

A Computational Analysis of Interaction Patterns in the Acquisition of Turkish

Ayşe Pinar Saygın

Published online: 1 March 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract We provide a corpus-based computational approach to analyzing acquisition data on Turkish, a richly inflected language. We describe the process by which transcripts from the CHILDES database for 16 children aged 2;0 to 3;0 were morphologically tagged and parsed. We computed a number of imitation, overlap, and repetition measures on the transcripts using CLAN and CHIP programs. These measures tended to decrease as a function of mean length of utterance, which was broadly consistent with previously published work on English-speaking children. The data also revealed additional usage patterns, where the adult utterances provided children with rich morphosyntax in the input, while at the same time helping them to maintain discourse. Children on the other hand, tended to omit optional constituents and repeat morphemes from the previous utterance. The Turkish data and previously published English data showed cross-linguistic differences in repetition patterns that were congruent with the typological differences between the two languages. More generally, the data were consistent with a usage-based model for the acquisition of Turkish as a first language. The corpora and methods provided here can be extended to future applications.

Keywords Turkish · Acquisition · Morphology · Parsing · Computational linguistics · CHILDES · Interaction · Input · Crosslinguistic

A. P. Saygın (✉)

Department of Cognitive Science, Center for Research in Language, University of California,
San Diego, La Jolla, CA 92093-0515, USA
e-mail: asaygin@cogsci.ucsd.edu

1 Introduction

Turkish is an agglutinative language characterized by rich inflectional and derivational morphology, typological features that distinguish it sharply from English and many of the other commonly studied languages in acquisition research. Many features of Turkish, notably its morphosyntax, make it difficult to readily use some of the commonly used computational or corpus-based analytical methods in studying this language. In particular, CLAN (Computerized Language Analysis) programs, which are often used with the CHILDES (Child Language Data Exchange System) database, have been particularly useful toolsets in the study of language acquisition in many languages. However, applications of corpus based analyses to Turkish child language data have been limited (see [Acarlar and Johnston 2006](#)).

Turkish has extensive agglutinative morphology, vowel harmony, and free word order of major constituents ([Kornfilt 1997](#); [Underhill 1976](#)). The morphotactics of Turkish are highly regular and very flexible. On the other hand, the morphology is also complex and operates under phonotactic (e.g., vowel harmony), syntactic (e.g., agreement, case) and even semantic (e.g., case, tense, aspect) constraints. Despite this, Turkish-speaking children learn the morphological system rapidly and most children make productive use of morphology by the time they are 2 years old ([Aksu-Koc and Slobin 1985](#)). Turkish-speaking children also make errors much less frequently than children learning many other languages. As [Kuntay and Slobin \(1999\)](#) summarize, “the degree of morphological productivity exhibited by Turkish-speaking preschoolers goes far beyond what has been reported for child speech in Indo-European languages, even those with the morphological complexity of Slavic languages.”

How do Turkish-speaking children acquire the rich morphology of the language so early and seemingly without many errors? Research by Aksu-Koc and Slobin suggests that the relatively free word order may help to provide children with rich input from which to learn the morphology. Turkish is sometimes characterized as being a subject-object-verb (SOV) language, but speakers use other orders for constituents. Since case marking determines the grammatical function of a noun phrase in the sentence, changing the order of the constituents generally does not affect the grammaticality of a sentence. Sentences with non-typical orders usually have pragmatic contributions to discourse ([Taylan 1984](#)). Turkish-speaking children use a variety of word orders by the age of 2, and comprehend all six orders of subject, object and verb ([Aksu-Koc and Slobin 1985](#); [Kuntay and Slobin 1999](#); [Slobin and Bever 1982](#)). Child-directed Turkish contains a wide variety of pragmatic word order variations ([Aksu-Koc and Slobin 1985](#); [Kuntay and Slobin 1996](#); [Slobin and Bever 1982](#)).

