very best performance which learners can achieve under ideal conditions, but should probably not be expected in a more realistic communicative situation.

7.6.3 A comprehensive view of morphosyntactic skills

The EIT and the comprehension tasks were not originally designed to be paired: their correlation was made necessary by the EIT technical faults referred to above. As a result, the two tests are not directly comparable, the main difference being that they use words belonging to different paradigms. While the nouns used in the EIT all belong to the paradigm of feminine nouns in -a, the same is not true of the comprehension test, in which *siostra* is indeed a feminine noun in –a, but *brat* is a masculine animate noun in non-palatalised consonant. As case endings frequently encode different meaning across paradigms, the accusative form of masculine nouns is identical to the NOM of feminine ones, a situation which is not encountered in the EIT. On the other hand, it must be said that this is not an unusual situation in natural languages, so much so that a sentence like *siostra woła brata* 'the sister calls the brother' is perfectly normal in Polish. At the same time, the contrast between words belonging to different paradigms may be easier to perceive. While the ACC case of masculine nouns is one syllable longer and requires a stress shift, as in NOM *polak* /'polak/ but ACC *polaka* /po'laka/, 'Pole', the contrast between [a] NOM and [e] ACC as found in the feminine paradigm involves no stress shift, but only opposes two vowels which are not too different from an articulatory and perceptual point of view (Sisinni et al. 2013).

7.6 Limitations of the study and future directions of research

Finally, it cannot be ignored that the comprehension task only includes two lexical items. The test target sentences did not occur in the input, which eliminates the risk of processing based on chunks. Nevertheless, greater lexical variability would be advisable in future work.

8 Conclusion

The present chapter summarises the results of the analysis presented so far and attempts to present a coherent, comprehensive picture of the initial acquisition of a morphosyntactic contrast.

The methodological heart of both the repetition and the comprehension test is the 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, only the former strategy will work with OS targets.

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. Even though the present study only considered a minimal sub-system of Polish grammar, comprising only two forms and two functions, the VILLA project is not a laboratory experiment exclusively targeting this point, and Polish is not an artificial language: the learner's 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.

The analysis of comprehension errors points to a very clear and not unexpected effect of word order, whereby OS targets are characterised by a much higher error rate, whereas SO targets exhibit a ceiling effect for virtually all learners. These findings are in line the predictions and observations of numerous theoretical frameworks, including Processability Theory (Pienemann 1998; Baten 2013; Artoni & Magnani 2015), Processing Instruction (Van Patten 1984; 1996; VanPatten et al. 2013) and many others (e.g. Kempe & MacWhinney 1998; Jackson 2007; Henry et al. 2009; Rankin 2014). There is clear processing advantage for SO word order, which is in accordance with its greater diffusion in the languages of the world and indeed in the VILLA participants' L1s.

The EI task and the semi-spontaneous production task also involve a production component, which makes it possible to observe the principles through which

learners attempt to express syntactic relations. While the nominative case in -/a/ is hardly ever repeated incorrectly, 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/. This appears to be the single, invariable "basic word form" (Perdue 1993) for those learners who do not inflect nouns as required by the target sentence. The choice of the ending -a suggests that it is based on NOM case, presumably because of the more favourable distribution of the latter in the input. Indeed, cases in which the invariable word form is based on other inflected forms have been reported in the literature (e.g. Broeder et al. 1993; Garðarsdóttir & Þorvaldsdóttir in preparation), but this question is beyond the scope of the present work.

Since the EI test did not include a comprehension or translation component, learner production left some questions unanswered. In order to achieve a clearer picture, the results of the EI task were confronted with those of the comprehension task, which makes it possible to identify a hierarchy of target structure difficulty:

OS repetition ⊃ OS comprehension ⊃ SO repetition ⊃ SO comprehension

While the two extreme points are quite incontrovertible, the hierarchy seems a little uncertain in its medial compartment, as OS comprehension and SO repetition are correctly processed by a roughly equal number of learners. In any case, the generally tendency shows a facilitatory effect for SO word order and comprehension as opposed to OS and repetition. The reasons behind the learners' preference for SO have been discussed above; regarding the EI task, it can be argued that it is more complex than the comprehension test in that it encompasses it, while at the same time exerting other demands on the learner. According to most of the literature available, the EI task requires learners not just to repeat a string of sounds, but to decode it (just like in the comprehension test) and then re-produce it, both operations being performed on the basis of the present stage of interlanguage development.

If one further considers the results of the semi-spontaneous production task, the target-like use of inflectional morphology is even less frequent than in OS repetition, which places this task at the left end of the hierarchy presented above. However, in the production task all transitive utterances have an SO structure. While nothing can thus be said as to the learners' potential ability to use OS structures in spontaneous production, it is also the case that some learners proved able to accurately use inflectional morphology in OS repetition, but failed to the same in SO structures in the production task. In this respect, the production task is certainly more complex than the EI task because it requires participants to concentrate not only on the form of the message, but also on its content and integration

in the interaction. It is not surprising that inflectional morphology should not be given priority (Klein 2002). Despite the absence of inflectional morphology in the speech of most participants, the meaning of utterances is usually retrievable through alternative means, such as semantic and positional principles. The former rely 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. This observation too is fully compatible with vast evidence on utterance structure in the early stages of acquisition.

