Teachability and Learnability across languages
Publisher: John Benjamins
Editors: Ragnar Arntzen, Gisela Håkansson, Arnstein Hjelde, Jörg-U Keßler (eds.)

Title: The Elicitation of Oral Language Production Data: An Exploration of the *Elicited Imitation Task*

Author: Kristof Baten

Affiliation: Ghent University / Research Foundation Flanders (FWO)

The Elicitation of Oral Language Production Data: An Exploration of the *Elicited Imitation Task*

Introduction

This chapter will discuss the so-called *Elicited Imitation Task* (EIT) as a possible useful task to collect oral language production data in Processability Theory (PT) research (Pienemann, 1998). In order to test the developmental hypotheses based on PT, language data is needed which relates to "the learners' capacity to utilise their interlanguage (IL) grammar under the time constraints of *spontaneous oral language production"* (Pienemann, 1998: xvi). The most important, and at the same time also the most difficult, aspect of collecting spontaneous oral language production data is of course eliciting enough relevant data. It is known to happen that certain tasks are not successful in eliciting enough relevant data. Pienemann and Mackey (1993), for example, found that a communicative task, such as the habitual action task, is more successful in eliciting third-person-s in L2 English than a free interview. Communicative tasks have an important advantage over tests of oral free production in that they allow for the elicitation of specific language structures and, what is more, achieve this in a relatively short timeframe. Using a range of communicative tasks within *Rapid Profile*, Keßler (2006, 2007) indeed reached high data density in only 12.5 minutes (speech samples ranged between 7 and 17 minutes). The development of suitable communicative tasks can be labour-intensive, however, and some structures may remain difficult to elicit.

In this chapter the so-called *Elicited Imitation Task* (EIT) will be put forward as a possibly useful technique for the elicitation of grammatical structures that are difficult to elicit. It will be argued that the EIT can elicit *any* structure in a short timeframe. A case in point with regard to German as a Second/Foreign Language, is the use of case markers in topicalized sentences (Baten, 2015).

(1) Den Sohn trägt der Vater
    The$_{acc}$ son carries the$_{nom}$ father

Sentences like (1) prove difficult to elicit. Importantly, if such sentences do not occur in the learner data, this does not necessarily mean that the learner is not able to use case in topicalized sentences. Rather, the data elicitation technique might very well have failed to elicit these kind of structures. The absence of relevant data has, of course, repercussions for the findings, because no conclusions can be drawn as to the emergence or non-emergence of a stage. The use of an EIT can avoid this problem. The present chapter will therefore present an EIT that was used to elicit case use in topicalized (and other) structures in L2 German. The next section will describe the procedure of the EIT in general (Section 2). Then, in Section 3, the structure under study will be presented. Section 4 will bring the two together and present the results of an empirical study using the EIT to elicit case markers in L2 German. Finally, Section 5 discusses the pros and cons of the EIT.

The *Elicited Imitation Task*: procedure

In an *Elicited Imitation Task* the participant hears a spoken stimulus sentence and then attempts to repeat it as exactly as possible. Between the stimulus and the response, the participant has to conduct a distraction exercise (see below). The stimulus sentences are either grammatical or ungrammatical. To illustrate the procedure, consider the following examples from a study on German negation and the German sentence bracket (Winkler, 2011, 531). The sentences in (2a) and (3a) are target-like. The sentences in (2b) and (3b) are not, because the negation particle in (2b) should appear after the finite verb and the past participle in (3b) should appear in sentence-final position.

(2)	a.	Der Fotograf		ist	nicht	in Deutschland
		the photographer		is	not	in Germany
	b.	*Die Oma	nicht	ist	im Supermarkt	
		the grandma	not	is	in.the supermarket	
(3)	a.	Der Opa	hat	ein Buch	gelesen	
		the grandpa	has	a book	read	
	b.	*Der schöne Mann		hat	geküsst	die Frau.
		The beautiful man		has	kissed	the woman

If a learner is able to reconstruct the deviant structures in (2b) and (3b) into target-like structures, following the model in (2a) and (3a), then it can be is assumed that that particular linguistic rule has been acquired.

The procedure of the EIT builds on the role of the working memory and its limited capacity for processing information (McLaughlin et al., 1983; Baddeley et al., 1998). In normal foreign language processing, meaningful items (i.e., content lexical items) are processed before less meaningful or non-meaningful items (i.e., grammatical morphemes) (Pienemann, 1998; VanPatten, 2004). The limited capacity of the working memory, so to speak, forces the foreign language learner to strategically allocate resources. In other words, regarding the EIT, when a learner hears the stimulus sentences, then s/he will process for meaning first, and only later for form. Furthermore, memory span research has demonstrated that the memory of the form (with respect to syntax, morphology and lexicon) quickly disappears after a sentence has been understood; the memory of meaning is retained longer (see Sachs (1967) and McDade et al. (1982)). This is the reason why in the EIT a short time-span between stimulus and response is installed by means of a distraction exercise. The implication is that after the distraction exercise the meaning can be retrieved from working memory, but not the linguistic form. Instead, the linguistic form (i.e., the syntactic structure and the morphemes used) needs to be reconstructed. For this reconstruction the learner will draw on her/his foreign language knowledge. As such, the reconstruction is assumed to grant insight in the learner's interlanguage grammar, in that a linguistic rule is probably not acquired when no successful reconstructions occur. On the other hand, the linguistic rule is probably acquired, when a learner successfully reconstructs a deviant structure into a target-like structure.