[Kuntay and Slobin \(1996\)](#) have found that Turkish child-directed speech features repetitions of the same set of lexical items, or *variation sets*. Variation sets provide important cues for language learning ([Onnis et al. 2008](#)), and their structure has been shown to correlate with children’s acquisition of syntax ([Hoff-Ginsberg 1986](#); [Waterfall 2006](#)). In Turkish, variation sets featured a change in word order 25% of the time; when the verb was repeated in a successive utterance, it changed position 37% of the time ([Kuntay and Slobin 1996, 2002](#)). Kuntay and Slobin have noted: “If you listen to [the variations], even without knowing Turkish, the verb stands out as an acoustic unit. It is a sort of acoustic gestalt, which achieves saliency as it stands out against a shifting

background. The root, too, begins to stand out, against an array of different suffixes.” (1996, p. 276).

Here we aimed to provide methods that facilitate exploring Turkish child language corpora. We present a morphologically tagged and disambiguated version of the CHILDES Turkish corpus, containing transcripts of Turkish-speaking children aged 2;0 to 3;0¹ (MacWhinney 1999, 2000). This allows the CLAN tools to be readily used on the corpus to address questions regarding the acquisition of Turkish whilst having access to the morphology of the language (see Acarlar and Johnston 2006 for a similar approach on a corpus from older children using a different computational tool). As an example of the kinds of questions that can readily be probed with this method, we explored input and interaction patterns in the corpus using a computational approach (see Methods; Sokolov and Moreton 1994; Snow and Ferguson 1977; Snow 1994), which had been an area calling for further attention in Turkish acquisition research (Kuntay and Slobin 2002).

Consider an utterance “*gitmemeliydiniz*”, meaning “you (plural) should not have gone”.² And consider the response in the next utterance “*gitmemeliydik*” (we should not have gone). Since CHIP operates at the word level, if we used CHIP on this pair (see Methods for details), we would have concluded these two utterances have no overlap whatsoever. We would have no way of picking up on the fact that these utterances share 4 out of 5 morphemes. Instead, if we had simply written out the morpheme boundaries as if they were word boundaries as in: “*git me meli di niz*” (the y is not a distinct morpheme but orthographic convention), we could run the input and overlap analysis and be able to pick up the overlap with “*git me meli di k*”.

In this paper, we used an automated parser with manual semantic disambiguation to parse the corpus in a manner similar to the example above, and enable it to be used with CLAN and CHIP fairly straightforwardly. Morphological parsing of Turkish is an independent and active research problem of its own. We used a parser developed by Oflazer and colleagues, which is explained in detail in previous publications (Oflazer 1994; Hakkani-Tür et al. 2002; Eryigit et al. 2008).

Interaction patterns can reveal or highlight aspects of grammar to the children. Using previous utterances as frameworks for their own, children can improve communication in the short term and learning in the long term (Clark 1977; Masur 1995). Imitation of adult utterances can aid the learning of new words (Bloom et al. 1974, 1975; Masur 1995; Rogdon and Kurdek 1977; Reger 1986), the acquisition of grammar (Bloom et al. 1974; Reger 1986), or constitute an intermediate stage between comprehension and production (Whitehurst and Vasta 1975). Imitation is thought to facilitate lexical and grammatical development (e.g., Bloom et al. 1974; Masur 1995; Reger 1986) as well as aid discourse coherence (Keenan 1977; Ramer 1976).

To summarize the main findings, we found that adults provided children with rich morphosyntactic information in their utterances, from which children could copy and use morphemes in their own output. Our analyses of interaction patterns in the corpora

¹ Age is denoted in the Years;Months format.

² A more impressive example of agglutination is the word *Muvaffakiyetsizleştiricileştiriveremeyebileceklerimizdenmişsinizcesine*, but we chose a simple example that is likely to actually be uttered.

thus provide some clues regarding why Turkish-speaking children appear so advanced, even at early ages.

2 Methods

2.1 Corpora

We used 16 transcripts from the CHILDES database (MacWhinney 1999). The system is made up of three components: a standardized format for transcription of spoken language (CHAT), computer programs to aid in the analysis of these transcriptions (CLAN), and a database of transcriptions provided by various laboratories around the world.