The detailed analysis of the input reveals that OS targets only represent a small subset of all transitive structures, even though the input had been specifically manipulated in order to provide learners with sufficient evidence as to this target structure. Moreover, strong tendencies were found regarding the associations between syntactic functions (hence case endings) with animacy, the vast majority of subjects being instantiated by personal pronouns or person names and most objects being represented by inanimate nouns. Structures with referents not differing in animacy are rare or absent altogether. The targets of the structured tests thus required a certain degree of generalisation involving new lexemes and semantic classes. Due to their object-like semantics, for instance, lexical items like *matematyka*, 'maths' only occurred in the input in their accusative form, yet, learners often produced them in an invariable word-form modelled on the nominative case, just like all other lexical items belonging to the same inflectional class. This result adds a precious piece of information to the debate on 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). In the case of the VILLA input, the predominance of the ending [a] in learner output may be explained with reference to its strongest association to the meaning NOM compared to [e] ACC. In terms of construction learning, it thus seems that higher-level, more abstract constructions like the association between case endings and syntactic functions overcome more specific constructions, like that between a given referent and a specific, inflected word form. Alternatively, it cannot be excluded that the preference for [a] may be due to factors somewhat independent of the input, such as the fact that lexical items were always introduced using their citation form (the NOM in [a]).

Turning to the role of the L1, it was predicted that speakers of a morphologically complex language would 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 influence 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. It could be hypothesised 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. These findings lead to two interesting observations. Firstly, it highlights the importance of perception and perceptual prominence in the (perhaps apparent) processing of morphology (Gallimore & Tharp 1981; Peters 1985). Secondly, it raises stimulating doubts as to the nature as well as the validity of the EI task for our research purposes (Vinther 2002; Erlam 2006; Van Moere 2012).

8.1 A comprehensive view

It is now possible to finally outline a fully comprehensive picture of learner processing skills in the earliest hours of SLA. First, a few learners proved able to process inflectional morphology in a structured test after only a few hours of exposure to the input, probably facilitated in this by their L1 (or possibly by other additional languages, e.g. Latin). The amount of input required to reach such results in the structured test is variable, but a small group of participants was able to achieve target-like results by the first test time (9 hours). In contrast, the majority of participants consistently applied a positional principle throughout the experiment in both comprehension and production. All L1 English learners fall within this group, which suggests a clear L1 effect. In between these extreme scenarios, a variety of evolutional patterns may be observed. Results tend to become more target-like over time, which witness to the beneficial effect of further input, although no clear pattern could be identified.

The use of inflectional morphology is rarest in the production task, in which even learners who proved able to successfully process OS targets in comprehension and repetition switch back to a positional mode, in which nouns only exhibit an invariable ending. 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. It thus appears that while the structured tasks elicited the very best performance

which the learner was capable of under laboratory conditions, in which the productive use of the target structure may well emerge. The "actual" competence, i.e. what the learner can do under pressure in a real communicative situation, or otherwise what the speaker *needs* to master in order to be, if not correct, at least effective. Actual production is very likely to have a very different structure from laboratory production, in many respects reflecting that of spontaneous learner varieties.

The main results obtained reported in this book are not revolutionary or surprising *per se*. The greater difficulty of production compared to comprehension, differences in grammatical accuracy depending on the task and the existence of marked forms are all facts long acknowledged or at least suspected by both linguistics and language teachers. In this respect, the present work confirms existent observations and brings additional details or domains of application: to name but a few, the acquisition of Slavic languages as L2s have been poorly researched so far, and first exposure studies are only limited to a very short time-span under strictly laboratory conditions. The VILLA project attempted to apply the same rigorous rationale to a communicative situation to a certain extent comparable to existing language teaching practices.

What gives new value to the results presented in this book is precisely the thorough methodology through which they were collected. While the tendencies which emerged from the analysis were mostly known to SLA and language teaching research, the doubt remained that what appeared to be a property of the target structure or a shared acquisitional fact would in fact be due to factors beyond experimental control, among which chiefly the learner's previous exposure to the target language (as well as to other languages) and input varying in amount and quality. These factors of variability were either eliminated or experimentally controlled in the VILLA project, which makes it possible to focus on the actual acquisitional facts thanks to the reduced disturbance from extra-linguistic factors.

Input control is particularly essential in the debate between nativism and generativism as to its role in the shaping the interlanguage: with respect to the former, the results show that learners do not always conform to the patterns found in the input, but on the contrary are able to generalise them in an innovative way in order to create new structures, perhaps partly reproducing structures belonging to the L1. This is particularly evident in semi-spontaneous production, in which structures occur which are ungrammatical in Polish and as such never occurred in the input. This observation is indeed consistent with the learner variety approach, which shows that learners manipulate the building blocks of input (words and constructions) in a manner that is not always in line with the target

language, but that is largely shared cross-linguistically. Again, however, no input control was attempted in these studies, so that this crucial variable inevitably remained a possible source of explanation.

Rather than answer new questions, the present study made it possible to answer existing questions in a more rigorous and comprehensive manner. The wealth of data collected for each learner within the VILLA project describes a rich picture comprising a variety of factors which are not usually found together in a single experiment. Although the present work only used a subset of the data, it conclusions can be further refined or expanded in light of other thoroughly controlled variables. It is hoped that the present analysis made a useful contribution towards the identification of what really matters in SLA, by controlling some of the many variables impacting on each individual learning experience.

9 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).