The use of the EIT appeals to a long research tradition, going back to the 1960s, and peaking in the 1970s and early 1980s. A number of reviews on the topic indicate that there is a consensus among researchers on the usefulness of applying the EIT. Nevertheless, after its auspicious start the use of the EIT declined, because a number of aspects related to its construct validity were unclear and because the field also began to embrace more communicative approaches (see, Bley-Vroman & Chaudron, 1994; Jessop et al., 2007; Vinther, 2002; Yan et al., 2016). It would go beyond the scope of this chapter to recount this debate here. Instead, two issues will be singled out: (i) the length of the stimulus sentence and (ii) the duration of the delay between stimulus and response. Both issues are related to the working memory span. The question is where to draw the line between rote repetition on the one hand and failure to repeat on the other hand. In other words, what is the appropriate sentence length and the adequate duration of delay in order to investigate the participants' language knowledge and not their working memory capacity. If sentences and/or the delay are too short, participants might very well just parrot the sentences literally, without any kind of reconstruction taking place. If sentences and/or the delay are too long, then it might occur that the participants are not able to repeat the sentence at all. With regard to (i) the length of the stimulus sentence, Vinther (2002) noted that sentence length in an EIT is often established on flimsy grounds, and that EIT studies usually pilot their own items due to the lack of guidelines on the appropriate sentence length. In Winkler (2011), for example, the appropriate sentence length was, after piloting, set at six to ten words. With regard to (ii) the length of the delay, McDade et al. (1982) observed that learners could no longer literally parrot a sentence after three to five seconds (see also Sachs, 1967).

As stated, the use of the EIT diminished in the 1980s and 1990s (with Bley-Vroman & Chaudron (1994) and Munnich et al. (1994) as notable exceptions), because a number of questions related to the EIT were not adequately answered. In recent years, however, the EIT has been experiencing a revival. In this respect, Graham et al. (2008) distinguish two lines of approach: the first approach relates to estimating an L2 learner's overall oral languages proficiency, whereas the second approach deals with the nature of the L2 competence itself. Empirical studies within the first approach examine whether EIT scores can statistically account for pre-existing differences among participants (usually in terms of institutional level differences) and whether these scores can be correlated with other measures of linguistic skills (e.g., fluency, vocabulary capacity, syntactic complexity, etc.).[1] These studies suggest that EITs offer a good estimate of L2 learners' global oral proficiency level.

The other line of approach questions whether the EIT can be used to reveal the reconstructive nature of the interlanguage and whether it taps into the *implicit* linguistic knowledge of the L2 learner. Within this approach a number of validation studies have recently provided empirical support for the EIT as a measure of implicit linguistic knowledge (Bowles, 2011; Ellis, 2005; Erlam, 2006; Spada et al. 2015). For example, in a correlation study, Erlam (2006) observed a strong positive correlation between the repetition of grammatical items and the correction of ungrammatical items, which indicates that the EIT is indeed reconstructive in nature. This finding counters earlier claims that the EIT merely represents rote repetition (see McDade et al, 1982).[2] Furthermore, it also suggests that in

[1] The validity of the EIT has been investigated with regard to a number of languages, e.g., Chaudron, Ngyuen & Prior (2005) for L2 Vietnamese, Graham et al. (2008) for L2 English, Tracy-Ventura, McManus, Norris & Ortega (2013) for L2 French, Wu & Ortega (2013) for L2 Chinese.

[2] It should be noted that the that original EITs used in the 1960s to 1980s most of the times only included grammatical items. In such design it is right to question the reconstructive nature of the EIT. As Erlam (2006) demonstrated, this consideration is no longer relevant when also ungrammatical items are included.

order for the L2 learner to be able to reconstruct an ungrammatical item into a grammatical one, the underlying linguistic rule must be part of his/her interlanguage. In other validation studies, Ellis (2005) (for L2 English) and Bowles (2011) (for L2 Spanish) used exploratory factor analyses. These studies revealed that the tests intended to measure implicit knowledge loaded on one factor while the tests intended to measure explicit knowledge loaded on another factor. Recently, Spada et al. (2015) generated a similar two-factor model with the EIT loading on one factor and an error correction task on the other. Interestingly, their study only examined one language feature (i.e., the English *be*-passive), whereas the other validation studies (Bowles, 2011; Ellis, 2005; Erlam, 2006) included no less than 17 features.

Further support for the validity of the EIT as a measure of implicit knowledge comes from Ellis (2008), which investigated whether EIT data can also be used to examine interlanguage *development* as defined by Processability Theory. Using EIT data of four L2 English grammatical features located at different stages in the PT hierarchy (category -, phrasal -, sentence -, and subordinate clause procedure), he found the same implicational scale that was found in PT research using naturally occurring data. In other words, the study by Ellis (2008) shows that the EIT, similar to spontaneous oral production tasks, measures the learners' implicit knowledge, i.e. the type of knowledge that is involved in interlanguage development and precisely the type of knowledge on which PT predictions are based. Therefore, in the same vein, the present chapter will investigate whether the EIT can also be used to uncover the developmental stages of L2 German case marking that were observed in free L2 speech data (Baten, 2013). The next section will describe these stages as well as zoom in on the language structure that proved difficult to elicit. Different from Ellis (2008), which investigated separate, independent grammatical features of different PT stages, the present study concentrates on one grammatical feature that locates at different stages in the PT hierarchy.

Methodological challenge: topicalization and functional case use in L2 German

Previous research on German case acquisition within PT has established that learners acquire the German case system in developmental stages (Baten, 2013). Broadly speaking, four stages have been observed. The first stage is a nominative stage. Learners only use nominative case markers and rely on other linguistic means (canonical word order, prepositions, etc.) to indicate grammatical functions. In the second stage, learners start to use accusative and/or dative case markers, but their use coincides with the canonical position of the arguments; deviations from the canonical positions lead to case errors. In the third stage, a distinction emerges between accusative and dative in prepositional phrases. Finally, the fourth stage shows target-like use of case markings in canonical as well as non-canonical sentences. At this stage, target-like case use in topicalized structures (as in (1)) occurs. In other words, case development follows the following path: nominative-only < positional marking < prepositional marking < functional marking (see table 1).

Table 1. Developmental sequence for L2 German case acquisition (Baten, 2013, 121)

STAGE	PROCEDURE	C-to-F-MAPPING	CASE
4	Sentence procedure	Non-linear (Topic = Object)	4) Functional marking
3	Phrasal procedure		3) Pcase
2	Category procedure	Linear (Topic = Subject)	2) One-to-one positional marking 1) No case marking (nominative only)

These stages also featured in many other studies that have dealt with the acquisition of German, but only the studies with findings relevant to the present paper will be discussed below.[3] First of all, the above-listed developmental stages are based on the emergence criterion (Pallotti, 2007; Håkansson, 2012). Other studies on the acquisition of German case markers have used similar criteria. For example, on the basis of the first and early occurrence of a few dative markers in prepositional phrases, Clahsen et al. (1994) and Czepluch (1996) argued that case is acquired sooner in prepositional phrases than in nominal phrases governed by the verb (i.e., the direct and indirect objects). In contrast, Eisenbeiß et al. (2005) and Schmitz (2006) claimed that case is actually acquired first in relation to objects instead of prepositions. Their claims were based on higher accuracy scores in nominal phrases compared to prepositional phrases. In PT research, however, accuracy scores are not considered as valid or reliable indicators of developmental stages (Pienemann, 1998: 135-138; Pallotti, 2007: 362-363). They rather reflect differences in mastery levels of a structure, which can be the results of all kinds of influencing factors (e.g., the complexity of the task, the level of fatigue of the learner, etc.). Therefore, the present study will adopt the emergence criterion (see below).