We used the Turkish corpora contributed to the project by Slobin and colleagues. These are transcriptions of conversations with normally developing Turkish children gathered in 1972 and 1973 (Slobin 1982). The children were living in Istanbul and at least one parent had a university education. Children were visited in their homes. There was always an experimenter present, and sometimes also other adults and children. The activities in the transcripts varied and included play, mealtime, and social interaction. All files were transcribed in the CHAT (Codes for the Human Analysis of Transcripts) format, which allows them to be analyzed using CLAN (MacWhinney 2000). After parsing, we computed several input and interaction measures as detailed below.

The average age of the children was 30.5 months (st. dev.: 4.4; range 2;0 to 3;0) and the average mean length of utterance (MLU, in morphemes) was 3.73 (st. dev.: 0.66). We used MLU in morphemes as an index of language development (Eisenberg et al. 2001) since chronological age is generally not as reliable a predictor of language ability. Indeed, the highest MLU and the lowest MLU in our corpus belonged to children who were the same age (24 months). Overall though, chronological age and MLU correlated significantly but modestly at $r = 0.33$ in the Turkish corpus used here. The MLU and chronological age correlation may be stronger at later stages in development (Ege et al. 1998 for Turkish); or is sometimes not observed (Klee and Fitzgerald 1985 for English).

We also used a subset of a previously analyzed and published English corpus (Sokolov and Moreton 1994) in order to link the present data to previous work using the same measures, as well as to provide a cross-linguistic comparison. These data are also from CHILDES and were analyzed using the same methods detailed in Sect. 2.3. The corpus contained transcripts from children grouped into three levels of MLU (1.5, 2.5, and 3.5). In the quantitative comparisons reported, we selected subsets of the corpora such that the Turkish and English-speaking children were matched for MLU. For these analyses we used 13 Turkish-speaking children (mean MLU = 3.56) and 18 English-speaking children (mean MLU = 3.5). We used MLU in morphemes, though MLU in words and morphemes are very highly correlated in English indicating either can be used as a measure of gross language development (Parker and Brorson 2005). It was unfortunately not possible to compare English and Turkish at other points in language development because there were no transcripts of Turkish-speaking children

with lower MLUs, even though the transcripts started at age 2;0, like most data on CHILDES. Also note that when matched for MLU, it is not possible to also match Turkish and English-speaking children in chronological age. For a discussion of the difficulties that can arise when comparing MLUs between languages, see [Schnell de Acedo \(1994\)](#).

2.2 Parsing, Tagging and Disambiguation

One of our main goals was to enable computational access to the rich morphological structure of Turkish utterances. CLAN can be used to add a morphologically parsed tier to the transcripts in CHILDES (e.g., [Sagae et al. 2004](#)) but a grammar for Turkish was not available at the time of this study.

To explain the modification that the utterances underwent more clearly, we will use an example utterance from our corpus (Example 1 in “Appendix”). Upper-case letters are used to denote surface forms that vary (e.g., based on vowel harmony): A stands for a or e (e.g., the plural suffix can take the form -ler or -lar, and is denoted -lAr); H stands for i, ı, u, or ü (e.g., the genitive suffix is denoted as -Hn, and can be realized as -in, -ın, -un, or -ün); and D stands for d or t (e.g. the past tense suffix is denoted -DH).

Here, the *CHI tier contains the original utterance of the child in CHAT format, the %COM tier shows the different morphemes in the utterance (where | denotes a word boundary and + a morpheme boundary that is not a word boundary), the %ENG tier is a translation of the utterance, the %PAR tier is the end result of the parsing process. The %ENGMOR tier is not part of CHILDES and is added by the author to provide a morpheme by morpheme English translation. Abbreviations used in the %PAR tiers are also translated and expanded in the %ENGMOR tier, except for agreement markers (where 2sg stands for 2nd person singular, 3pl for 3rd person plural and so on). Not all of these layers will be shown for all of the examples.