9 Pronunciation guide

Orth.	IPA	notes	example	IPA
a	a		*ale* 'but'	/ale/
ą	ɔN[1]	before stops and affricates	*początek*	/pɔtʃɔntek/
ą	ɔw̃	before fricative and word-final	*mamą* 'mum[INS]'	/mamɔw̃/
b	b		*niebieski* 'blue'	/nʲebo/
c	ts		*co* 'what'	/tso/
ć	tɕ		*pić* 'to drink'	/pitɕ/
ch	x		*dach* 'roof'	/dax/
ci	tɕ		*ciągnie* 'pulls'	/tɕɔŋgnʲe/
cz	tʂ		*czarny* 'black'	/tʃarnɨ/
d	d		*dom* 'home'	/dom/
dz	dz		*bardzo* 'very'	/bardzo/
dź	dʑ		*dźwięk* 'sound'	/dʑvʲenk/
dż	dʒ		*drożdże* 'yeast'	/droʒdʒe/
dzi	dʑ		*gdzie* 'where'	/gdʑe/
e	ɛ		*we* 'in'	/ve/
ę	ɛN	before stops and affricates	*między* 'between'	/mʲendʑɨ/
ę	ɛw̃	before fricative	*męża* 'husband'	/mɛw̃ʒa/
ę	ɛ	word-final	*kawę* 'coffee'	/kave/
f	f		*flaga* 'flag'	/flaga/
g	g		*gra* 'game'	/gra/
h	x		*hotel* 'hotel'	/xotel/
i	i		*i* 'and'	/i/
j	j		*ja* 'i'	/ja/
l	l		*kolor* 'colour'	/kolor/
ł	w		*mały* 'small'	/mawɨ/
m	m		*mama* 'mum'	/mama/
n	n		*komin* 'fireplace'	/komin/
ń	ɲ		*dzień* 'day'	/dʑeɲ/
ni	ɲ		*niemiec* 'german'	/ɲemʲets/
o	ɔ		*to* ''this'	/to/
ó	u		*mówić* 'to say'	/muvitɕ/
p	p		*pokój* 'room'	/pokuj/
r	r		*rower* 'bike'	/rover/
rz	ʒ		*dobrze* 'good'	/dobʒe/
s	s		*jest* 'is'	/jest/
ś	ɕ		*śpi* 'sleeps'	/ɕpi/
si	ɕ		*siostra* 'sister'	/ɕostra/
sz	ʃ		*proszę* 'please'	/proʃe/
t	t		*tam* 'there'	/tam/
u	u		*tu* 'here'	/tu/
w	v		*w* 'in'	/v/
y	ɨ		*ty* 'you'	/tɨ/
ź	ʑ		*jeździ* 'goes'	/jeʑdʑi/
ż	ʒ		*mąż* 'husband'	/mɔnʃ/
zi	ʑ		*zielony* 'green'	/ʑelonɨ/

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'matɨka/; *chodziliśmy* 'we went' /xo'dʑiliɕmɨ/;

- Nasals and stops followed by pre-vocalic /i/ are palatalised to various degrees, e.g. *niebieski* 'blue' /nʲebʲeski/. The letter <i> in this case effectively functions as a diacritic.

References

Aldai, Gontzal & Søren Wichmann. 2018. Statistical observations on hierarchies of transitivity. *Folia Linguistica* 52(2). 249–281. https://doi.org/10.1515/flin-2018-0006.

Anderson, Richard C. 1972. How to construct achievement tests to assess comprehension. *Review of Educational Research* 42(2). 145–170. https://doi.org/10.3102/00346543042002145.

Armon-Lotem, Sharon & Natalia Meir. 2016. Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: Evidence from Russian and Hebrew: Accuracy of repetition tasks for identification of SLI in bilingual children. *International Journal of Language & Communication Disorders* 51(6). 715–731. https://doi.org/10.1111/1460-6984.12242.

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* (Eurosla Monograph Series), 177–193. Amsterdam: EuroSLA.

Baayen, Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

Bachman, Lyle F. & Adrian S. Palmer. 2009. *Language testing in practice: Designing and developing useful language tests* (Oxford Applied Linguistics). Oxford: Oxford University Press.

Baddeley, Alan. 1986. *Working memory.* Oxford: Oxford University Press.

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. 1992. The use of adverbials and natural order in the development of temporal expression. *IRAL - International Review of Applied Linguistics in Language Teaching* 30(4). 299–320.

Bardovi-Harlig, Kathleen. 2000. *Tense and aspect in second language acquisition: Form, meaning, and use.* Oxford: Wiley-Blackwell.

References

Baten, Kristof. 2011. Processability Theory and German case acquisition. *Language Learning* 61(2). 455–505.

Baten, Kristof. 2013. *The acquisition of the German case system by foreign language learners*. Amsterdam: John Benjamins Publishing Company.

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Bernini, Giuliano. 2003. The copula in learner Italian: Finiteness and verbal inflection. In Christine Dimroth & Marianne Starren (eds.), *Information structure and the dynamics of language acquisition*, 159–185. Amsterdam: John Benjamins Publishing.

Bernini, Giuliano. 2016. La réalisation phonique des mots en début d'apprentissage de langues secondes. *Revue Française de Linguistique Appliqueé* 21(2). 139–151.

Bernini, Giuliano. 2018a. Dalla fonetica alla fonologia e alla morfologia: La varietà iniziale di polacco L2 del progetto VILLA. In Marina Chini & Pierluigi Cuzzolin (eds.), *Tipologia, acquisizione, grammaticalizzazione [Typology, acquisition, grammaticalization]*, 205–218. Milano: Franco Angeli.

Bernini, Giuliano. 2018b. The sound pattern of initial learner varieties. *Linguistica e Filologia* 38. 85–110. https://doi.org/10.6092/LeF_38_p85..

