Production methods in language acquisition research

Sonja Eisenbeiss

1. Introduction

Collecting and analyzing samples of spoken speech from learners has played a central role in acquisition research since its beginnings. Initially, such speech samples were mostly collected in naturalistic settings, where researchers recorded learners in spontaneous interactions with family members, friends or researchers. Many of these naturalistic or spontaneous speech samples are now freely available, typically in written form, i.e. as transcripts; and they continue to be used by (psycho)linguists with a wide range of research interests.

Since the 1950s, acquisition researchers have supplemented naturalistic speech samples with production experiments. In these experiments, researchers systematically manipulate one or more variables and measure whether any changes with respect to these variables affect speakers' behavior.
Other researchers have developed semi-structured elicitation techniques: they keep the communicative situation as natural as possible, but use interviewing techniques, videos or games to encourage the production of rich and comparable speech samples. In the following, I will discuss the rationale of naturalistic speech sampling, production experiments and semi-structured elicitation and show for which phenomena and learner types they are typically used. I will also give a brief overview of data collection and analysis procedures and discuss the advantages and disadvantages of different production methods.

2. Naturalistic studies

2.1 Rationale

In naturalistic studies, researchers only interfere by audio/video-recording learners and other participants – sometimes without them even knowing that they are being recorded. Thus, this type of production study is least likely to be affected by the observer's paradox (Labov 1972: 113), i.e. the fact that the objects of research are susceptible to change because of the ongoing research process, in particular the presence of researchers, stimuli or recording equipment. Even when naturalistic samples are collected repeatedly, learners are not very likely to develop particular response strategies – if anything they become more used to being recorded and tend to forget about it. Hence, naturalistic studies have a high ecological validity as the recording situation closely approximates the real-life situation under investigation. Moreover, naturalistic sampling does not require prior in-depth linguistic knowledge to create stimuli; and naturalistic samples can in principle be obtained from any learner, independently of age, cognitive and linguistic ability. Finally, recordings where learners interact with their regular conversation partners also provide input samples.

This makes naturalistic studies an ideal starting point. They show which forms and constructions learners hear or produce themselves, and they show deviations from the target language that can help us investigate learners' acquisition mechanisms. For instance, production studies have demonstrated that children sometimes combine the regular past tense morpheme *-ed* with irregular verbs (**sing-ed*), which suggests that children do not simply imitate, but productively generalize inflections.

While minimizing researcher intervention enhances ecological validity, it can lead to incomparable samples, as learners may talk about different topics or differ with respect to vocabulary and levels of detail. Moreover, linguistic phenomena must occur quite frequently if researchers want to carry out

quantitative analyses of their acquisition. This is not always the case. For instance, in naturalistic samples from young German children, I found an average of twenty-eight contexts for articles or quantifiers per hundred utterances – but only an average of 0.2 contexts for the German possessive marker -s (*Annas Auto* 'Anna's car'; Eisenbeiss 2003). Similarly, noun phrases with adjectives are comparatively rare. One could pool data from several learners, but this can lead to sampling errors and ignores inter-learner variation.

The lack of researcher control can also hamper studies of fine-grained semantic distinctions, for instance distinctions between different manners of motion (*hop, walk, run*, etc.). Speakers might not talk about the semantic domain under study or they might focus on only one sub-domain. For instance, they might not mention motion at all or only refer to caused motion (*put*, *lay*, etc.). Moreover, investigating semantics requires access to detailed contextual information, but learners may choose to talk about non-present objects and events for which crucial contextual information is unavailable.

Finally, it is difficult to assess whether learners have acquired the forms or constructions under study. If these elements do not occur, this might simply result from a lack of appropriate contexts. For instance, speakers might not require information - and hence not ask questions. For some elements, one can identify "obligatory contexts", i.e. contexts where a competent native speaker would use this element (Brown 1973). For example, if speakers ask wh-questions, omitting the wh-element makes the questions ungrammatical (**are you doing*?). However, even conclusions based on omissions in obligatory contexts are limited as one cannot systematically vary variables to determine which factors cause such omissions: a lack of lexical or grammatical knowledge, processing or articulation problems, etc.

Neither can one simply take the occurrence of an element as evidence for its acquisition: many early child utterances in naturalistic samples consist of recurring word-forms and phrases that are also frequent in the input (e.g. *where's the doggie/car/cat*). Such early grammatical morphemes might be parts of (semi)formulaic patterns (Eisenbeiss 2000, Tomasello 2001, Radford 1990). Thus, one might overestimate learners' knowledge. Conversely, one might underestimate learners' knowledge when they are engaged in routine activities that are not linguistically challenging (bedtime rituals, meals, ...), or in activities like picture-book reading, which often only require imitations, object naming or elliptical answers.

To summarize, naturalistic studies are versatile, have high ecological validity and offer an excellent starting point for research on a broad range of phenomena, including learners' input. However, the lack of researcher control can lead to incomparable samples and make it difficult to study low-frequency phenomena, semantics and the productivity of learners' utterances.

2.2 Linguistic variables

Naturalistic speech samples are widely available: the CHILDES-database provides the largest collection of naturalistic corpora (<http://childes.psy.cmu.edu/>, MacWhinney & Snow 1990). The Max-Planck-Institute for Psycholinguistics in Nijmegen has digitized its L1/L2-corpora (<http://corpus1.mpi.nl/ds/imdi_browser/>). L2-Data are available from a number of web pages (<http://leo.meikai.ac.jp/~tono/lcresource.html>). Naturalistic studies cover a broad range of phonological, morphological, lexical, syntactic, semantic, and pragmatic phenomena; see the CHILDES bibliographies (<http://childes.psy.cmu.edu/bibs/>), Behrens (2008), Guasti (2002), MacWhinney (2000), MacWhinney & Snow (1990), Myles (2005), Sokolov & Snow (1994), Menn & Ratner (2000). See Vihman (1996) and Leather (1999) for phonology; Eisenbeiss (2009a), Chang, Dell & Bock (2006) for morpho-syntax; and Clark (1993), Blom (2007), Bloom (2000), Pinker (1989) for lexicon and semantics.

The fact that naturalistic data are ecologically valid and involve spontaneous informal conversations makes them ideal for calculating general measures for linguistic development (see below) and for studying discourse coherence and social conventions for speakers (Gallaway & Richards 1994, Pan & Snow 1999). Moreover, one can study bilingual learners' code-switching and their responses to other speakers' language use and choice in everyday multi-lingual situations (Wei & Moyer 2008). However, many naturalistic studies focus on routine activities that do not contain complex interactions and longer narratives. Thus, in studies of discourse and narrative development, naturalistic sampling is sometimes supplemented by elicitation (Berman & Slobin 1994).

When naturalistic samples are obtained from learners interacting with family members, caretakers, friends, etc., we can also investigate learners' input and the feedback they receive (Cameron-Faulkner, Lieven & Tomasello 2003, Chouinard & Clark 2003, Gallaway & Richards 1994, Küntay & Slobin 1996, Marcus 1993). Such naturalistic input studies have been supplemented by experiments that investigate the effectiveness of different types of input and feedback (see below). However, such studies cannot replace naturalistic studies that show whether a type of feedback that is effective in experiments is actually available to learners.