We used a parser for Turkish developed with the aid of *Xerox Finite State Tools* ([Oflazer 1994](#); [Hakkani-Tür et al. 2002](#)). The aim was not to substitute a full grammatical decomposition in place of the data with a claim regarding the speakers’ knowledge or actual usage of the morphosyntactic categories the parser provides. The morphosyntactic analysis was not based on a theoretical underlying grammar. Rather, like [Parisse and LeNormand \(2000\)](#), we used the output of morphosyntactic parsing on actual utterances in order to study patterns in child and adult language. In line with what Ingram has called the *competence assumption*, according to which researchers should “not propose a linguistic construct until there is evidence for it in the child’s performance” (1989, p. 76), we only used morphemes that were realized on the surface. While there are good reasons, both from linguistic and computational perspectives, to utilize an elaborate morphological tagging scheme for Turkish ([Hakkani-Tür et al. 2002](#)), the parser’s raw output contained some grammatical information that was not appropriate for the current study. In particular, for conducting an input and interaction analysis, it is not reasonable to include linguistic units that are never uttered nor heard. Thus, any zero morphemes that were in the parser’s output (as Zero) were removed

from the analyses using a custom computer program written in C, which also converted the utterances into CHAT format.

Finally, there were ambiguous parses involving multiple possibilities underlying surface realization of morphemes (e.g., *evi* can mean ‘house’ in accusative case, or ‘house’ in possessive third person singular). The sentences become unequivocally anomalous (semantically and/or syntactically) with the incorrect parse. The parser cannot disambiguate these since its scope does not extend to sentential semantics. These words were disambiguated manually by the author (a native speaker of Turkish), taking sentential context into account. There were no instances for which the context was insufficient to disambiguate and select the intended parse, so all words in the corpus were eventually represented with a single parse of morphemes. In order to validate the disambiguation, two of the transcripts (one from a 24 month old, one from a 32 month old) were disambiguated additionally by a 59 year old native Turkish speaker who teaches Turkish as a foreign language. The native speaker was linguistically knowledgeable enough to do the task, but did not know about the specific goals of the research. She was told these were real world language transcripts parsed into morphemes with a computer program, and where there were multiple parses provided, she was asked to choose one that best fit the context. The second native speaker was also able to disambiguate all multiple parses. The agreement between this native speaker’s choices and the author’s was 100% on both transcripts.

2.3 Measures

We used the CLAN program CHIP, which was developed to analyze conversational interactions by comparing the utterances of a speaker to the previous utterances, coding the amount of lexical overlap and change between utterances (Sokolov and MacWhinney 1990; Sokolov and Moreton 1994). Since the Slobin Turkish corpus was now morphologically parsed, CHIP could operate on it, computing morphological overlap measures instead of lexical overlap, a more appropriate manner in which to study richly inflected languages (see Examples 2a and 2b in the Appendix).

CHIP computes a number of statistics, six of which are especially informative in looking at interactivity and overlap. The first three are general measures of imitation and overlap:

1. *Repetition Index*: The average proportion of overlapping words between the child utterance and the previous adult utterance.
2. *Overlap*: The proportion of overlapping utterances (with an overlap of at least one word with the previous adult utterance) to the total number of child utterances.
3. *Imitations*: The proportion of imitative utterances to the total number of child utterances.

The remaining three measures describe the type of imitation:

4. Proportion of *Exact* Imitations: The words in the source and response match exactly.
5. Proportion of *Expanded* Imitations: The response contains repetition, additions, but no deletions.

6. Proportion of *Reduced* Imitations: The response contains partial repetition with deletions but no additions.

For each transcript, these six measures were computed using CHIP. We also computed MLU in morphemes for each child. Since both MLU and the overlap measures are inherently continuous, we did not sort the children into groups, but instead computed regression analyses. Reported *p* values are corrected for multiple comparisons when necessary.

2.4 Examples

In the Appendix, we provide several examples from the corpus to illustrate how CHIP is used on the morphologically parsed corpus in order to acquire computational measures of overlap between utterances in Turkish. The examples also emphasize individual differences in children's imitiveness, which can be explored in the future using the measures calculated here (e.g., by identifying outliers, Bates et al. 2005).

In all examples, CHI denotes the child's utterances, EXP denotes the experimenter's utterances, and underlining is used to highlight repetitions.