A second important finding relates to the variation that exists within developmental stages. For instance, with regard to the category procedure stage the learner can employ different structural options to indicate the grammatical function of the arguments of a sentence.[4] A sentence like 'she gives the boy the apple' could be expressed in the following ways. Only sentence (6) is target-like.

(4) *Sie gibt der Junge der Apfel
She gives the-NOM boy the-NOM apple

(5) ?Sie gibt den Apfel an den Jungen
She gives the-ACC apple to the-ACC boy

(6) Sie gibt dem Jungen den Apfel
She gives the-DAT boy the-ACC apple

[3] The interested reader is referred to the following research overviews: Eisenbeiß et al. (2005) for first language acquisition, Marx (2014) for second language acquisition, Baten and Lochtman (2014) for foreign language acquisition, and Motsch and Riehemann (2008) for speech therapy.

[4] It should be noted though that the functional marking is actually underspecified, because the three structural option in (2)-(4) are restricted to a fixed canonical order.

In (5) a prepositional construction is used to signal the dative object. In (4) only nominative markers are used, and as a result, the canonical position of the arguments reveals their grammatical function. The variation of structures at a particular developmental stage is in line with PT, because PT assumes a two-dimensional space with (i) developmental stages that are determined by increasing processing capacity and (ii) variational solutions for every stage that operates at the same level of processability (see 'hypothesis space' in Pienemann, 1998). The variation within a stage implies that morphological and syntactic features do not have to develop in tandem. It is possible, and actually very common, that learners have already acquired the syntactic structure of a certain stage, but without the morphological features that are located at that same stage.

In addition to syntactic variation, the development of case is also characterized by morphological variation in terms of gender. In previous studies on L1 acquisition (Szagun, 2004; Bittner, 2006) and L2 acquisition (Baten, 2013) it has been observed that the feminine article *die* 'the' is frequently overgeneralized. According to Bittner (2006, 127), learners overgeneralize *die* because of its high frequency in the input and because of its underspecification with respect to case features. The same might hold true for *das* 'the', but then again, *die* is also underspecified for number. Notwithstanding this variation, case development is to be kept separate from the gender issue. Following PT, "'gender' is an idiosyncratic diacritic feature of German nouns, the value of which has to be acquired individually for every lexical entry" (Pienemann, 1998: 159). The implication for the research on German case acquisition, in general, and the present chapter, in particular, is that conclusions regarding the developmental stages for German case cannot be drawn from the underspecified forms, but only from the specified, masculine forms.

Finally, a third and probably the most important finding for the present study is that both children and learners mainly seem to use 'easy' sentences, i.e., canonical sentences or sentences with prepositional objects. The L1 studies as well as the L2 studies have used an array of data elicitation techniques, e.g. diary notes, natural conversations, interviews, free or prompted writing tasks, translation tasks, and picture descriptions. However, none of these techniques seem to be very successful in eliciting the more difficult topicalized sentences (as in (1)) that are needed to determine unambiguously whether or not learners are capable of employing case markers to mark grammatical functions (Stage 4), instead of only argument positions (Stage 2). The absence of this kind of topicalized sentences does not allow for the conclusion that Stage 4 has not emerged, because the absence of proof does not equal proof of absence. Assuming that L2 learner can do more than what they have shown by the writing and speaking prompts that were used so far, research should find a more controlled manner to elicit these kind of topicalized sentences.

An attempt was undertaken in Baten (2013) and evidence of the above-listed Stage 4 could be provided, albeit only minimally. Baten (2013) referred to Tomlin's (1995) so-called *Fish Film*, which is a well-established task in SLA research. For example, PT research had adopted this film to elicit passive constructions (Keatinge & Keßler, 2009). The film shows two fish, each of a different colour, swimming toward each other. When they reach each other, one of the fish eats the other. The learner is asked to describe this action, but is linguistically constrained in his options due to a little arrow that accompanies one of the fish. The learner is instructed to start his utterance with the fish associated with the little arrow, meaning that the fish with the arrow is either eating or being eaten by the other fish. In other words, the arrow determines whether the fish is considered the agent or

the patient and, as such, whether the sentence reflects an active or a passive voice (e.g., *the green fish eats the red fish* vs. *the green fish is eaten by the red fish*). In Baten (2013) this method was adopted to elicit preverbal object. Learners were instructed to start their sentence with the object or person that was accompanied by the arrow. The study showed that the 'arrow method' elicited topicalized objects with both correct (7) and incorrect case use (8).

(7) Den fisch nehmen sie mit nach hause
The-acc fish take they with to home
'they take home the fish.'

(8) Die dritte hund kann ich nicht sehen
The-nom/acc.fem dog can I not see
'I can't see the third dog.'

However, these topicalizations only occur to a very minimal extent: only 71 of the 1823 objects appear in non-canonical position, which is not even 4%.

A similar method for the elicitation of topicalized objects was used in PT research on L2 Russian case acquisition (Artoni & Magnani, 2013). This study also found a development from marking the position (Stage 2) to marking the function (Stage 4). Evidence of Stage 4 was provided through a data elicitation task, called 'The Party'. The learners were told that a number of people were attending a party and that they would bring something with them. On a computer screen, the learners were shown two pictures, one showing a character, the other showing an object. The learners were instructed to tell who brings what, with the condition to start their sentence with the picture on the left. Because the picture on the left sometimes showed a person and sometimes an object, it was expected that learners would produce both canonical and non-canonical sentences (with topicalized accusative objects). The results indeed showed that this method elicited topicalizations among all participants.