2.3 Subjects

Naturalistic samples can be obtained from any learners that are able to produce speech sounds, independently of their age, linguistic experience or linguistic, cognitive and physical abilities.

Participants of different ages will receive different levels of information: when learners cannot give informed consent due to their age or cognitive impairments, guardian consent will be obtained and one could conceal recording equipment. This minimizes observer effects, i.e. learners will be less likely to change their behavior or adapt to the researcher's language because of the ongoing research process. Note, that learners' behavior might still be affected if their interaction partners are aware of the recordings. Participants that can give informed consent should be informed about recordings. However, they might be willing to record themselves when no researcher is present or to allow researchers to conceal equipment and switch it on without further notice, which makes it easier to "forget" about recordings.

Recording bilingual speakers involves additional considerations (Quay 1995, Wei & Moyer 2008). Firstly, learners' own language use and their code-switching patterns will be affected by knowing that they are interacting with monolingual or bilingual researchers. Secondly, the presence of researchers might make participants self-conscious and lead to more self-monitoring and less code-switching. Thirdly, learners who are aware that researchers are interested in code-switching might start switching languages deliberately and produce "unnatural" types of code-switching.

2.4 Description of procedure

For all production methods, researchers have to make decisions about (i) the way in which learners are encouraged to speak, (ii) the recording situation and (iii) the number of participants and recordings. Here, we will discuss these decisions in general and for naturalistic studies. Specific considerations for semi-structured elicitation and experiments will be discussed in sections 3.4 and 4.4.

The aim of naturalistic studies is typically a representative and varied sample of learners' everyday speech. Hence, many naturalistic corpora involve a mixture of free play or other unstructured activities, meal-time conversations, and semi-structured activities that are not directly aimed at eliciting language (board games, bedtime rituals etc.; Menn & Ratner 2000).

In order to encourage learners to speak, some L1-researchers use multi-piece toys that allow for a range of engaging activities. While some argue that such toys encourage de-contextualized conversation about non-present events, others find that children actually talk less while playing with such toys (Eisenberg, Fersko & Lundgren 2001). In my experience, if tasks are challenging, but can be achieved by children alone (e.g. complicated puzzles), children are more likely to focus on the task than on talking. In contrast, multi-piece toys encourage speaking if learners need to verbally coordinate their actions with others.

With respect to the recording situation, researchers have to decide how they themselves will be involved. When the learner is only interacting with researchers, recording permission only has to be obtained for the learners themselves, not for other participants; and there is minimal interference with learners' private or professional lives. However, recordings with researchers do not provide representative input data and observer effects are stronger. In order to reduce such effects, one can give learners a chance to familiarize themselves with the researcher before the actual recordings. One can also exclude data from the first 5 minutes of each recording session when participants are settling into the recording situation.

Recordings can take place in learners' homes, in institutions they are familiar with (day-care centres, schools, work places, etc.) or in the researcher's institution. Home recordings are often preferable for naturalistic studies because learners are observed in their typical daily routines. However, it can be intrusive and requires researchers to travel. Recording in research settings reduces travel time for researchers, but puts learners into unfamiliar environments and requires them to travel themselves. Hence, it is typically only considered for adult learners that live closely to the researcher's institution. Recording in schools or workplaces is a good solution when several learners attend the same institution and institutional permissions can be obtained: learners are familiar with the setting; common interaction partners can be recorded as well, and the travel time is reduced for researchers while learners need not make additional trips.

While earlier naturalistic studies typically involved audio recordings, the availability of cheaper video-recorders has made video-recordings more common. Many web-pages provide up-to-date guidance (<http://www.talkbank.org/>, <http://www.mpi.nl/DOBES/help>, <http://www.mpi.nl/world/corpus/a4guides/>; <http://www.hrelp.org/languages/resources/>, and <http://www.camcorderinfo.com/>). Additional information can be found in Wei & Moyer (2008) and postings on the LINGUIST-list: <http://www.linguistlist.org/>.

Video-recordings provide additional situational information for understanding object, deictic and temporal references; and video-recordings are necessary for studies of multimodal interaction and links between speech, gestures, and actions. When learners or their guardians object to being video-taped, one can try to make this more acceptable for them: faces can be blurred, cameras can be positioned so that they do not capture the learner, but focus on the things learners talk about, etc. Moreover, learners can be given the right to cut out scenes. If they still object to video-recordings, note-taking and photographs of the recording site are helpful.

When speakers move around, one can use wireless broadcast microphones, as clip-ons for adults, or integrated into vests for child learners (<http://talkbank.org/da/fleck.html>). For audio-recordings, one can use digital recorders with tapes, minidisks or cards. If one has a laptop/PCs with a quiet fan and a good external microphone, one can use it to record with free audio-recording and editing software (e.g. Audacity: <http://audacity.sourceforge.net/>). Microphones must have high sampling rates. It is not advisable to use long-play options for video-recordings or to rely on MP3 or similar compression schemes for audio recordings as compressed formats are lower in quality. External hard-drives or DVDs provide cheap and portable storage.

In order to ensure good sound quality, one should minimize vibrations and outside noise by using microphone stands and camera tripods, avoiding locations with outside traffic or hard floors, and switching off televisions, washing machines, computer fans, etc. To reduce noise resulting from interference, researchers should use batteries instead of main power, keep cables short, use balanced-line connections for microphones, keep audio cables away from power cables, and use digital interconnections and optical cables whenever possible. Moreover, the recording quality should be checked at the recordings site before each individual recording, by recording short samples and listening to them. All tapes should be labeled clearly and uniquely, following a consistent format, and the relevant information should also be recorded in the video/audio-tape itself (<http://www.mpi.nl/world/corpus/a4guides/tapelabeling.pdf>). When recordings are stored in digital form, phonetic analysis tools like PRAAT can be used (Boersma & Weenik 2009). Moreover, segments can be played in a loop at varying speeds, which facilitates transcription. Finally, transcriptions can be carried out on digital copies, not the original tapes, which should be stored and backed up in separate locations.

It is advisable to keep logs of all digital files, with relevant metadata, i.e. information about participants, recording situation, etc. Such meta-data files should be stored in XML format, i.e. using the markup meta-language of the world wide web. This offers access to search tools that are independent of individual word processors or computer systems (<http://www.mpi.nl/tools/>). CHILDES-contributors also include minimal meta-data in the transcription/annotation files themselves.

With respect to the number of participants and recordings, naturalistic studies offer several options (Wei &Moyer 2008, Behrens 2008): one can conduct cross-sectional studies in which larger learner groups are recorded once or a few times within a short time interval. Such studies can be restricted to a particular population or provide a representative sample, stratified according to socio-economic status, age, sex, education. In both cases, we can study inter-individual variation. We can also get an initial idea of linguistic development by comparing L1 learners from different age ranges or L2 learners with different amounts of target-language exposure. Alternatively, one can look for implicational relationships: for instance, construction A might only be used by learners who also produce construction B; and there are learners who use construction B, but not A. This suggests that construction A is acquired later than construction B. However, cross-linguistic studies do not allow us to investigate the time-course of individual learners' development to gain insights into their acquisition mechanisms. This can be achieved in longitudinal studies where individual learners are recorded over longer time-periods, typically for 1-2 hours every week or month, but sometimes more frequently (Tomasello & Stahl 2004, Behrens 2008). Such studies usually only involve one or a few learners as they are time-consuming. Thus, they are not based on representative samples; and insights into inter-individual variation are limited. Therefore, some researchers combine longitudinal and cross-linguistic sampling; see e.g. Clahsen, Vainikka & Young-Scholten (1990) or Bol & Kuiken (1990).