If we ran CHIP for the exchange in Example 2a, the results for all measures would be zero. That the child repeats *seviyor* 'love-PROG', which contains two morphemes realized on the surface, would not be detected because CHIP is sensitive only to the word level. In the morphologically parsed corpus (Example 2b), CHIP can pick up the overlap (sev+Prog1) between the utterances.

Consider Example 3a, in which the *Repetition Index* is quite high (the underlined morphemes denote the overlap). Since Turkish is a pro-drop language, the child could have omitted both onlar (those) and kuşun (the bird's). In fact, this is what the majority of the children do (see below). This child on the other hand, seems to make use of the construction she just heard, possibly because it is easy to reproduce it from working memory. However, we can also tell that she is not just parroting, as she substitutes kanat (wing) for ne (what) in the question, and transforms the demonstratives appropriately (bunlar (this-plural) into onlar (that-plural)).

Another exchange is shown in Example 3b. Here, the child correctly substitutes herself as the agent in her response, and marks the agreement correctly on the verb. The *Repetition Index* in Example 3b is not zero, but is lower than that for Example 3c. In this example, the child correctly eliminates the question morpheme from her response, but she fails to change the agreement marker to match the agent (herself). The same child (the only child who made notable errors in the entire corpus), made the same agreement error multiple times in the same transcript. When asked a question about herself, which contains the second person singular agreement marker, she answered using second person singular. Thus this is a case in which high imitiveness led to errors.

3 Results

We explored relationships between the six overlap measures and language development in our sample of 16 children of age 2;0 to 3;0, with mean MLU of 3.73

Table 1 Means and standard deviations for the six measures for the Turkish corpus as a whole ($N = 16$; Mean MLU = 3.77), for the reduced corpus ($N = 13$; Mean MLU = 3.56) and for English-speaking children (Sokolov and Moreton 1994, $N = 18$, Mean MLU = 3.5)

Measure	Turkish ($N = 16$) MLU = 3.73	Turkish ($N = 13$) MLU = 3.56	English ($N = 18$) MLU = 3.5	Turkish vs. English
Overlap	0.47 (0.11)	0.46 (0.11)	0.21 (0.07)	***
Repetition index	0.51 (0.09)	0.52 (0.09)	0.48 (0.07)	n.s.
Imitations	0.08 (0.06)	0.09 (0.06)	0.04 (0.02)	*
Exact	0.007 (0.005)	0.007 (0.006)	0.01 (0.01)	n.s.
Expanded	0.007 (0.006)	0.006 (0.006)	0.02 (0.01)	***
Reduced	0.068 (0.059)	0.074 (0.058)	0.02 (0.01)	**

The last column shows p values for the statistical comparison of MLU-matched Turkish-speaking and English-speaking children

*** $p < 0.0001$; ** $p < 0.01$; * $p < 0.05$

(Slobin 1982). The mean values and standard deviations for the six overlap measures are reported in Table 1. We observed a significant relationship between MLU and *Repetition Index* ($r = -0.746$, $p < 0.001$, Fig. 1a). MLU was also related to *Imitations* ($r = -0.532$, $p < 0.05$, Fig. 1b) and *Reduced* imitations ($r = -0.583$, $p < 0.05$, Fig. 1c). The remaining measures were not significantly related to MLU.

These results are broadly consistent with those from a previous study with English-speaking children by Sokolov and Moreton (1994), where *Repetition Index*, *Imitations*, *Exact* and *Reduced* imitations were lower in the higher MLU group. To provide a more direct link to previous work using the same measures, we also present our data on Turkish in relation to the Sokolov and Moreton data (1994). In order to match MLU across languages (see Sect. 2.1), we selected a reduced corpus consisting of 13 Turkish-speaking children (mean MLU = 3.56).

As shown in Table 1, *Overlap* and *Imitations* were significantly higher in the Turkish corpus compared with English, whereas no significant difference was found in *Repetition Index* between the languages. *Exact* imitations did not differ between the two corpora. There was a striking difference between *Expanded* imitations between the two languages; the English-speaking children had the higher mean in only this measure. *Reduced* imitations were significantly higher in the Turkish-speaking group.