Even though both methods proved successful in eliciting topicalizations, they have their shortcomings. The 'arrow method' in Baten (2013), for example, was not always successful in that it also elicited passive sentences, in which in accordance with the instructions the noun with the arrow came in sentence-initial position. Also, the learners appeared to overlook the arrow, or maybe they deliberately ignored it. The reason for this might be that a set of multiple pictures had to be described, some of which were 'arrowed' and some of which were not. Possibly the results might have been different if participants had been asked to describe the pictures without an arrow first, and only then the pictures with an arrow (i.e., as a separate task). This was the case in Artoni and Magnani (2013). Their 'Party' was a separate task (in fact, the fifth in a set of five tasks) and as a result, by that very fact, it could elicit topicalized objects among all learners. However, their method is limited to accusative objects. The reported Elicited Imitation Task in the present chapter will also include dative objects.

The Study

Participants. The study reported in this chapter involves data that was collected from 15 migrant learners living in Germany. Data collection took place at three time points (November; February; June) at an adult education centre in Bremen. The learners were enrolled in an integration course that included language classes of German as a Second Language at a beginner's level. The learners came from various L1 backgrounds and also had acquired various L2s during their life. Their experience with German ranged from 'very recent' to 'less than two years'. On a questionnaire, the learners indicated that they hardly speak German in everyday life, except at work (when they are employed) or when doing groceries. Some learners also indicated speaking German sometimes with non-German-speaking friends. In contrast, their receptive contact with German is more common. On the same questionnaire the learners indicated to watch German television or read German newspapers on a regular basis, ranging from once a day to once a week. Referring to the self-assessment grid of the CEFR (http://www.coe.int/en/web/portfolio/self-assessment-grid), most of the participants judged their own listening comprehension to be on level B1, and their speaking skills on level A2/B1. It should be noted that these levels do not allow for an evaluation of the participants' case knowledge, because the language-independent CEFR levels are not explicitly linked to particular aspects of grammar knowledge (and of vocabulary knowledge, for that matter) (Hulstijn et al., 2010). However, current research tries to identify the linguistic elements corresponding to the competence descriptors of the 6 CEFR levels in so-called *Reference Level Descriptions* (http://www.coe.int/t/dg4/linguistic/dnr_EN.asp; for L2 German, see Glaboniat et al., 2005).

Data elicitation task. The Elicited Imitation Task consisted of 40 stimulus sentences. Analogous to Winkler (2011), the length of these sentences ranged from five to nine words.

- 8 transitive and 8 ditransitive sentences with canonical word order to elicit Stage 2;
- 8 sentences with prepositional phrases to elicit Stage 3;
- 8 transitive and 8 ditransitive sentences with non-canonical word order to elicit Stage 4.

Half of the 40 sentences featured ungrammatical case use. In other words, in 20 sentences the learners had to reconstruct a case error into a target-like case marker. In the other 20 sentences, the learner only had to repeat the sentence. The difference between reconstructions and repetitions is important for the data analysis. In PT research, the emergence of stages is determined on the basis of creative language use. In the case of repetitions, however, it is unclear whether the language use really is creative. The repeated form might reflect the learner's developmental level, but it might also be mere parroting. In other words, repetitions do not necessarily show what the learner knows on the basis of his/her acquired L2 grammar. In contrast, when forms and structures are reconstructed, learners will draw on their L2 knowledge. So, reconstructions will grant us insight into the learner's interlanguage grammar. According to Munnich et al. (1994) the acceptance and rejection of grammatical violations (i.e., either repeating or reconstructing the ungrammatical case forms) are indeed a powerful indication of the learner's grammar knowledge.

In line with common practice, the EIT in this study also included a distraction between the stimulus sentence and the learner's response sentence. The distraction involved a picture-choice task. The

drawings used for this task depicted simple events, such as a man reading a letter, a train going through a tunnel, and so on. The concrete procedure was as follows and involved five steps:

I. Ten drawings lie on the table in front of the learner;
II. The learner hears a stimulus sentence (e.g., *Die Verkäuferin zeigt *DER Mann einen Pulli, 'the saleswoman shows the-NOM man a sweater'*);
III. The learner searches the drawing that corresponds to the sentence just heard and puts it aside;
IV. The participant repeats or reconstructs the sentence (e.g., *Die Verkäuferin zeigt DEM Mann einen Pulli, 'the saleswoman shows the-DAT man a sweater'*);
V. A new drawing is put on the table, so that again ten drawings are on the table. The procedure starts again from step I.

Results. Table 2 presents the results of the reconstructions. The columns show the three data collection points for the three case contexts. The maximum number of reconstructions that could be reached was eight in the positional and functional context and four in the prepositional contexts (i.e., 20 in total). Considering the 15 learners on the left and taking into account the missing data (indicated by '-'), the dataset consists of 780 items needing a reconstruction. Judging from the low numbers in this table, it is no surprise that the total number of correct reconstructions represents only 10%. The percentage of correct repetitions of grammatical items was slightly higher (31%), but as said, these mere repetitions will not be considered, because it is unclear whether or not they actually indicate the implicit knowledge on which PT's developmental hypotheses are based.

Table 2. Number of reconstructions in positional, prepositional and functional case use.

	POSITIONAL (8)			PREPOSITIONAL (4)			FUNCTIONAL (8)		
	T1	**T2**	**T3**	**T1**	**T2**	**T3**	**T1**	**T2**	**T3**
FRA	-	0	2	-	0	0	-	0	0
BIA	0	0	-	0	0	-	0	0	-
DIA	0	0	0	(1)	0	0	0	0	0
MON	1	4	-	0	0	-	0	0	-
SUG	0	0	0	0	0	2	0	0	0
NEL	1	2	1	0	2	2	0	0	0
MOH	0	0	0	1	0	1	0	0	0
AHM	0	0	-	1	0	-	0	0	-
PIO	0	0	-	1	1	-	0	0	-
GAB	0	0	0	1	2	1	0	0	0
KHA	0	1	0	2	3	3	0	0	0
DYA	-	3	1	-	1	2	-	0	0
PAR	1	1	3	1	1	2	0	0	0
TER	3	1	6	1	0	2	0	0	0
SAA	2	1	6	1	1	2	0	0	2