2.5 Analysis and outcomes

Analyzing naturalistic data typically involves the transcription and annotation of the data; frequency, distributional and error analyses; and the assignment of recordings to developmental stages. When

spoken language is transcribed, i.e. rendered into written form, it is recommended that the recording is transcribed independently by 2 native speakers and that the reliability of the transcription is checked by comparing the 2 transcripts (see Wei & Moyer 2008 for discussion). However, researchers can decide whether they want to create full transcripts with all utterances of all speakers or only transcribe utterances that are relevant for the current study. One can transcribe phonetically, which is time-consuming, but provides more information than purely orthographic transcriptions. This is crucial for studies on phonology. For studies on other phenomena, most researchers make orthographic transcriptions and combine them with additional conventions for capturing deviations from the target. In addition, most researchers also include target forms and encode properties of spoken speech that can be relevant for its interpretation – pauses, hesitations, gestures, etc.

The most common transcription standard is the CHAT-format for the CHILDES-database (<http://childes.psy.cmu.edu/manuals/chat.pdf>). For adaptations of CHAT for multi-lingual data, see Wei & Moyer (2008). CHAT makes use of Unicode; so researchers can use their keyboard to represent different character sets and writing systems (Chinese, International Phonetic Alphabet, etc.) and work with special characters for discourse and conversation analysis.

Researchers can store their transcripts in separate text or XML files. Alternative, they can time-link parts of a transcript (utterance, conversational turns, etc.) to the corresponding segment of the recording, using the tools provided by CHILDES or the multi-media annotator ELAN (<http://www.lat-mpi.eu/tools/elan/>). Time-linking allows one to move quickly from one part of the media file to another by searching and selecting different transcribed segments. Thus, one can first transcribe only those utterances that are relevant for the current analysis and later transcribe others. Moreover, each speaker is assigned an individual transcription tier and dependent tiers for additional annotations (morpho-syntactic coding, etc.) and one can calculate overlap times, pauses between turns, etc.

For digital transcripts, tools such as the CLAN-tools for CHAT-transcripts can automatically create word lists and search for particular strings of characters or words (<http://childes.psy.cmu.edu/manuals/clan.pdf>). However, additional annotations are required for more complex automatic searches, for instance for a search for all obligatory article contexts. Thus, many researchers enrich their transcriptions by parts-of-speech tags, error codes, codes for grammatical relations such as subject, etc. The CLAN-tools allow researchers to semi-automatically add a tier with morphological and part-of-speech annotations. This reduces annotation times, though manual checks are recommended.

The (annotated) transcripts allow researchers to calculate frequencies. For instance, they might calculate how many different verbs (type frequency) and how many individual verb forms (token frequency) the learner produces. As the number of available contexts for a form can vary in naturalistic studies, frequency information has to be interpreted in relation to the number of contexts for this form. For some elements, obligatory contexts can be established, i.e. contexts where competent native speakers would use them. Then, researchers calculate how frequently the element is supplied in such obligatory contexts and provides percentages for target-like and non-target-like uses. Sometimes, it is not possible to determine obligatory contexts. For instance, one cannot say whether learners should have used a passive or an active construction. However, one can calculate how frequently these constructions appear per hundred utterances, so that one can compare recordings of different sizes.

When children deviate from the target language, one should distinguish between errors of omission (e.g. omitting inflections) and errors of commission (e.g. incorrect uses of inflections). While omissions might simply be due to processing limitations, errors of commission can provide information about learners' non-target-like generalizations – as the discussion about morphological overgeneralizations has shown. Moreover, one should investigate whether "errors" can be attributed to underspecifications of learners' grammatical representations. For instance, German children initially use feminine forms of articles in feminine context, but masculine forms in both masculine and neuter

contexts (Eisenbeiss 2003). This could be attributed to an initial underspecification for the feature GENDER, with a lack of the [+/-MASCULINE]-distinction.

For target-like forms, further distributional analyses are required. Firstly, one has to identify lexical gaps – as far as this is possible from recordings that do not cover everything a learner produces. For instance, a learner might omit the wh-word *why*, but use all other wh-elements when required. This suggests that the learner does not know *why*. Secondly, one can check whether linguistic elements are over- or under-extended. For instance, a learner might use articles in 80% of obligatory contexts, but never combine them with adjectives. Or a learner might use articles in all obligatory contexts, but also in contexts where they are inappropriate. This suggests non-target-like article representations. Thirdly, one can determine whether the respective linguistic element is used contrastively. For instance, a learner might correctly use singulars of demonstratives (e.g. *this, that*) in all and only in obligatory contexts, but there might not be any contexts for plurals (*these, those*). Then, no conclusions about the acquisition of number distinctions for demonstratives can be drawn. Fourthly, one can determine whether the linguistic element under study is combined with many different items or only co-occurs with a few lexical items – even though it could have occurred with others. Lexical restrictions suggest that learners have not yet acquired all generalizations for the respective element. In addition to frequency, error and distributional analyses, one can investigate developmental curves (Eisenbeiss 2000, Ellis & Larsen-Freeman 2006). For instance, English-speaking children initially produce correct regular and irregular past tense forms, then over-apply regular endings to irregulars (e.g. *go-ed*), and eventually performed adult-like (Marcus, Pinker, Ullman, Hollander, Rosen & Xu 1992). This U-shaped development in the percentage of correct forms can be interpreted as evidence for a shift from full-form storage for inflected word forms to morphological generalizations, but the nature of these generalizations is still debated.

Both distributional and developmental analyses can be employed to assign individual recordings to developmental "stages", which is useful for comparing different learners (see Eisenbeiss 2000 for discussion). For instance, one can distinguish between a stage in which the morpheme or construction under study is missing completely, a stage in which it appears, but is not used in all obligatory contexts, and a later stage, where it is used in all obligatory contexts and with a wide range of lexical elements. Alternatively, one can use U-curves to distinguish an early stage before a drop in performance, a stage of development during and immediately after the drop in performance and a stage of target-like performance. For cross-sectional data, one can investigate implicational relationships (see section 2.4)

Some researchers do not base their stage-assignment on the phenomenon itself, but use independent measurements of general linguistic development. These criteria can be independent of the researcher's data analysis: L1 researchers typically refer to children's chronological age and sometimes tests of verbal proficiency – especially for children with language impairments (see Paradis, this volume). L2 researchers employ proficiency tests or use information about age of arrival, length of L2 exposure, years of schooling, etc (see Unsworth & Blom, this volume).