4 Discussion

In this paper, we described the morphological parsing, tagging and disambiguation of transcripts of Turkish-speaking children aged 2;0 to 3;0 from the CHILDES database (MacWhinney 1999, 2000). The modified corpus can now readily be used to make further contributions from this highly agglutinative language to work using corpora in language development.

Here, we provided an application of the morphologically parsed corpus, and explored input and interaction in Turkish child language. Although input, interaction and imitation has been widely discussed in relation to language use and learning

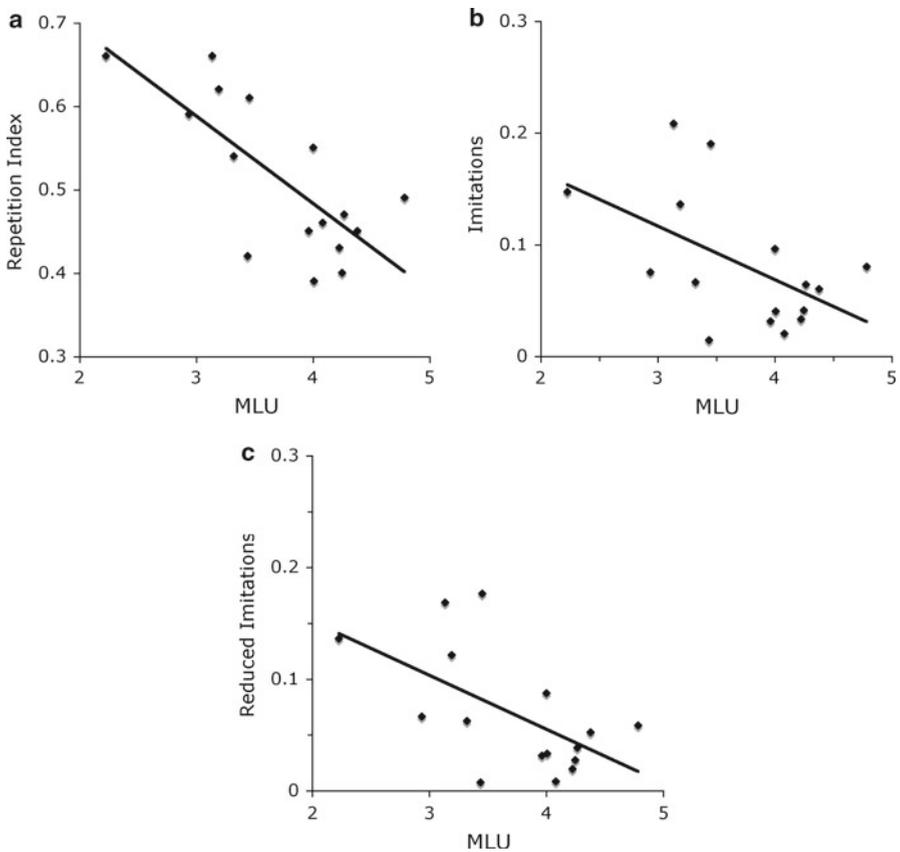


Fig. 1 Significant correlations were found between MLU and **a**) Repetition Index ($r = -0.746$, $p < 0.001$); **b**) Imitations ($r = -0.532$, $p < 0.05$); and **c**) Reduced Imitations ($r = -0.583$, $p < 0.05$)

(e.g., Ramer 1976; Snow and Ferguson 1977; Snow 1994; Stine and Bohannon 1983; Valian 1999), it had been understudied in Turkish child language (Kuntay and Slobin 1999). We computed measures of interaction and overlap between adults and children in the Turkish corpus using the CHIP program, which yielded 6 measures of overlap and imitation of utterances (see “Methods”).

Overall, whereas children tended to use the bare minimum of constituents in their speech, adults provided even the optional ones, possibly in order to enrich the input and facilitate language acquisition (Kuntay and Slobin 2002; Cameron-Faulkner et al. 2003; Snow 1983).