Let's now discuss the table according to data point. At Time 1, the number of reconstructions is very minimal (i.e., 6.5%). There are a few reconstructions in the context of positional and prepositional case marking, but no reconstructions in the contexts of functional case marking. Three learner profiles can be distinguished. First, three learners (SUG; BIA; DIA) do not make reconstructions as

they do not use case markers other than nominative. This means that their interlanguage does not show a case system, because they actually only use articles (which happen to be nominative forms).[5] Second, two other learners (MON; NEL) show the burgeoning of a case system, because they reconstruct an incorrect nominative-marked object into a correct accusative object. Even though this reconstruction occurs only once, it shows that these two learners start to make a distinction between nominative subjects and accusative objects (or, more in general, non-nominative objects). These two learner profiles are situated at Stage 2 in the PT hierarchy. The third learner profiles includes the other learners and coincides with Stage 3 in the PT hierarchy. These learners show the ability to make some reconstructions in prepositional phrases, and therefore are at the prepositional marking stage.

The same picture occurs at Time 2 and Time 3, but the number of reconstructions increases to 8% at Time 2 and to 17% at Time 3. More interesting, however, are the stage gains that take place. At Time 2, the learner NEL moves up a stage; now reaching the prepositional case marking stage. At Time 3, the learner SUG also moves to the prepositional case marking stage, and the learner SAA even moves to the functional case marking stage. These stage gains are important, because on a methodological level they demonstrate that the EIT is useful and valid to reveal inter-stage development.

In terms of intra-stage developments, the EIT also reveals the variation that occurs within a stage. If correct reconstructions occur only in slightly more than 10% of the items, then in about 90% the learners do something else. Before summing up these variable solutions, it should be noted that sometimes the learners were not even able to reproduce the stimulus sentences. In fact, in 10% of the items, they merely strung together some of the words they had heard.

With regard to the remaining 80%, the most frequent solution was to use the nominative masculine form on the target item (47%). This should not come as a surprise, because this was exactly the ungrammatical form they had heard in the stimulus sentence in the first place. In other words, this means that the learners just repeated the sentence without any attempt of reconstructing it. In some cases, the learners did attempt to reconstruct the incorrect masculine nominative form. However, the attempts proved unsuccessful, because the reconstructions involved the indistinct neuter or feminine forms (26%). Taking together these two solutions, the results clearly indicate that the learners reproduced the syntactic pattern they had heard (i.e., Stage 2: SVO; Stage 3: SV PP; Stage 4: OVS), but did not reconstruct the ungrammatical case form into the direction of a target-like case form.

In addition, in a minimal number of cases, the learners also reconstructed the syntactic pattern, and as such reveal the constraints of their interlanguage grammar. This applies to two contexts. First, in ditransitive sentences, the targeted dative object was reconstructed into a prepositional phrase (especially, the prepositions *an* 'to' and *vor* 'for' were used). For example, the ditransitive sentence *die verkäuferin zeigt *der mann einen pullover* 'the saleswoman shows the-NOM guy a sweater' was changed into *die verkäuferin zeigt eine pullover vor der mann* (instead of the targeted *die verkäuferin zeigt dem mann einen pullover*). This reconstruction shows that the learner is not able to use the positions of the objects relative to each other to signal the grammatical function, let alone to use the correct (positional) case makers. Instead s/he uses a preposition to indicate the argument's

[5] Diehl et al. (2000) referred to this stage as a one-case system, but because there is no opposition to other cases, the use of nominative *markers* only cannot really be considered as showing the existence of nominative *case*.

grammatical function. Previous research has found that L2 learners quite often make use of this strategy in free spoken and written language production (Baten, 2013; Diehl et al., 2000). However, in the present study, the use of a PP instead of a dative object only makes out 3% of the EIT data. In all likelihood, the participants in the present study use prepositional phrases more in free speech as well, but the syntactic structure of stimulus sentence will most certainly have primed their response.

Syntactic priming research has indeed shown that the syntactic structure of a prime sentence is echoed in the syntactic structure of a response sentence (for L2 German, see Scheepers & Corley (2000) and Loebell & Bock (2003)). In other words, the double object construction in the stimulus sentence primes its use in the response sentence. Analogously, the phenomenon of syntactic priming makes that the participants in the present study produce a high number of topicalized sentences. The OVS structure of the stimulus sentence primes the use of the same structure in the response. This is exactly the structure which previous studies so often failed to elicit. Nevertheless, in 3% of the EIT data, the syntactic pattern of the sentences with topicalized objects was not repeated. For example, the stimulus OVS sentence *DER Brief liest der Lehrer* 'the teacher is reading the letter' (remark again the incorrect nominative form on the object), was reconstructed into a canonical SVO sentence, *die lehrerin lesen die brief* (instead of the targeted *den brief liest der lehrer*). The reconstruction shows that the learner is not able to reproduce a Stage 4 syntactic structure, and as a result, the context for functional case marking is actually lacking.

Summing up, the EIT replicated the same developmental stages and revealed similar variation patterns as the ones that were found in previous research using free production data (see, Diehl et al., 2000; Baten, 2013). The EIT might therefore be considered as a useful and promising method to examine the development of learners' implicit knowledge (as was also shown for L2 English in Ellis (2008)). Clearly, one of the advantages of the EIT is that it can elicit structures that are difficult to elicit in free speech. However, the EIT also has a number of disadvantages, which brings us to the discussion of the pros and cons.

Pros and cons of the *Elicited Imitation Task*

Based on the results of this small-scale study a number of pros and cons can be derived. The major advantage of the EIT is that it can be used to elicit (i) *any* structure in (ii) a short timeframe. Regarding (i) the structure, the learners in this study were pushed to produce 'higher-staged' structures, which they would otherwise not produce. Recall that every learner had to produce 8 non-canonical sentences at each data point. Taking all learners and data points together, the EIT provided 312 non-canonical sentences, which had to be reconstructed. This number represents no less than 40% of the corpus. In 269 cases, the learners actually reproduced a non-canonical sentence. Only in 43 cases the learners failed to do so, in that they either produced a canonical sentence or an incomplete sentence (see discussion on variation above). Nevertheless, this number still represents 34% of the corpus. For the sake of comparison, in Baten (2013), the free picture description task only elicited a few non-canonical sentences, which only comprised 4% of the corpus. As a major methodological consequence, the EIT-based evidence of the (non-)emergence of the stage of functional case assignment is stronger. In Baten (2013) the majority of the participants did not use non-canonical sentences, which suggests that functional case marking was in all likelihood not acquired. However, the absence of proof is not necessarily proof of absence. The free picture

description task may simply not always have been successful in eliciting the relevant structures. In contrast, the EIT used in the present study made the learners to produce non-canonical sentences, and except for one learner (SAA), they failed to use case markers functionally. In other words, in the present study the non-use of functional case markers really is proof of the non-emergence of the functional case marking stage.