Other general measures of linguistic development are based on the transcripts themselves, but do not focus on the phenomenon under study. In L1 research, the most common of these measures is the so-called MLU (mean length of utterance). It is calculated by dividing the total number of morphemes in a sample by the number of utterances (Brown 1973). This can be done automatically using CLAN-tools on annotated corpora. The use of the MLU is based on the idea that an increase in average sentence length reflects an increase in morpho-syntactic complexity. However, even long utterances might not be target-like and thus not provide evidence for advanced linguistic knowledge. Moreover, MLU-calculations involve difficult decisions about the segmentation of speech into utterances, the types of utterances that should be excluded (e.g. memorized songs) and the criteria for determining morphemes in typologically different languages (Brown 1973, Eisenberg et al. 2001, Jackson-Maldonado & Conboy 2007). Therefore, some researchers have used word-based instead of morpheme-based MLUs.

Moreover, MLU values above 3 or 4 do not seem to correlate well with other measures of grammatical development, which makes the MLU an inappropriate measure for older learners. Thus, other measures have been used in addition to the MLU, for instance, the maximal length of utterance, measures of vocabulary size and lexical diversity (Behrens 2008).

Calculations of such developmental measures – as well as error analyses - are typically only carried out for the learners themselves. In contrast, frequency and distribution analyses and some developmental analyses are often also conducted on input data: many studies calculate percentages of different types of feedback for learners and study the contexts in which these feedback types are used, or they look for correlations between input frequencies and age of acquisition (see section 2.2 for references).

3. Semi-structured elicitation

3.1. Rationale

In order to overcome the limitations of naturalistic studies, some researchers use semi-structured elicitation techniques or "games". These techniques encourage speech production and increase researcher's control of the recording situation, but keep the communicative situation as natural as possible (Eisenbeiss 2009b). Broad-spectrum techniques are used to obtain rich speech samples that allow researchers to compare speakers with different ages, and linguistic or cultural backgrounds. Some of these tasks involve stimuli that encourage participants to describe displays of events or objects. For instance, the Frog Story book (Mayer 1969) has been used with L1- and L2 learners, and speakers with language impairments (Berman & Slobin 1994). L2 researchers have employed Charlie Chaplin movies and other silent movies to elicit narratives (Stutterheim & Carroll 2006, Jucker 2008). Other tasks involve shared activities. For instance, the Bag Task involves a bag for blocks and animals of different sizes and colors (Eisenbeiss 2009b). The bag has pockets that match the animals in color and have colored buttons, ties, etc.; and children refer to colors, sizes and locations when they ask other players to help them hide or find animals in the pockets. For studies of the bilingual lexicon, children are sometimes given the same set of toys when they play with speakers of their two languages (Quay 1995). This makes them more likely to use translational equivalents of words and ensures the comparability of samples from the two languages.

Meaning-focused elicitation techniques target a particular semantic space that is encoded differently cross-linguistically, offering contexts for expressing fine-grained semantic distinctions and providing contextual information for the interpretation of speakers' productions. For instance, the "cut-and-break" video stimulus was created for cross-linguistic studies of "separation and material destruction" events (Bohnemeyer, Bowerman & Brown 2001). Naturalistic studies had found that children sometimes overextend the verb *open* to describe separating two Frisbees, etc., but richer data samples were required for more systematic studies. As the precise semantic distinctions for this domain in the target language were yet not known, no fixed set of variables was used to create controlled experiments. Rather, the videos simply showed events that differed with respect to properties that were assumed to be relevant for at least some languages.

Form-focused elicitation techniques target specific low-frequency forms or constructions. For instance, games with (pictures of) objects in different colors and sizes have been used to elicit noun phrases with attributive adjectives (Eisenbeiss 2009b); and sentence completion tasks have been used to elicit infinitival complements (Eisenberg 2005). As comparisons with naturalistic sampling have shown, such techniques do not lead to task-based errors or strategies, but produce richer and more varied data sets that can be reanalyzed for other phenomena (Eisenbeiss 2009b).

Thus, semi-structured elicitation techniques are useful for studies of low-frequency phenomena, fine-grained semantic distinctions or the productivity of learners' utterances. Moreover, elicitation techniques with flexible procedures and variable stimuli can also be used repeatedly as they typically do not lead to training effects. However, as semi-structured elicitation does not provide representative input or frequency data, it can only supplement, but not replace, naturalistic sampling. Neither do elicitation techniques provide the systematic control of materials and procedures that experiments involve. This makes them suited for exploratory studies, but reduces the power of (quantitative) analyses.

3.2 Linguistic variables

Semi-structured elicitation techniques can be used to study most phenomena that can be investigated in naturalistic studies, but they are particularly useful if one wants to study low-frequency phenomena or investigate learners' use of grammatical morphemes with a broad range of lexical elements. Moreover, semi-structured elicitation is ideal for explorative cross-linguistic studies of particular morpho-syntactic or semantic domains. Here, researchers often know too little about the properties of the languages to construct a controlled experiment, but want rich and comparable data sets. Elicitation studies are not appropriate for input or frequency studies: due to the elicitation task, the input samples and the frequency distribution of words, forms and constructions in the learners' production are not representative.

3.3 Subjects

Semi-structured elicitation can be used with L1 and L2 learners, with and without linguistic or cognitive impairments. As they involve higher task demands than naturalistic sampling, one cannot use elicitation techniques with new-born babies, but some tasks can be used with one or two-year old children, at an age when controlled production experiments are not possible (Eisenbeiss 2009b). However, stimuli and procedures must be adapted to the abilities of participants (see below).

3.4 Description of procedure

In contrast to naturalistic studies, where researchers interfere as little as possible, researchers in elicitation studies pro-actively create semi-structured contexts for the production of relevant linguistic data by providing a communicative function. In speaker/listener-tasks, speakers are asked to provide information for someone who does not have access to this information. For instance, speakers can be asked to relate life events or describe their home. Learners with limited linguistic or cognitive abilities produce little output in such tasks; and the lack of stimuli makes it more difficult to determine intended meanings. Therefore, many speaker/listener tasks involve verbal or visual stimuli (stories for retelling, pictures or videos for descriptions, etc.). Verbal stimuli can be more encouraging for unconfident learners, but if learners closely reproduce what they heard it is not clear whether their production is based on their own grammar. If one introduces delays between presentation and production and does not ask learners to imitate exactly, it is more likely that speakers' production is based on their own grammar - especially if learners use lexical items that had not appeared in the stimulus.

In director-matcher tasks, learners do not simply pass on information to passive listeners, but "direct" "matchers" in such a way that the matchers can actively find a particular stimulus in a set of stimuli or follow the directors' instructions. For instance, matchers have to follow route descriptions through toy landscapes, recreate spatial arrangements, etc. (Senft 2007).

In both speaker/listener- and director/matching tasks, it must be made clear to learners that the listeners/matchers do not have access to the relevant information. This can be achieved by blindfolds, physical barriers or seating arrangements that prevent listeners/matchers from seeing the stimuli. Alternatively, listener/matcher can pretend to be inattentive or they can be absent when the speaker/director receives the relevant information. However, younger children sometimes ignore the differences in knowledge between speakers/directors and listeners/matchers and it can prove difficult to prevent them from removing barriers, blindfolds, etc. Therefore, co-player tasks in which no "tricks" are used to create differences in knowledge are sometimes better suited for younger learner. In such tasks, players have to exchange information and coordinate actions to achieve goals, either in everyday activities such as cooking or in games.