Bernini, Giuliano. 2018c. The sound pattern of initial learner varieties. *Linguistica e Filologia* 38. 85–110.

Bettoni, Camilla & Bruno Di Biase. 2015. Processability Theory: Theoretical bases and universal schedules. In Camilla Bettoni & Bruno Di Biase (eds.), *Grammatical development in second languages: Exploring the boundaries of Processability Theory*, 19–80. Amsterdam: EuroSLA.

Bittner, Dagmar, Wolfgang Dressler & Marianne Kilani-Schoch (eds.). 2000. *First verbs: On the way to mini-paradigms*. Berlin: ZAS.

Bley-Vroman, Robert. 1983. The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning* 33(1). 1–17.

Bondaruk, Anna. 2013. *Copular clauses in English and Polish: Structure, derivation and interpretation*. Lublin: Wydawnictwo KUL.

Bordag, Denisa & Thomas Pechmann. 2008. Grammatical gender in translation. *Second Language Research* 24(2). 139–166. https://doi.org/10.1177/0267658307086299.

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.

Bresnan, Joan. 2001. *Lexical-function syntax*. Malden, MA: Blackwell.

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. Cambrdidge: Cambridge University Press.

Brown, Gillian & George Yule. 1983. *Discourse analysis* (Cambridge textbooks in linguistics). Cambridge ; New York: Cambridge University Press.

Brown, H. Douglas & Priyanvada Abeywickrama. 2010. *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.

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.

Buck, Gary. 2001. *Assessing listening* (Cambridge language assessment series). Cambridge: Cambridge University Press.

Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5). 425–455.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733.

Bybee, Joan. 2013. Usage-based theory and exemplar representations of constructions. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of construction grammar*, vol. 1, 49–69. Oxford: Oxford University Press.

Bybee, Joan & Sandra Thompson. 2000a. Three frequency effects in syntax. *Berkeley Linguistics Society* 23(1). 378–388.

Bybee, Joan & Sandra Thompson. 2000b. Three frequency effects in syntax. *Berkeley Linguistics Society* 23(1). 378–388.

Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins Publishing.

Bygate, Martin. 2001. Effects of task repetition on the structure and control of language. In Martin Bygate, Peter Skehan & Merrill Swain (eds.), *Task-based learning: Language teaching, learning and assessment*, 23–48. London: Longman.

Cadierno, Teresa & Soren Wind Eskildsen (eds.). 2015. *Usage-based perspectives on second language learning* (Applications of Cognitive Linguistics). Berlin: De Gruyter Mouton.

Carroll, Susanne. 2012a. First exposure learners make use of top-down lexical knowledge when learning words. In Kurt Braunmüller & Christoph Gabriel (eds.), *Hamburg studies on multilingualism*, vol. 13, 23–46. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/hsm.13.04car.

References

Carroll, Susanne. 2012b. 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. 2012c. When is input salient? An exploratory study of sentence location and word length effects on input processing. *Internation Review of Applied Linguistics* 50. 39–67.

Carroll, Susanne E. & Elizabeth N. Widjaja. 2013. Learning exponents of number on first exposure to an L2. *Second Language Research* 29(2). 201–229.

Carroll, Susanne Elizabeth. 2014. Processing 'words' in early-stage foreign language acquisition: A comparison of first exposure and low proficiency learners. In ZhaoHong Han & Rebekah Rast (eds.), *First exposure to a second language: Learners' initial input processing*, 107–138. Cambridge: Cambridge University Press.

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.

Chang, Winston. 2014. *extrafont: Tools for using fonts*.

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.

Chun, Christian W. 2006. An analysis of a language test for employment: The authenticity of the PhonePass Test. *Language Assessment Quarterly* 3(3). 295–306. https://doi.org/10.1207/s15434311laq0303_4.

Conklin, Kathy & Norbert Schmitt. 2008. Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics* 29(1). 72–89. https://doi.org/10.1093/applin/amm022.

Corder, S. P. 1967. The significance of learner's errors. *International Review of Applied Linguistics* 5(4). 161–170.

Crocco Galeas, Grazie. 1998. *The parameters of natural morphology*. Padova: Unipress.

DeKeyser, Robert. 2001. Automaticity and automatization. In Peter Robinson (ed.), *Cognition and second language instruction*, 125–151. Cambridge: Cambridge University Press.

Di Biase, Bruno & Camilla Bettoni (eds.). 2015. *Grammatical development in second languages: Exploring the boundaries of Processability Theory*. Amsterdam: EuroSLA.

Diehl, Erika, Sandra Leuenberger, Isabelle Pelvat & Thérèse Studer (eds.). 2000. *Grammatikunterricht, alles für der katz? Untersuchungen zum zweitsprachenerwerb Deutsch* (Reihe Germanistische Linguistik). Tübingen: Niemeyer.

Dietrich, Rainer, Wolfgang Klein & Colette Noyau. 1995. *The acquisition of temporality in a second language* (Studies in bilingualism). Amsterdam ; Philadelphia: J. Benjamins.

Dimroth, Christine. 2012. Videoclips zur Elizitation von Erzählungen: Methodische Überlegungen und einige Ergebnisse am Beispiel der "Finite Story". In Bernt Ahrenholz (ed.), *Einblicke in die Zweitspracherwerbsforschung und ihre methodischen Verfahren*, 77–98. Berlin: De Gruyter.