There was an inverse relationship between *Repetition Index* and MLU (Fig. 1; also seen for English, Sokolov and Moreton 1994), which is not unexpected, since this measure is indirectly proportional to length of utterance. E.g., an utterance of length 3 will have a *Repetition Index* of 0.66 if two words are repeated, whereas an utterance of length 4 will have a *Repetition Index* of 0.5.

We also found that imitations decreased with increasing MLU (Fig. 1). Specifically, *Exact* and *Reduced* imitations tended to decrease, whereas *Expanded* imitations tended to increase, although among these only *Reduced* imitations reached significance. Imitation is thought to facilitate lexical and grammatical development (e.g., Bloom et al. 1974; Masur 1995; Reger 1986) as well as aid discourse coherence (Clark 1977; Keenan 1977; Ramer 1976). The decrease in *Imitations* as MLU increases is consistent with this research, and suggests that as their language skills develop, children rely less on previous utterances in conversation (Ramer 1976; Stine and Bohannon 1983). It is possible that recent utterances help reduce processing load for the children as they are more readily accessed from working memory (Bohannon and Stanowicz 1989; Speidel and Herreshoff 1989), and that this may be particularly needed in earlier stages of development.

We also observed that Turkish-speaking children differed qualitatively and quantitatively from English-speaking children. In comparison to a group of English speaking children matched for MLU, Turkish-speaking children produced significantly more utterances in which at least one morpheme was repeated from the previous adult utterance (*Overlap*). On the other hand, the average proportion of overlap between the child utterance and a previous adult utterance (*Repetition Index*) did not differ between the languages. In other words, Turkish-speaking children did not repeat more morphemes from previous utterances, but they did so in more utterances. The difference in *Overlap* is likely caused by the regularity of Turkish verbal morphology, especially that of tense/aspect/mood suffixes. Children learning Turkish acquire case marking early and make use of it in their utterances with surprisingly low error rates (Aksu-Koc and Slobin 1985). Also, Turkish relies on obligatory overt morphemes (case marking) for denoting agents and patients, whereas English relies on word order.

The distributions of the different types of imitative utterances in the two languages are consistent with typological differences between the languages. *Reduced* repetitions are more common in Turkish child language compared with English. On the other hand, the proportion of *Expanded* repetitions is significantly less in Turkish than in English. As Turkish is a pro-drop language, speakers tend not to use overt subjects, as the agreement marker on the verb already carries this information. In conversation, when context provides sufficient information, it is also common to omit the object(s). Utterances in which children use all constituents in the previous utterance (e.g., Example 3a) are very rare. Most of the time, a child can use the verb from the previous utterance with the same tense/aspect/mood markers (thus repeating the verb stem as well as one or more other morphemes), the agreement suffix being the only morpheme she needs to change. An utterance in which the child does just that was provided in Examples 2a and 2b. In fact, some children overuse this strategy and make errors by repeating the agreement marker in addition to the rest of the morphemes (e.g., Example 3c). Overall, *Imitations* in the Turkish-speaking children correlated with *Exact* imitations at $r = 0.69$ ($p < 0.01$) and with *Reduced* imitations at an extremely high $r = 0.97$ ($p < 0.0001$), indicating that Turkish children are making use of the aforementioned characteristics of the language they are learning.

Conversely, in the Turkish corpus, 89% of adult imitations were *Expanded*. It has been reported that in talking to children, adults often provide slightly varying repetitions, which by providing the children with statistical information, can aid their lan-

guage development (Hoff-Ginsberg 1985, 1986, 1990; Onnis et al. 2008). In Turkish, adults tend to use the subject, the object(s) and the verb in their utterances (Aksu-Koc and Slobin 1985; Kuntay and Slobin 1996; Slobin and Bever 1982) which leads to the high proportion of *Expanded* imitations. Here, we were able to reveal these patterns with a computational approach. The corpus can be used in a similar fashion to extract *variation sets* in Turkish child-directed speech (Brodsky et al. 2007; Kuntay and Slobin 1999, 2002; Waterfall 2006).