Co-player tasks can be used as part of broad-spectrum techniques such as the Bag Task (see section 3.1 and Eisenbeiss 2009b), but such tasks can also be used to target particular meanings or forms. For instance, in the Puzzle Task, children ask for puzzle pieces with pictures on them, which they can then put into cut-outs of a puzzle board that show matching pictures (Eisenbeiss 2009b). The pictures differ minimally from each other, so that children must express the differences verbally in order to identify individual puzzle pieces. The differences between the individual pictures can be chosen so that children are forced to use specific forms or constructions. For instance, color and size differences can be used to elicit noun phrases with adjectives: *the big red balloon* vs. *the big blue balloon* vs. *the small red balloon*, etc. Alternatively, one can have contrasting pictures that focus on particular meanings and can use them to investigate how learners encode them. For instance one can depict possession transfer events and study whether learners use dative or prepositional constructions (*give someone something* vs. *give something to someone*).

Static stimuli, such as pictures or photographs, are easier to create and display than videos or animations. However, it is sometimes better to use dynamic stimuli as it can be difficult to reconstruct events from pictures or photographs - or even from a series of static displays that show processes and their results. This is particularly true if the focus is on the manner of motion (*hop* vs. *jump*, etc.). Using video stills instead of photographs or drawings makes events easier to recognize. In contrast, using conventions from comic strips, such as wiggly lines to indicate movement, is only possible if all learners are familiar with these conventions - which may be difficult in cross-cultural or L1-studies, where some participants might not (yet) have encountered pictures involving these conventions.

Both static and dynamic stimuli can be depictions (drawings, photos, videos, etc.) or involve toys or real objects (pots, bottles, etc.). Toys or real objects do not require any knowledge of artistic conventions, and are easier to obtain in fieldwork situations (Eisenbeiss 2006). Depictions can be naturalistic or more abstract (e.g. photos or videos vs. drawings or animations). Abstract depictions are ideal when colours, sizes or manners of motion are manipulated systematically or when unrealistic scenes are shown – e.g. animals acting in a "funny" way. Moreover, creating additional stimuli for follow-up studies tends to be easier for abstract stimuli as one does not have to worry whether actors or locations for photo/video-shootings might change or become unavailable. Finally, researchers should take into account adult learners sometimes find cartoon-like stimuli or toys too childish whereas children are usually happy to look at photographs or real objects. Thus, for L1/L2-comparisons, more realistic stimuli are preferred as they are acceptable for all participants. In order to ensure acceptability, one must also respect cultural taboos. This rules out pictures where body parts are shown that have to be covered in the culture of the participants – or pictures that show the consumption of foods that are taboo.

Whichever toys or stimuli are used, they should be non-toxic, waterproof (or laminated), easy to transport, easily replaceable if they are lost, and age-appropriate; and potential noise emissions from Velcro fasteners, metal components, etc. should be considered. Before each recording, researchers should ensure that all props are stable and fully functional. Picture or video stimuli can be found in the publications mentioned above, on the webpage of the International Picture-Naming Project

(<http://crl.ucsd.edu/~aszekely/ipnp/>) and in the field-work manuals of the Language and Cognition Group at the Max-Planck-Institute for Psycholinguistics <http://ww.mpi.nl/research/groups/LanguageAndCognition>).

With respect to the recording situation and the number of participants and recordings, semi-structured elicitation is similar to naturalistic sampling (see section 2.4). Due to the semi-naturalistic nature of the tasks, elicitation games can be used longitudinally when stimuli are sufficiently varied. The main difference to naturalistic studies is that learners in elicitation studies are typically recorded with researchers and not with their natural interaction partners. However, in longitudinal studies with children, one can train older family members to play games such as the Puzzle Task and regularly supply new stimulus pictures.

3.5 Analysis and outcomes

Due to their semi-naturalistic nature, semi-structured elicitation data can in principle be subjected to the same distributional, error and developmental analyses as naturalistic data (see section 2.5). However, in order to detect potential artifacts one should consider checking whether any systematic errors in the elicited data also occur in naturalistic data. One should also check for training effects such as a particularly frequent use of formulaic structures. Moreover, one has to acknowledge that elicitation does not provide representative data for MLU-calculations, frequency or input analyses as researchers have interfered with the recording situation.

4. Production experiments

4.1 Rationale

Researchers who want to systematically manipulate some variables and control for the effects of others favor experiments with standardized procedures and stimuli. The use of standardized procedures helps to avoid accidentally providing learners with linguistic models or feedback that might influence their behavior; and the use of stimuli allow us to determine learners' intentions and facilitates data interpretation.

In elicited imitation experiments, participants are asked to imitate spoken sentences (Lust, Flynn & Foley 1996, Bernstein Ratner 2000, Gallimore & Tharp 2006, Vinther 2002). When stimuli are sufficiently long and complex, participants cannot memorize them holistically, but have to employ their own grammar to recreate them. Hence, comparing target utterances and learners' actual productions might shed light on learners' grammatical knowledge. Early studies have indeed indicated that children modify targets in ways that closely resemble their own spontaneous use (Brown 1973). However, learners do not always imitate sentences they are known to have produced spontaneously (Bernstein Ratner 2000, Gallimore & Tharp 2006). Thus, non-target-like performance might not necessarily reflect non-target-like knowledge. Moreover, learners with particularly good memory and vocabulary knowledge might recall target sentences at least partially - and hence perform well even when they could not produce the target using their own grammar. Performance might also be affected by task-induced strategies. Given these interpretation problems, some researchers only use elicited imitation to gain a first glimpse of learners' production or to determine appropriate age/proficiency ranges. Others employ it in teaching or therapy (Gallimore & Tharp 2006).

Elicited production experiments do not involve explicit models for imitation, but only prompts to produce forms (Menn & Bernstein Ratner 2000, Thornton 1996, Crain & Thornton 1998). This can be one variant of the target and a context to produce another variant (*This is a door. These are two …?*). Alternatively, learners can be instructed to turn sentences into questions (*The dog is eating something,*

but I cannot see what. Can you as ask the puppet?), negated sentences *I'll say something and then you say the opposite*), etc. Both versions can be used to investigate whether learners can produce rare forms or construction when appropriate contexts are provided. Moreover, deviations from the target-form can be analysed to further investigate acquisition mechanisms (see section 4.5). Some experiments involve novel words (Menn & Ratner 2000, Theakston, Lieven & Tomasello. 2003). For instance, Berko (1958) showed children pictures of made-up creatures and asked them to produce novel inflected word forms (*This is a wug. These are two …*?). The rationale for using novel words is that learners do not have any stored representations of these words or formulas containing them. Thus, when learners actually produce novel words in the respective forms or constructions, this suggests they have the relevant grammatical knowledge. However, participants tend to do more poorly in elicitation experiments than in naturalistic studies. This might be (partially) due to task demands such as the need to remember novel words. Hence, one should be cautious with conclusions based on poor performance.