Dimroth, Christine. 2018. Beyond statistical learning: Communication principles and language internal factors shape grammar in child and adult beginners learning Polish through controlled exposure. *Language Learning* 68(4). 863–905. https://doi.org/10.1111/lang.12294.

Dimroth, Christine, Rebekah Rast, Marianne Starren & Marzena Watorek. 2013. Methods for studying the learning of a new language under controlled input conditions: The VILLA project. *EUROSLA Yearbook* 13. 109–138.

Doughty, Catherine J. S & Jessica Williams. 1998. *Focus on form in classroom second language acquisition.* Cambridge: Cambridge University Press.

Downey, Ryan, Hossein Farhady, Rebecca Present-Thomas, Masanori Suzuki & Alistair Van Moere. 2008. Evaluation of the usefulness of the versant for English test: A response. *Language Assessment Quarterly* 5(2). 160–167. https://doi.org/10.1080/15434300801934744.

Dressler, Wolfgang. 2011. The rise of complexity in inflectional morphology. *Poznan Studies in Contemporary Linguistics* 47(2). 159–176.

Dressler, Wolfgang & Annemarie Karpf. 1995. The theoretical relevance of pre- and proto- morphology in language acquisition. In Geert Booij & Jaap van Marie (eds.), *Yearbook of morphology 1994*, 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.

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.

References

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.

Dziubalska-Kołaczyk, Katarzyna. 1997. Pre- and proto- in Polish phonology and morphology and their interrelations. *Papers and Studies in Contrastive Linguistics* 33(9). 159–171.

Ellis, Nick. 2001. Memory for language. In Peter Robinson (ed.), *Cognition and second language instruction*, 33–68. Cambridge: Cambridge University Press.

Ellis, Nick. 2006a. Language acquisition as rational contingency learning. *Applied Linguistics* 27(1). 1–24. https://doi.org/https://doi.org/10.1093/applin/ami038.

Ellis, Nick. 2006b. 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 & 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 C. & Nuria Sagarra. 2010. 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. 2011. Learned attention in adult language acquisition: A replication and generalization study and meta-analysis. *Studies in Second Language Acquisition* 33(04). 589–624. https://doi.org/https://doi.org/10.1017/S0272263111000325.

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). 141–172.

Ellis, Rod, Shawn Loewen & Rosemary Erlam. 2006. Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition* 28(2). 339–368.

Erlam, Rosemary. 2006. Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics* 27(3). 464–491.

Fellows, Ian. 2014. *Wordcloud: Word Clouds. R package version 2.5.*

Ferrari, Stefania & Elena Nuzzo. 2009. Task per l'elicitazione di strutture opzionali in italiano L2. *Rassegna italiana di linguistica applicata* 1/2. 235–250. https://doi.org/10.1400/128064.

Gallimore, Ronald & Roland G. Tharp. 1981. The interpretation of elicited sentence imitation in a standardized context. *Language Learning* 31(2). 369–392.

Garðarsdóttir, María & Sigríður Þorvaldsdóttir. in preparation. *A processability approach to the acquisition of case in L2 Icelandic.*

Gathercole, Susan E. & Alan D. Baddeley. 2004. *Working memory and language.* Hove: Erlbaum.

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.

Giacalone Ramat, Anna. 2003. *Verso l'italiano: Percorsi e strategie di acquisizione.* Roma: Carocci.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure* (Cognitive theory of language and culture). Chicago: University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language.* Oxford: Oxford University Press.

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.* Modern Approaches to Language 25(2). 70–86.

Granger, Colette. 2004. *Silence in second language acquistion: A psychoanalytic reading.* Clevedon: Multilingual Matters.

Gries, Stefan Th. & Stefanie Wulff. 2005. Do foreign language learners also have constructions? *Annual Review of Cognitive Linguistics* 3. 182–200. https://doi.org/10.1075/arcl.3.10gri.

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.

Gussman, Edmund. 2007. *The phonology of Polish.* Oxford: Oxford University Press.

Håkansson, Gisela. 1989. The acquisition of negative placement in Swedish. *Studia Linguistica* 43(1). 47–58.

Hamann, Silke. 2002. Postalveolar fricatives in Slavic languages as retroflexes. In Sergio Baauw, Mike Huiskes & Maaike Schoorlemmer (eds.), *OTS yearbook 2002*, 105–127. Utrecht: Utrecht Institute of Linguistics.

References

Hamann, Silke. 2003. *The phonetics and phonology of retroflexes*. Utrecht: LOT.

Hamann, Silke. 2004. Retroflex fricatives in Slavic languages. *Journal of the International Phonetic Association* 34(01). 53–67. https://doi.org/10.1017/S0025100304001604.

Hamayan, Else, Joel Saegert & Paul Larudee. 1977. Elicited imitation in second language learners. *Language and Speech* 20(1). 86–97. https://doi.org/10.1177/002383097702000109.

Han, ZhaoHong & Stephen Peverly. 2007. Input processing: A study of ab initio learners with multilingual backgrounds. *International Journal of Multilingualism* 4(1). 17–37.

Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42(01). 25–70.

Hatch, Evelyn Marcussen & Anne Lazaraton. 1991. *The research manual: Design and statistics for applied linguistics*. Nachdr. Boston, MA: Heinle & Heinle.

Henry, Nicholas, Hillah Culman & Bill VanPatten. 2009. More on the effects of explicit information in instructed SLA: A partial replication and a response to Fernández (2008). *Studies in Second Language Acquisition* 31(4). 559–575.

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.