In terms of the effectiveness of data elicitation tasks to provide relevant L2 data of specific structures, the observed difference between the free picture description task and the elicited imitation task in this study is reminiscent of the findings in Pienemann and Mackey (1993). In this study with a group of child ESL learners it was found that the habitual-action task was far more successful to provide linguistic contexts for the use of third-person-s than a free interview. The same applied to the picture-difference task, which elicited far more questions than the free interview. Clearly, the methodological implication of these observations is that specific tasks should be designed to target specific linguistic structures, in order to avoid unproductive data collection sessions. In this respect, the findings in Pienemann and Mackey (1993) and in the present study demonstrate that it *is* possible to elicit specific structures in tasks that are designed appropriately.

The findings also throw an interesting light on L2 teaching and classroom interaction. The ability of learners to produce 'higher-staged' structures or particular linguistic contexts (e.g., thirds-person-s, questions, etc.) in specific tasks may also fail to appear in classroom discourse. Indeed, in a classroom study on question formation and negation in ESL, Keßler and Schwab (2015) observed that the majority of the learners was well beyond Stage 2, as shown by out-of-class elicitations tasks. None of them, however, produced structures from stages that are higher than stage 2 during the lessons that were recorded. One possible explanation, formulated by Keßler and Schwab (2015), was that interaction in the classroom was mainly teacher-led, leaving little opportunity for the learners to use questions or negated structures themselves. In the context of the present chapter, another possibility is that the tasks used in the classroom did not prompt the learners to use these kinds of structures. Extrapolating these observations to German as a Second Language (GSL), and more in particular, to the use of case markers, it can be assumed that in GSL-classrooms the use of case markers is often restricted to canonical sentences (i.e., Stage 2). Indeed, Thielmann (2007) found that example sentences in GSL textbooks were almost exclusively restricted to canonical word order. Also, in a recently suggested lesson plan, based on a Cognitive Grammar account of German case, the targeted use of German case markers is again limited to canonical sentences (Arnett & Jernigan, 2014). This might not be very useful, because the function of case markers does not really show in canonical sentences. Thielmann (2007), therefore, suggests to also provide examples of case use in topicalized constructions in the GSL classroom, as in (9).

(9) Ist die Küche denn schon komplett?
Nicht ganz. Einen Herd und einen Kühlschrank habe ich.

In the same vein, Fischer (2010) suggests to use journalistic texts in the GSL classroom, because this text type often contains a lot of topicalized objects. In addition to journalistic texts, the procedure of

an EIT could be integrated in order to show the free word order, typical for German, and the accompanying functionality of the case markers.[6]

Turning back to the advantages of the EIT, a second positive aspect was the relatively short timeframe, in which relevant data was collected. The execution of the EIT only took about ten minutes. By comparison, the free picture descriptions in Baten (2013) lasted between 20 and 25 minutes. Not only did data collection happen much faster, it also proved much less time-consuming to transcribe and annotate the 40 sentences of the EIT used in the present study, compared to transcribing and annotating the free speech in Baten (2013). In other words, the use of an EIT significantly reduces the labour in second language acquisition research. To reduce the labour even more, a possible next step would be to computerize the EIT which would yield the extra advantage that different learners can take the EIT in a classroom at the same time, instead of one learner after another (as in *Rapid Profile*).

Notwithstanding the advantages discussed so far, the use of the EIT is not without its setbacks. A number of disadvantages exist (although it should be noted that some of the disadvantages might be rather due to limitations of the present study than to the EIT in itself). A first problematic issue relates to the instruction that comes with the EIT, i.e. "repeat as exactly as possible". Foreign language learners might not expect native speakers or language teachers to make mistakes. Not expecting this, foreign language learners may in fact be "tricked" into making mistakes they might not make in free speech. What is interesting in this respect is that native speakers have been found to repeat ungrammatical sentences quite strictly (see Kaplan (1996), cited in Vinther (2002) and Yan et al. (2016)). In other words, they minutely followed the instructions to repeat what had been said ("Say as I say"). Likewise, in the present study, the repeated use of the ungrammatical nominative forms (in 47% of the cases) is actually in line with the instructions of the EIT. Unfortunately, it is difficult to find out whether the learners repeated the ungrammatical nominative case marker because they had been told to repeat what they had heard as exactly as possible or because they were not aware of the ungrammatical form. It can be assumed that the instructions influence how learners react and that differences in the instructional phrasing (e.g., "repeat as exactly as possible" vs. "use a correct sentence") may result in different responses. Clearly, more research on this issue is necessary.

In addition to the possible influence of the precise instructional phrasing other factors affect the reconstructive imitation process as well. The EIT does not only involve language production, but also builds on language comprehension ability, lexical knowledge and working memory (see, Jessop et al., 2007). Inadequate functioning in one of these domains will certainly have its effects on the learners' responses. To check for language comprehension and lexical knowledge the present study used the picture-choice task. If learners were able to match the stimulus sentence with the right picture, then it was assumed that there was sufficient lexical knowledge to understand the particular stimulus sentence. As described above, the picture-choice task also served the purpose of controlling the working memory, because the pause it created made that language competence and not working memory was tested. However, sometimes the duration of the pause was rather long, as it took time

[6] It goes beyond the present chapter to discuss how this can be done, but I refer to Winkler (2011) for an example of the use of an EIT in the GSL classroom, targeting German negation and the German sentence bracket.

to find the right picture. In some cases, the participants were even unable to point to the right picture, which may indicate a lack of lexical knowledge or comprehension difficulties. This situation does not invalidate the method of EIT in itself. It rather points to the practical difficulties that go together with task design which researchers should reflect upon when developing EITs (for an instrument providing guidelines for EIT development, see Tomita et al., 2009). Luckily, it is always possible to remedy disadvantages of this kind.