In speeded production experiments, learners are asked to produce morphologically complex forms as fast as possible (see Clahsen, Hadler & Weyerts 2004 for an overview). For instance, they might produce past tense forms like *walked* after hearing infinitives like *walk*. Researchers measure how long it takes learners to produce the complex form (e.g. *walk-ed*). This onset time for speech production indicates whether the respective form is stored holistically or computed by a morphological rule (Past:-*ed*): if they were stored in the mental lexicon, then high-frequency forms should have stronger memory traces, due to additional exposure. Therefore they should be retrieved and produced faster than low-frequency forms. In contrast, if morphological complex forms are computed from stems and affixes, the production latencies should only be affected by the frequency of these components (e.g. *walk* and –*ed*), not by the frequency of the complex form.

Another method from native-speaker processing research is syntactic priming (Branigan 2007). "Syntactic priming" refers to speakers' tendency to repeat syntactic structure across otherwise unrelated utterances. For instance, speakers are more likely to use passives after hearing or producing passives than after active sentences. This effect can be employed to explore learners' syntactic representations: learners are given a picture description as a prime and asked to describe another picture. Researchers then determine whether the presentation of the prime makes learners more likely to use the same construction. In adult native speakers, such effects occur even when primes and targets contain different lexical material. If learners show the same effect, this indicates that they have abstract target-like grammatical representations that can be activated by priming. In contrast, if priming only occurs when prime and target involve the same verb or other lexical material, this suggests that the grammatical representation that the prime pre-activates is not abstract, but rather lexically bound. Thus, syntactic priming experiments can be used to study L1/L2-learner's grammatical representations (Chang et al. 2006, Kim & McDonough 2008 and studies cited there).

Input/Feedback-experiments investigate which type of input or feedback affects learners' linguistic development. They thus supplement the naturalistic input studies discussed above. For instance, Saxton, Kulscar, Marshall & Rupra (1998), presented children with correct past tense forms of novel irregular words (e.g., *Look what happened! The spider pold the grasshopper*), i.e. positive input, but never corrective recasts in response to errors that they might make. These children performed worse than children who also received corrective reformulations (Child: *He pelled him*. Adult: *Yes, he POLD him*). Such studies have been carried out with L1 and L2 learners and populations with language impairments (McDonough & Mackey 2006, Proctor-Williams & Fey 2007).

4.2 Linguistic variables

Production experiments are used to systematically explore the role of phonological, semantic, morphological and syntactic linguistic variables as well as the role of input (Menn & Bernstein Ratner 2000). They are appropriate for low-frequency constructions and can provide insights into learners'

representations that other studies cannot offer: novel word experiments demonstrate productive language use; speeded production studies help us determine whether forms are decomposed or stored holistically; and syntactic priming effects tell us whether learners' grammatical representations are lexically bound or more abstract.

4.3 Subjects

Production experiments can be used with L1 and L2 learners, with and without linguistic or cognitive impairments. Which participants can take part in production experiments depends on the tasks involved, but most of these experiments are only conducted with learners from the age of 3. As with elicitation techniques, stimuli and procedures must be adapted to the linguistic and cognitive abilities of participants (see above). Similarly, for studies where input frequency is crucial for stimulus construction, frequency measures should be based on appropriate samples - e.g. the CELEX frequency database for studies with adult learners of English (Baayen, Piepenbrock & Van Rijn 1993), and child-directed speech from the CHILDES-corpora for studies with children. For some experiments, it may be helpful to select participants who are already familiar with the task. For example, elicited imitation tends to be more successful with children used to imitation games.

4.4 Description of procedure

All types of production experiments involve standardized procedures; and auditory stimuli are typically pre-recorded to ensure that every learner hears exactly the same. Moreover, items must be ordered in a way that avoids experimental artifacts. Specifically, series of auditory or visual stimuli tend to be interpreted as stories, which may cause problems. For instance, learners who hear several stimulus sentences with the same subject may change the subject to a pronoun in their own production (Levinsky & Gerken 1995). However, such effects can be reduced by filler sentences or breaks between items - or by using different subjects for consecutive utterances. Such measures can also minimize unintended priming effects.

Production experiments involve different procedures: in elicited imitation experiments, learners have to imitate stimulus sentences and researchers compare their production to the stimulus. Here, it is crucial to find complexity levels that allow learners to recall the content and lexical material of the sentence, but requires them to use their grammatical knowledge to reconstruct the sentence. For instance, children with an MLU between 1.5 and 3.5 typically attempt to imitate sentences with 3 to 5 syllables, often omitting at least 1 of them (Lust et al. 1996, Gallimore & Tharp 2006). Moreover, task effects should be minimized by keeping the communicative situation natural and engaging. For instance, in L1-studies, one can use puppets that "produce" recorded utterances; and children have to repeat these utterances for researchers who pretend not to understand the puppet. Typically, the target sentence is repeated twice, in a neutral context without rising intonation: *"The dog was chased by the cat". Say that: "The dog was chased by the cat".*

In elicited and speeded production experiments, learners can either be asked to complete an incomplete sentence (e.g. *This is a door. These are two …?*), or they can be asked to transform sentences or forms (e.g. *I'll say something and then you say the opposite* or *Can you give me the plural form for this?*). In order to evaluate learners' knowledge based on their performance in such a task, researchers should minimize task demands by giving learners a chance to familiarize themselves with any visual stimuli. They can also ask questions about stimuli that help learners to focus on the relevant aspects of these stimuli. Moreover, researchers must ensure that learners understand the cue words in their prompts, e.g. temporal adverbs such as *yesterday* in prompts for tense markers.

In syntactic priming experiments, learners are typically given picture descriptions as primes and asked to describe another picture; and researchers investigate whether the prime makes learners more

likely to use the primed construction (Chang et al. 2006, Kim & McDonough 2008). For such experiments, researchers must ensure that both constructions are pragmatically appropriate for describing prime and target stimuli. In order to avoid unintentional priming effects, many researchers use between-group designs: they compare a group hearing 1 prime type to another group hearing the other prime type. Here, all learners respond to the same set of target pictures, but each learner only hears 1 particular prime type. Alternatively, participants can be exposed to different prime types, but in blocked form and on different days. Then, the order in which the blocks are presented should be randomized across participants. Potential additional variables in priming experiments are the way in which participants interact with the prime (simply hearing vs. repeating) and the degree of lexical overlap between prime and target (same vs. different verb; same vs. different argument noun phrases).
In input/feedback-experiments, learners receive different types of feedback for deviations from the target (Saxton et al. 1998). Thus, it must be ensured that they produce such deviations: researchers can select phenomena for which learners of this developmental level tend to produce non-target-like forms. Alternatively, researchers use novel words, for which learners can be expected to apply general rules (e.g. past-tense *pell–ed*). Then, experimenters can treat novel words as irregularly inflected words and confront learners with "correct" irregular forms (*pell-pold*).