Hoffmann, Thomas & Graeme Trousdale (eds.). 2013. *The Oxford handbook of construction grammar*. Oxford: Oxford University Press.

Housen, Alex & Hannelore Simoens. 2016. Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in Second Language Acquisition* 38(2). 163–175. https://doi.org/10.1017/S0272263116000176.

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. 1997. Testing SLA Theory in the research laboratory. *Studies in Second Language Acquisition* 19(s2). 131–143.

Hulstijn, Jan. 2015. Discussion: How different can perspectives on L2 development be?: Perspectives on L2 development. *Language Learning* 65(1). 210–232.

Jacennik, Barbara & Matthew Dryer. 1992. Verb-subject order in Polish. In Doris Payne (ed.), *Pragmatics of word order flexibility*, 209–241. Amsterdam: John Benjamins.

Jackson, Carrie N. 2007. The use and non-use of semantic information, word order, and case markings during comprehension by L2 learners of German. *The Modern Language Journal* 91(3). 418–432.

Janssen, Dirk P, Balthasar Bickel & Fernando Zúñiga. 2006. Randomization tests in language typology. *Linguistic Typology* 10(3). https://doi.org/10.1515/LINGTY.2006.013.

Jassem, Wiktor. 2003. Polish. *Journal of the International Phonetic Association* 33(1). 103–107.

Jessop, Lorena, Wataru Suzuki & Yasuyo Tomita. 2007. Elicited imitation in second language acquisition research. *Canadian Modern Language Review* 64(1). 215–238.

Jezek, Elisabetta. 2003. *Classi di verbi tra semantica e sintassi.* Pisa: ETS.

Jezek, Elisabetta. 2011. *Lessico: Classi di parole, strutture, combinazioni.* Bologna: Il Mulino.

Jezek, Elisabetta. 2016. *The lexicon: An introduction.* Oxford: 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 & Brian MacWhinney. 1999. Processing of morphological and semantic cues in Russian and German. *Language and Cognitive Processes* 14(2). 129–171.

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.

Klein, Wolfgang. 1986. *Second language acquisition.* Cambridge: Cambridge University Press.

Klein, Wolfgang. 2002. Why case marking? In Ingrid Kaufman & Barbara Stiebels (eds.), *More than words: Festschrift for Dieter Wunderlich*, 251–273. Berlin: Akademie Verlag.

Klein, Wolfgang & Clive Perdue. 1992. *Utterance structure: Developing grammars again.* Amsterdam: John Benjamins.

Klein, Wolfgang & Clive Perdue. 1997. The basic variety (or: Couldn't natural languages be much simpler?) *Second Language Research* 13(4). 301–347.

Krashen, Stephen D. 1981. *Second language acquisition and second language learning.* Oxford: Pergamon Press.

References

Krashen, Stephen D. 1985. *The input hypothesis: Issues and implications.* London & New York: Longman.

Kruschke, John K. & Nathaniel J. Blair. 2000. Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review* 7(4). 636–645.

Landau, Ernestina, Mijo Lončarić, Damir Horga & Ivo Škarić. 1999. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet.* Cambridge: Cambridge University Press.

Lantolf, James & 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 Long. 1991. *An introduction to second language acquisition research.* London ; New York: Longman.

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.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levinson, Stephen C. 1983. *Pragmatics* (Cambridge textbooks in linguistics). Cambridge ; New York: Cambridge University Press.

Lewkowicz, Jo A. 2000. Authenticity in language testing: Some outstanding questions. *Language Testing* 17(1). 43–64. https : / / doi . org / 10 . 1177 / 026553220001700102.

Łuczyński, Edward. 2002. Fleksja języka polskiego z punktu widzenia ontogenezy mowy. *Biuletyn Polskiego Towarzystwa Językoznawczego* 58. 157–165.

Łuczyński, Edward. 2004. *Kategoria przypadka 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.

Luoma, Sari. 2004. *Assessing speaking* (Cambridge language assessment series). Cambridge: Cambridge University Press.

MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk.* III. Mahwah, NJ: Lawrence Erlbaum.

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.

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.

Marinis, Theodoros & Sharon Armon-Lotem. 2015. Sentence repetition. In Sharon Armon-Lotem, de Jan Jong & Natalia Meir (eds.), *Assessing multilingual children disentangling bilingualism from language impairment*, 95–122. Bristol: Multilingual Matters.

Mcdonough, Kim & Youjin Kim. 2009. Syntactic priming, type frequency, and EFL learners' production of why-questions. *The Modern Language Journal* 93(3). 386–398.

Meara, Paul. 2005. *LLAMA Language Aptitude Tests: The manual.* Swansea: University of Wales.

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.

Munnich, Edward, Suzanne Flynn & Gita Martohardjono. 1994. Elicited imitation and grammaticality judgment tasks: What they measure and how they relate to each other. In Elaine Tarone, Susan M. Gass & Andrew D. Cohen (eds.), *Research methodology in second-language acquisition* (Second language acquisition research), 227–245. Hillsdale, N.J: L. Erlbaum Associates.

Nyqvist, Eeva-Liisa. 2018. Mastering complex Swedish NPs: A comparison of non-immersion pupils and immersion L1 Finnish pupils. *Journal of the European Second Language Association* 2(1). 14. https://doi.org/10.22599/jesla.33.

Oksanen, Jari, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs & Helene Wagner. 2019. *vegan: Community ecology package.*

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, TX: Cognitive Science Society.