Conclusion

The aim of the present chapter was to contribute on a methodological level to data elicitation in second/foreign language research. More specifically, with reference to the acquisition of L2 German case marking, the chapter explored the possible usefulness of the EIT in PT research. Methodological limitations notwithstanding, the results show that the data obtained from the EIT is comparable to the spontaneous oral production data that is usually collected in PT research, because the EIT data replicated the same developmental sequence and the same pattern of variation that was observed in previous PT research on L2 German case. This finding suggests that the EIT can be utilized as a research tool in PT. Naturally, more meta-methodological research is needed to validate the task, especially with regard to the instructional phrasing and the intervening variables (comprehension, lexical knowledge and working memory). In addition to its usefulness for research, the chapter also hinted at the possible application of the EIT to language teaching. However, although language teaching might benefit from the psycholinguistic approach explored in this chapter, it is unlikely that language teachers (but also teacher trainers, educational advisors and educational policy makers) will look favourably towards PT and tests of morphosyntactic development. In current language teaching and didactics the focus lies on communicative competence. Linguistic competence is often completely ignored, and in the rare case that it is not, the emphasis lies on accuracy and not so much on the emergence of linguistic competence. Seeing that "what learners can do with language is to a very considerable extent dependent on what language they know" (Ellis, 2008, 18), foreign language teaching should try to combine the communicative approach with a developmental linguistic one. Therefore, future research should design EITs that include different grammatical structures from different PT levels (i.e., not only case), which then could be used as a task in the classroom, in addition to the other performance tasks that are used.

References

Arnett, C. & H. Jernigan (2014). A Cognitive Grammar Account of Case for L2 Students of German. *German as a Foreign Language 15*, 68-93.

Artoni, D. & M. Magnani (2013). LFG Contributions In Second Language Acquisition Research: The Development Of Case In L2 Russian. In M. Butt & T. Holloway King (eds.), *Proceedings of the LFG13 Conference* (pp. 69-89). CSLI Publications.

Baddeley, A., Gathercole, S. & C. Papagno (1998). The phonological loop as a language learning device. *Psychological Review 105*, 158–173.

Baten, K. (2013). *The Acquisition of the German Case System by Foreign Language Learners*. Amsterdam: John Benjamins.

Baten, K. (2015). Methodologische Herausforderungen der empirischen Fremdsprachenerwerbsforschung am Beispiel des Kasuserwerbs: Die „Elicited Imitation Task". *Zeitschrift für Fremdsprachenforschung 26,* 161-181.

Baten, K. & Lochtman, K. (2014). Das deutsche Kasussystem im Fremdsprachenerwerb: Ein Forschungsüberblick. *Muttersprache 124*, 1-25.

Bittner, D. (2006). Case Before Gender in the Acquisition of German. *Folia Linguistica: Acta Societatis Linguisticae Europaeae, 40*, 115–134.

Bley-Vroman, R. & C. Chaudron (1994) Elicited imitation as a measure of second-language competence. In E. Tarone, S. Gass & A. Cohen, *Research methodology in second-language acquisition* (pp. 245–61). Hillsdale: Lawrence Erlbaum.

Bowles, M. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition 33*, 247–271.

Chaudron, C., Ngyuen, H., & M. Prior (2005). Manual for the Vietnamese elicited imitation test. *National Foreign Language Resource Center, University of Hawai'i,* 1-33.

Clahsen, H., Eisenbeiß, S., & A. Vainikka. (1994). The Seeds of Structure: A Syntactic Analysis of the Acquisition of Case Marking. In T. Hoekstra & B.D. Schwartz (Eds.), *Language Acquisition Studies in Generative Grammar* (pp. 85–118). Amsterdam: John Benjamins.

Czepluch, H. (1996). Case Morphology and Case System in L1 acquisition: A Pilot Study of a German Child. In R. Sackmann, M. Budde, & A. Lieb-Doczy (Eds.), *Theoretical Linguistics and Grammatical Description* (pp. 89–108). Amsterdam: John Benjamins.

Diehl, E., Christen, H., Leuenberger, S., Pelvat, I., & Studer, T. (2000). *Grammatikunterricht: alles für der Katz? Untersuchungen zum Zweitsprachenerwerb Deutsch*. Tübingen: Niemeyer.

Eisenbeiß, S., Bartke, S., & H. Clahsen (2005). Structural and Lexical Case in Child German: Evidence from Language-Impaired and Typically Developing Children. *Language Acquisition: A Journal of Developmental Linguistics 13*, 3–32.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition 27*, 141–172.

Ellis, R. (2008). Investigating grammatical difficulty in second language learning: Implications for second language acquisition research and language testing. *International Journal of Applied Linguistics 18*, 4-22.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics 27*, 464-491.

Fischer, K. (2010). Kodiert das Deutsche funktional? Mythen, Fakten, Sprachvergleich und Deutsch als Fremdsprache. In K. Fischer, E. Fobbe, S. Schierholz (eds.), *Valenz und Deutsch als Fremdsprache* (pp. 13–40). Frankfurt am Main.

Glaboniat, M., Müller, M., Rusch, P., & Wertenschlag, L. (2005). *Profile Deutsch: gemeinsamer europäischer Referenzrahmen: Lernzielbestimmungen; Kannbeschreibungen; Kommunikative Mittel; Niveau A1-A2. B1-B2. C1-C2.* Berlin: Langenscheidt.

Graham, C. R., Lonsdale, D., Kennington, C., Johnson, A. & J. McGhee (2008). Elicited imitation as an oral proficiency measure with asr scoring. *Proceedings of LREC 2008,* Marrakech (Morocco). 1604–1610.

Håkansson G. (2012). The emergence criterion. In P. Robinson (ed.), *The Routledge Encyclopedia of Second Language Acquisition* (pp. 213-214). London/New York: Routledge.