4.5 Analysis and outcomes

Production experiments provide researchers with a controlled way of studying how independent variables such as frequency affect dependent variables that offer measurements of speakers' production. The production experiments discussed above differ with respect to these measurements: In elicited imitation and production experiments, one calculates percentages of target-like and non-target-like responses and distinguishes between different error types; e.g. "overgeneralizations" (*go-ed*), omissions (*want* instead of *wanted*) or irregularizations (*truck* instead of *tricked*). The error analysis provides insights into acquisition mechanisms: For instance, overgeneralizations such as *go-ed* can be analyzed as evidence for learners' acquisition of linguistic generalizations. One can also investigate whether learners are more likely to overgeneralize regular endings to irregular forms that are similar to many regularly inflected word forms (*hold-held* vs. *fold-ed, mold-ed,* etc.) than to irregulars that are similar to many other irregular word forms (*sing-sang, ring-rang,…*). Such manipulations of form similarity can help us to determine whether learners' generalizations result from general rules - or whether they are analogical in nature, i.e. based on similarities to existing forms.

Percentages of target-like forms and error types can also be calculated for speeded production experiments. However, here the focus is on reaction times: if reaction times for low-frequency word forms are longer than those for high-frequency forms, this frequency effect is taken as evidence for full-form storage of this form whereas the lack of such an effect is viewed as evidence for decomposed entries (see section 4.1). For instance, Clahsen et al. (2004), found overgeneralizations of regularly inflected participles inflection and a frequency effect for irregulars only; which suggests that irregular word forms are stored as wholes, while regularly inflected word forms are computed.

In syntactic priming experiments, the focus is not on target-like or non-target-like performance. Rather, researchers analyse whether the percentage of a particular target construction is higher when it is presented after primes involving the same construction than when it is presented after primes involving different constructions.

5. Advantages and disadvantages

One of the major advantages of production methods is the fact that they can be applied to many different phenomena and learners and require comparatively small investments in equipment. Different

types of production studies complement one another: naturalistic production data are versatile, have high ecological validity and offer an excellent starting point for research on a broad range of phenomena, including learners' input. However, the lack of researcher control can lead to incomparable samples and make it difficult to study low-frequency phenomena, fine-grained semantic distinctions and the productivity of learners' utterances. Some of these problems can be overcome by using converging evidence from semi-structured elicitation studies that provide rich, but comparatively naturalistic data for low-frequency phenomena. Experimental studies allow researchers even more control and some types of production experiments can provide additional insights into underlying representations – in particular novel word, syntactic priming and speeded production studies. However, elicitation and experiments involve higher task demands and do not provide representative naturalistic speech samples for learners and their input.
Even when one combines different methods, one still has to consider the fact that comprehension seems to precede production in many domains of linguistic development. This suggests that, taken on their own, production methods do not provide a full picture of learners' linguistic representations.

6. Do's and don'ts

- Provide converging evidence from naturalistic samples, semi-structured elicitation and production experiments.
- Obtain representative input samples as well as representative samples of learners' own productions (if possible)
- Select talkative speakers with clear articulation and have a fall-back plan for dealing with attrition in longitudinal studies.
- Consider potential effects of researcher presence in naturalistic studies.
- Determine which situations encourage speech production and the use of target forms or constructions.
- Establish the appropriateness of stimuli and tasks in pilot-studies.
- Use state-of-the-art audio/and video-equipment.
- Use standardized formats for transcription, annotation and meta-data and at least 2 transcribers for reliability checks.
- Carry out a detailed distributional analysis that considers suppliance in obligatory contexts, over- and under-extensions, contrastive use, error types and shifts in the distribution of forms.
- Do not use compressed data format.
- Do not interpret all target-like forms or constructions as evidence for the availability of target-like representations.
- Do not interpret all non-target-like forms or constructions as evidence for a lack of target-like representations.

Acknowledgements

I would like to thank the editors and reviewers as well as Martin Henson, Bettina Landgraf, and Ingrid Sonnenstuhl for comments on this paper. The writing of the paper was supported by the Research Promotion Fund of the University of Essex.

References

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. 1993. *The CELEX Lexical Database* (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Behrens, H. (ed.). 2008. *Corpora in Language Acquisition Research. History, Methods, Perspectives.* Amsterdam: John Benjamins.

Berko, J. 1958. The child's learning of English morphology. *Word* 14: 150-177.

Berman, R. A., & Slobin, D. I. (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale, NJ: Lawrence Erlbaum.

Bernstein Ratner, N. 2000. Elicited imitation and other methods for the analysis of trade-offs between speech and language skills in children. In *Methods for Studying Language Production*. L. Menn & N. Bernstein Ratner (eds.), 291-312. Mahwah, NJ: Lawrence Erlbaum

Blom, E. 2007. Modality, Infinitives and finite bare verbs in Dutch and English Child Language. *Language Acquisition* 14: 75-113.

Bloom, P. 2000. *How Children Learn the Meanings of Words*. Cambridge, Mass: MIT Press.

Boersma, P. & Weenik, D. 2009. *Praat: Doing Phonetics by Computer* (Version 5.1.07). Retrieved May 12, 2009, from http://www.praat.org/ http://www.praat.org/

Bohnemeyer, J , Brown, P. & Bowerman, M. 2001. Cut and break clips. In *'Manual' for the Field Season 2001*, Levinson, S.C. & N.J. Enfield (eds), 90-96. Nijmegen: Max Planck Institute for Psycholinguistics.

Bol, G. W. & Kuiken, F. (1990). Grammatical analysis of developmental language disorders: A study of the morphosyntax of children with specific language disorders, with hearing impairment and with Down's syndrome. *Clinical Linguistics and Phonetics* 4: 77–86.

Branigan, H. (2007). Syntactic priming. *Language and Linguistics Compass* 1: 1-16.

Brown, R. 1973. *A First Language: The Early Stages.* London: Allen & Unwin.

Cameron-Faulkner, T Lieven, E. & Tomasello, M. 2003. A construction based analysis of child directed speech. *Cognitive Science* 27: 843-873.

Chang, F., Dell, GF. Bock, K. 2006. Becoming syntactic. *Psychological Review* 109: 35-54.

Chouinard, M. M. & Clark, E. V. 2003. Adult reformulation of child errors as negative evidence. *Journal of Child Language* 30: 637–69.

Clahsen, H., Hadler, M. & Weyerts, H. 2004. Speeded production of inflected words in children and adults. *Journal of Child Language* 31: 683-712.

Clahsen, Harald, Vainikka, Anne & Young-Scholten, Martha (1990). Lernbarkeitstheorie und Lexikalisches Lernen. Eine kurze Darstellung des LEXLERN-Projekts. *Linguistische Berichte* 130: 466-477.

Clark, E. 1993. *The Lexicon in Acquisition*. [Cambridge Studies in Linguistics]. Cambridge: Cambridge University Press.

Crain, S. & Thornton, R., 1998. *Investigations in Universal Grammar: A Guide to Experiments in the Acquisition of Syntax and Semantics.* Cambridge, MA: MIT Press.