Onnis, Luca, Heidi R. Waterfall & Shimon Edelman. 2008. Learn locally, act globally: Learning language from variation set cues. *Cognition* 109(3). 423–430.

References

Padgett, Jaye & Marzena Żygis. 2007. The evolution of sibilants in Polish and Russian. *Journal of Slavic Linguistics* 15(2). 291–324.

Pallotti, Gabriele. 2007. An operational definition of the emergence criterion. *Applied Linguistics* 28(3). 361–382.

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 Schwartz & Harald Clahsen. 2006. On the L2 acquisition of the morphosyntax of German nominals. *Linguistics* 42(3). 669–705.

Pawley, Andrew & Frances Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and native-like fluency. In Jack Richards & Richard Schmidt (eds.), *Language and communication*, 191–225. London: Longman.

Perdue, Clive (ed.). 1993. *Adult language acquisition*. Cambridge: Cambridge University Press.

Perdue, Clive. 1996. Pre-basic varieties: The first stages of second language acquisition. *Toegepaste Taalwetenschap in Artikelen* 55. 135–149.

Perdue, Clive. 2002. Development of L2 functional use. In Vivian Cook (ed.), *Portraits of the L2 user*, 123–142. Clevedon: Multilingual Matters.

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, NJ: Lawrence Erlbaum.

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. *Language Learning* 65(1). 123–151.

Pienemann, Manfred, Jörg Keßler & Anke Lenzing. 2013. Developmentally moderated transfer and the role of the L2 in l3 acquisition. In Anna Flymann Mattsson & Catrin Norrby (eds.), *Language acquisition and use in multilingual contexts: Theory and practice*, 142–159. Lund: Travaux de L'Institute Linguistique de Lund.

Plonsky, Luke & YouJin Kim. 2016. Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics* 36. 73–97. https://doi.org/10.1017/S0267190516000015.

R Core Team. 2017. *R: A language and environment for statistical computing*. Vienna.

Radloff, Carla F. 1991. *Sentence repetition testing for studies of community bilingualism* (Summer Institute of Linguistics and the University of Texas at Arlington publications in linguistics). Arlington, TX: Summer Institute of Linguistics: University of Texas at Arlington; Obtained from International Academic Bookstore.

Rankin, Tom. 2014. Word order and case in the comprehension of L2 German by L1 English speakers. *EUROSLA Yearbook* 14. 201–224.

Rankin, Tom & Sharon Unsworth. 2016. Beyond poverty: Engaging with input in generative SLA. *Second Language Research* 32(4). 563–572.

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.), *Grammatica applicata: Apprendimento, patologie, insegnamento*, 111–124. Milano: AItLA.

Révész, Andrea, Marije Michel & Roger Gilabert. 2016. MEASURING COGNITIVE TASK DEMANDS USING DUAL-TASK METHODOLOGY, SUBJECTIVE SELF-RATINGS, AND EXPERT JUDGMENTS: A validation study. *Studies in Second Language Acquisition* 38(4). 703–737. https : / / doi . org / 10 . 1017 / S0272263115000339.

Robinson, P. 2001. Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics* 22(1). 27–57. https://doi.org/10.1093/applin/22.1.27.

Robinson, Peter. 2005. Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *IRAL - International Review of Applied Linguistics in Language Teaching* 43(1). 1–32. https://doi.org/10.1515/iral.2005.43.1.1.

Robinson, Peter. 2010. Implicit artificial grammar and incidental natural second language learning: How comparable are they? *Language Learning* 60. 245–263.

Robinson, Peter. 2015. The Cognition Hypothesis, second language task demands, and the SSARC model of pedagogic task sequencing. In Martin Bygate (ed.), *Task-based language teaching*, vol. 8, 87–122. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tblt.8.04rob.

Sachs, Jacqueline S. 1967. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics* 2(9). 437–442.

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.

References

Sagarra, Nuria & Nick Ellis. 2013. From seeing adverbs to seeing verbal morphology. *Studies in Second Language Acquisition* 35(02). 261–290. https://doi.org/https://doi.org/10.1017/S0272263112000885.

Sasayama, Shoko. 2016. Is a 'complex' task really complex? Validating the assumption of cognitive task complexity. *The Modern Language Journal* 100(1). 231–254. https://doi.org/10.1111/modl.12313.

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. 2015. 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. 2018. Correzioni tra pari in varietà iniziali di L2: Uno strumento appropriato? In Roberta Grassi (ed.), *Il trattamento dell'errore nella classe di italiano L2: Teorie e pratiche a confronto*, 97–110. Firenze: Franco Cesati.

Saturno, Jacopo & Marzena Watorek. 2020. The emergence of functional case marking in initial varieties of Polish L2. *LIA* 11(1). 29–64. https://doi.org/https://doi.org/10.1075/lia.00007.sat.

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.

Service, Elisabet. 1992. Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology Section A* 45(1). 21–50.

Service, Elisabet & Viljo Kohonen. 1995. Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics* 16(02). 155.

Sharwood-Smith, Michael. 1993. Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition* 15(2). 165–179.

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. Oxford: Oxford University Press.

Sisinni, Bianca, Mirko Grimaldi, Barbara Gili Fivela, Francesco Sigona & Andrea Calabrese. 2013. Percezione e produzione di vocali non native da parte di parlanti adulti. In Vincenzo Galatà (ed.), *Multimodalità e multilingualità: La sfida più avanzata della comunicazione orale*, 355–372. Roma: Bulzoni.

Skehan, Peter. 1998. *A cognitive approach to language learning*. Oxford: OUP Oxford.