Hulstijn, J.H., Alderson, J.C. & R. Schoonen (2010). Developmental stages in second-language acquisition and levels of second-language proficiency: Are there links between them? In I. Bartning, M. Martin & I. Vedder (eds.), *Communicative proficiency and linguistic development: intersections between SLA and language testing research* (pp. 11-20). Eurosla Monographs Series.

Jessop, L., Suzuki, W., & Tomita, Y. (2007). Elicited imitation in second language acquisition research. *Canadian Modern Language Review 64*, 215-238.

Kaplan, T.I. (1996). *Elicited imitation tasks in L2 acquisition research: Lessons from native speakers*. Unpublished manuscript, University of Iowa.

Keatinge, D. & J.-U. Keßler (2009). The Acquisition of the Passive Voice in L2-English: Perception and Production. In J.-U. Keßler & D. Keatinge (eds.), *Research in Second Language Acquisition: Empirical Evidence Across Languages* (pp. 67-92). Newcastle upon Tyne: Cambridge Scholars Press.

Keßler, J.-U. (2006). *Englischerwerb im Anfangsunterricht diagnostizieren. Linguistische Profilanalysen und der Übergang von der Primar- in die Sekundarstufe I*. Giessener Beiträge zur Fremdsprachendidaktik. Tübingen: Narr.

Keßler, J.-U. (2007). Assessing EFL-development online: A feasibility study of Rapid Profile. In: F. Mansouri (ed.), *Second Language Acquisition Research. Theory-Construction and Testing* (pp. 119-143). Newcastle upon Tyne: Cambridge Scholars Press.

Keßler, J.-U. & ,G. Schwab (2015). PT meets CA: Second Language Research at the Crossroads. In K. Baten, A. Buyl, K. Lochtman & M. Van Herreweghe (eds.), *Theoretical developments in Processability Theory* (pp. 175-204). Amsterdam: John Benjamins.

Loebell, H. & K. Bock (2003). Structural priming across languages. *Linguistics 41*, 791–824.

Marx, N. (2014). Kasuszuweisung und Kasuslehre bei Schülern mit Migrationshintergrund: eine differenzierte Betrachtung. In B. Ahrenholz & P. Grommes (eds.), *Zweitspracherwerb im Jugendalter* (pp. 99-124). Berlin: de Gruyter.

McDade, H., Simpson, M. & D. Lamb (1982). The Use of Elicited Imitation as a Measure of Expressive Grammar: A Question of Validity. *Journal of Speech and Hearing disorders 47*, 19-24.

McLaughlin, B., Rossman, T. & B. McLeod (1983). Second Language Learning: An Information-Processing Perspective. *Language Learning 33*, 135-158.

Motsch, H.-J. & S. Riehemann (2008). Grammatische Störungen mehrsprachiger Schüler. Interventionsstudie zum Therapieziel Kasus. *Die Sprachheilarbeit* 53, 15-25.

Munnich, E., S. Flynn & G. Martohardjono (1994). Elicited imitation and grammaticality judgment tasks; what they measure and how they relate to each other. In E. Tarone, S. Gass & A. Cohen (eds.), *Research methodology in second-language acquisition* (pp. 227–43). Hillsdale: Lawrence Erlbaum.

Pallotti, G. (2007). An Operational Definition of the Emergence Criterion. *Applied Linguistics 28*, 361–382.

Pienemann, M. (1998). *Language Processing and Second Language Development: Processability Theory.* Amsterdam: John Benjamins.

Pienemann, M. & A. Mackey (1993). An empirical study of children's ESL development and Rapid Profile. In P. McKay (ed.), *ESL Development. Language and Literacy in Schools* (Vol.2, pp. 115-259). Canberra: Commonwealth of Australia and National Languages and Literacy Institute of Australia.

Sachs, J. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics 2*, 437-442.

Scheepers, C. & M. Corley (2000). Syntactic priming in German sentence production. In L. Gleitman & A. Joshi (eds.), *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society* (pp. 435–440). Mahwah, NJ: Lawrence Erlbaum.

Schmitz, K. (2006). Indirect Objects and Dative Case in Monolingual German and Bilingual German/Romance Language Acquisition. In D. Hole, A. Meinunger, & W. Abraham (Eds.), *Datives and Other Cases: Between Argument Structure and Event Structure* (pp. 239–268). Amsterdam: John Benjamins.

Spada, N., Shiu, J. L. J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning* 65, 723-751.

Szagun, G. (2004). Learning by Ear: On the Acquisition of Case and Gender Marking by German-Speaking Children with Normal Hearing and with Cochlear Implants. *Journal of Child Language 31*, 1–30.

Thielmann, W. (2007). Fallstudie: Kasus in Sprachtheorie und Sprachvermittlung. *Zielsprache Deutsch, 34*, 11–34.

Tracy-Ventura, N., McManus, K., Norris, J. M., & L. Ortega (2013). "Repeat as much as you can": Elicited imitation as a measure of global proficiency in L2 French. In P. Leclercq, H. Hilton & A.

Edmonds (eds), *Measuring L2 Proficiency: Perspectives from SLA* (pp.143-166). Bristol: Multilingual Matters.

Tomita, Y., Suzuki, W., & Jessop, L. (2009). Elicited imitation: Toward valid procedures to measure implicit second language grammatical knowledge. *TESOL Quarterly 43*, 345-350.

Tomlin, R.S. (1995). Focal Attention, Voice, and Word Order: An Experimental, Cross-Linguistic Study. In P. Downing & M. Noonan (Eds.), *Word Order in Discourse* (pp. 517–554). Amsterdam: John Benjamins.

VanPatten, B. (2004). Input Processing in Second Language Acquisition. In B. VanPatten (ed.), *Processing Instruction. Theory, Research, and Commentary* (pp. 5-31). Mahwah, NJ: Erlbaum.

Vinther, T. (2002). Elicited imitation: A brief overview. *International Journal of Applied Linguistics 12*, 54–73.

Wu, S-L. & L. Ortega (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals 46*, 680-704.

Winkler, S. (2011). Sprachstandsermittlung und Lernfortschrittstest mal anders. Die "Elicited Imitation Task" als ein Fenster zur lernersprachlichen Grammatik. *Info DaF 5*, 225-250.

Yan, X., Maeda, Y., Lv, J. & A. Ginther (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing 33*, 497-528.