Eisenbeiss, S. 2000. The Acquisition of the DP in German. In *Acquisition of Syntax. Issues in Developmental Linguistic*, L. Rizzi & M-A. Friedemann (eds.) *27-62.*

Eisenbeiss, S. 2003. Merkmalsgesteuerter Grammatikerwerb. Doctoral dissertation, University of Düsseldorf. "http://docserv.uni-duesseldorf.de/servlets/DerivateServlet/Derivate-3185/1185.pdf" http://docserv.uni-duesseldorf.de/servlets/DerivateServlet/Derivate-3185/1185.pdf)

Eisenbeiss, S. 2006. Documenting child language. In *Language Documentation and Description*, Volume 3, P.K. Austin (ed.), 106-140. London: SOAS, The Hans Rausing Endangered Languages Project.

Eisenbeiss, S. 2009a. Generative approaches to language learning. *Linguistics* 2: 273–310.

Eisenbeiss, S. 2009b. Contrast is the name of the game: contrast-based semi-structured elicitation techniques for studies on children's language acquisition. *Essex Research Reports in Linguistics* 57.7.

Eisenberg, S. 2005. When conversation is not enough. Assessing infinitival complements through elicitation. *American Journal of Speech-Language Pathology* 14: 92-106.

Eisenberg, S. L., Fersko, T. M., & Lundgren, C. 2001. The use of MLU for identifying language impairment in preschool children: A review. *American Journal of Speech-Language Pathology* 10: 323–342.

Ellis, N.C. & Larsen-Freeman, D. 2006. Language emergence: implications for applied linguistics - introduction to the special issue. *Applied Linguistics* 27: 558-589.

Gallaway, C. & Richards, B. (eds.) 1994. *Input and Interaction in Language Acquisition*. London: Cambridge University Press.

Gallimore, R. & Tharp, R. 2006. The interpretation of elicited sentence imitation in a standardized context. *Applied Linguistics* 27: 464-491.

Guasti, M.-T. 2002. *Language Acquisition.* Cambridge Mass: MIT Press.

Jackson-Maldonado, D. & Conboy, B.T. 2007. Utterance length measures for spanish-speaking toddlers: the morpheme versus word issue revisited. In *Communication Disorders in Spanish Speakers: Theoretical, Research and Clinical Aspects*, J.G. Centeno, R.T. Anderson & L.K. Obler (eds.), 142-155. Multilingual Matters.

Kim, Y. & McDonough, K. (2008). Learners' production of passives during syntactic priming activities. *Applied Linguistics* 29: 149-154.

Küntay, A. & Slobin, D.I. (1996). Listening to a Turkish mother: some puzzles for acquisition. In *Social Interaction, Social Context, and Language*, D.I. Slobin, J. Gerhardt, A. Kyratzis & J. Guo (eds.), 265-286. Mahwah, NJ: Lawrence Erlbaum.

Labov, W. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Leather, J. 1999. Second language speech research: an introduction. In *Phonological Issues in Language Learning*, J. Leather (ed.), 1-58. Oxford: Blackwell.

Levinsky, S., & Gerken, L. 1995. Children's knowledge of pronoun usage in discourse. In *Proceedings from the 26th annual Child Language Research Forum*, E. Clark (ed.), 89–104. New York: Cambridge University Press.

Lust, B., Flynn, S. & Foley, C. 1996. What children know about what they say : elicited imitation as a research method for assessing children's syntax. In *Methods for Assessing Children's Syntax.* D. McDaniel, C. McKee & H. Cairns (eds), 55-76. Cambridge, MA:MIT Press.

MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Lawrence Erlbaum.

MacWhinney, B., & Snow, C. 1990. The Child Language Data Exchange System: An update. *Journal of Child Language* 17: 457-472.

Marcus, G.F. 1993. Negative evidence in language acquisition? *Cognition* 46: 53-85.

Marcus, G.F., Pinker, S., Ullman, M., Hollander, M., Rosen, T.J., & Xu, F. 1992. *Overregularization in Language Acquisition*. [Monographs of the Society for Research in Child Development 57].

Mayer, M. 1969. *Frog, Where are You?* New York: Dial Book for Young Readers.

McDonough, K. & Mackey, A. 2006. Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning* 56: 693-720.

Menn, L. & Bernstein Ratner, N. (eds.). 2000. *Methods for Studying Language Production*. Mahwah, NJ: Lawrence Erlbaum,

Myles, F. 2005. Interlanguage corpora and second language acquisition research. *Second Language Research* 21: 373–391.

Pan, B., & Snow, C. 1999. The development of conversational and discourse skills. In *The development of language,* Barrett, M. (ed.), 229-250. Sussex: Psychology Press Ltd.

Papafragou, A., Hulbert, J. & Trueswell, J. 2008. Does language guide event perception? Evidence from eye movements. *Cognition* 108: 155-18.

Pinker, S. (1989). *Learnability and Cognition: The Acquisition of Argument Structure.* Cambridge, MA: MIT Press.

Proctor-Williams, K. & Fey, M.E. 2007. Recast density and acquisition of novel irregular past tense verbs. *Journal of Speech, Language and Hearin*g *Research* 50: 1029-1047.

Quay, S. 1995. The bilingual lexicon: implications for studies of language choice. *Journal of Child Language* 22: 369-387.

Radford, A. 1990. *Syntactic Theory and the Acquisition of English Syntax: The Nature of Early Child Grammars of English*. Oxford: Blackwell.

Saxton, M., Kulscar, B. Marshall & Rupra, M. 1998. Longer-term effects of corrective input: An experimental approach. *Journal of Child Language* 5: 701-21.

Senft, G. 2007. The Nijmegen space games: studying the interrelationship between language, culture and cognition. In *Experiencing New Worlds*, J. Wassmann & K. Stockhaus (eds.), 224-244.Oxford: Berghahn Books.

Sokolov, J. & Snow, C. (eds.). 1994. *Handbook of Research in Language Development Using CHILDES*. Hillsdale, NJ: Lawrence Erlbaum.

Stutterheim, C. von & Carroll, M. 2006. The impact of grammaticalised temporal categories on ultimate attainment in advanced L2-acquisition." In *Educating for Advanced Foreign Language Capacities: Constructs, Curriculum, Instruction, Assessment*, H. Byrnes (ed.), 40-53. Georgetown: Georgetown University Press.

Theakston, A.L., Lieven, E.V.M. & Tomasello, M. 2003. The Role of the Input in the Acquisition of Third Person Singular Verbs in English. *Journal of Speech, Language, and Hearing Research* 46: 863-877

Thornton, Rosalind (1996). Elicited production. In (Eds.), *Methods for Assessing Children's Syntax*. Dana McDaniel, Cecile McKee & Helen Smith Cairns Cambridge, MA: MIT Press, 77-102.

Tomasello, M. & Stahl, D. 2004. Sampling children's spontaneous speech: How much is enough? *Journal of Child Language* 31: 101-121

Tomasello, Michael (2001). The item-based nature of children's early syntactic development. In *Language Development. The Essential Readings*, Tomasello, M. & Bates, E. (ed.), 169-186. Oxford: Blackwell.

Vihman, M..M. 1996. *Phonological Development: The Origins of Language in the Child.* Oxford: Blackwell.

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

Wei, L. & Moyer, M.G. 2008. *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*. Oxford: Blackwell.