Skehan, Peter. 2009. Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics* 30(4). 510–532. https://doi.org/10.1093/applin/amp047.

Skehan, Peter & Pauline Foster. 2001. Cognition and tasks. In Peter Robinson (ed.), *Cognition and second language instruction*, 183–205. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.009.

Skehan, Peter & Pauline Foster. 2007. Complexity, accuracy, fluency and lexis in task-based performance: A synthesis of the Ealing research. In Alex Housen, Folkert Kuiken, Ineke Vedder & Van Daele (eds.), *Language learning & language teaching*, vol. 32, 199–220. Amsterdam: John Benjamins Publishing Company.

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. Przyswajanie form deklinacji rzeczownikowej przez dzieci w wieku młodszym. *Psychologia Wychowawcza* 5. 515–524.

Smoczyńska, Magdalena. 1985. The acquisition of Polish. In Dan Slobin (ed.), *The crosslinguistic study of language acquisition.* Vol. 1, 595–686. Hillsdale, NJ: Lawrence Erlbaum.

Smoczyńska, Magdalena. 1997. Przyswajanie systemu gramatycznego przez dziecko. In Halina Mierzejewska & Maria Przybysz-Piwko (eds.), *Rozwój poznawczy i językowy dzieci z trudnościami w komunikacji werbalnej*, 41–52. Warszawa: DiG.

Speciale, Giovanna, Nick Ellis & Tracey Bywater. 2004. Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics* 25(02). 293–321.

Starren, Marianne. 2001. *The second time: The acquisition of temporality in dutch and french as a second language*. Utrecht: LOT.

Swain, Merrill. 1985. Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In Susan M. Gass & Carolyn G. Madden (eds.), *Input in second language acquisition*, 235–253. London: Newbury House Publishers.

References

Tannen, Deborah. 2007. *Talking voices: Repetition, dialogue, and imagery in conversational discourse.* 2nd edn. (Studies in interactional sociolinguistics). Cambridge ; New York: Cambridge University Press.

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. *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard Univeristy Press.

Tyler, Andrea & Lourdes Ortega. 2016. Usage-based approaches to language and language learning: An introduction to the special issue. *Language and Cognition* 8(3). 335–345. https://doi.org/10.1017/langcog.2016.15.

Underhill, Nic. 1987. *Testing spoken language: A handbook of oral testing techniques* (Cambridge handbooks for language teachers). Cambridge ; New York: Cambridge University Press.

Van Moere, Alistair. 2012. A psycholinguistic approach to oral language assessment. *Language Testing* 29(3). 325–344.

Van Patten, Bill. 1984. Learners' comprehension of clitic object pronouns: More evidence for a word order strategy. *Hispanic Linguistics* 1. 57–67.

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, MA: Cascadilla Press.

VanPatten, Bill, Erin Collopy, Joseph E. Price, Stefanie Borst & Anthony Qualin. 2013. Explicit information, grammatical sensitivity, and the first-noun principle: A cross-linguistic study in processing instruction. *The Modern Language Journal* 97(2). 506–527.

Vinther, Thora. 2002. Elicited imitation: A brief overview. *International Journal of Applied Linguistics* 12(1). 54–73.

Watorek, Marzena, Marie Durand & Katarzyna Starosciak. 2016. L'impact de l'input et du type de tâche sur la production de la morphologie nominale en polonais par des apprenants francophones débutants. *Discours* 18.

Watorek, Marzena & Jacopo Saturno. 2016. *La variabilité inter- et intra-individuelle dans le cas des apprenants débutants du polonais.* Université de

Toulouse Jean Jaurès (FR): Journée d'Etude « La variabilité des parcours acquisitionnels en L2 ».

Watorek, Marzena, Marianne Starren, Rebekah Rast, Christine Dimroth & Giuliano Bernini (eds.). in prep. *VILLA: Field manual, results and future directions.*

Wells, John C. 1995. *Computer-coding the IPA: A proposed extension of SAMPA.*

Wells, John C. 1997. SAMPA computer readable phonetic alphabet. In Dafydd Gibbon, Roger K. Moore & Richard Winski (eds.), *Handbook of standards and resources for spoken language systems*, 684–691. Berlin: Walter de Gruyter.

Wichmann, Søren. 2015. Statistical observations on implicational (verb) hierarchies. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages*, 155–181. Berlin: De Gruyter Mouton.

Wichmann, Søren. 2016. Quantitative tests of implicational verb hierarchies. In Tarō Kageyama & Wesley Jacobsen (eds.), *Transitivity and valency alternations: Studies on Japanese and beyond*, 423–444. Berlin: De Gruyter.

Wickham, Hadley. 2017. *stringr: Simple, consistent wrappers for common string operations.*

Williams, John N. 2010. Initial incidental acquisition of word order regularities: Is it just sequence learning? *Language Learning* 60(S2). 221–244.

Wulff, Stefanie & Nick Ellis. 2018. Usage-based approaches to second language acquisition. In David Miller, Fatih Bayram, Jason Rothman & Ludovica Serratrice (eds.), *Bilingual cognition and language. The state of the science across its subfields.* 37–56. Amsterdam: John Benjamins.

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? *Language Learning* 65(1). 152–180.

Zipf, Georg. 1935. *The psycho-biology of language: An introduction to dynamic philology.* New York, NY: 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. Berlin: ZAS.

Żygis, Marzena & Silke Hamann. 2003. Perceptual and acoustic cues of Polish coronal fricatives. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 395–398.

References

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.