Whether or not this pattern is limited to child-directed language,³ our analyses revealed an interesting dynamic: Children omit the constituents in their responses to adults, whereas adults tend to include them in the first place, and to reproduce or introduce them in further utterances. It is possible that this approach by the adults allows the properties of the language (e.g., optional omission of subjects and objects) to become transparent to the child in earlier stages of learning, and reinforces it further during the course of language acquisition (Kuntay and Slobin 1996). Behind their precocious appearance (cf. the significant difference in MLU in comparison to the English-speaking children, the almost entirely error-free performance in the Slobin CHILDES corpus, even of the youngest children), children acquiring Turkish are not so different from children learning other languages. Our approach highlights that these children's speech output is aided by the adults' utterances in conversation (Bloom et al. 1974, 1975; Masur 1995; Ramer 1976; Reger 1986; Stine and Bohannon 1983), which in Turkish include many morphemes (or sets of morphemes, Saygin and Wilson 2002; Goldberg 2003) a child can simply repeat. As such, these patterns are consistent with accounts of language acquisition where representations are extracted from the patterns in language input, gradually becoming less item-specific and more abstract (e.g., Saffran et al. 1996; Tomasello 2003). Turkish-learning children may be taking advantage of the regular and transparent morphology of the language, plus the optional constituents, a process that could be achieved with modest reliance on working memory, and can aid development towards full productivity (Bloom et al. 1974; Bohannon and Stanowicz 1989; Speidel and Herreshoff 1989). It will be interesting to explore how this stage of acquisition may be used in bootstrapping further language development.

5 Conclusion and Future Work

We provided methods for morphologically segmenting the CHILDES Turkish corpus and have shown an example domain in which these methods can be used (input and interaction analysis). Conducting similar analyses with other languages could shed further light on the nature and function of input and interaction in language acquisition (e.g., Behrens 2006; Clark and Bernicot 2008). Another approach would be to assess children's language development in relation to interaction patterns (e.g., Hoff-Ginsberg 1985, 1990), which can shed light on the function of imitation in language acquisition. Our data also provide baseline measures for the acquisition of Turkish

³ Although unquantified, in this native speaker's life-long experience with the language, *Expanded* repetitions are nowhere nearly as frequent in adult conversations.

that can be used in future studies. The corpus can be used further to explore individual differences (Lieven 1984), where, for example, outliers in language learning can be identified and quantified (Bates et al. 2003).

Acknowledgments I would like to thank the late Elizabeth Bates for her support of the work presented here as well as many invaluable discussions; Kemal Oflazer for advice on using his parser; Jeremy Boyd and Holger Keibel for helpful comments.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix: Examples

- (1) *CHI: bunun kolu nerede?
 %COM: bu+nun | kol+u | nere+de?
 %ENG: where is this (one)'s arm?
 %PAR: bu+Gen | kol+P3sg | nere+Loc
 %ENGMOR: it+ genitive | arm+ possessive3sg | where+locative?
- (2a) *EXP: sen anneni seviyormusun?
 %ENG: do you love your mother?
 *CHI: seviyorum.
 %ENG: I do.
- (2b) *EXP: sen anne+ni sev+iyor+mu+sun?
 %PAR: sen anne+P2sg+Acc sev+Prog1+mQues+A2sg?
 %ENGMOR: sen = you | anneni = mother+possessive2sg+accusative |
 seviyormusun = love+progressive+question morpheme+agreement2sg
 *CHI: sev+iyor+um.
 %PAR: sev+Prog1+A1sg.
 %ENGMOR: seviyorum = love+progressive+agreement1sg
- (3a) *EXP: bun+lar kuş+un ne+ler+i?
 %ENG: What are these (of the bird)?
 %PAR: bu+A3pl kuş+Gen ne+A3pl+Acc?
 %ENGMOR: bunlar = this+plural (these) | kuşun = bird+genitive |
 neleri = what+plural+accusative
 *CHI: on+lar kuş+un kanat+lar+ı.
 %ENG: Those are the wings of the bird.
 %PAR: on+A3pl kuş+Gen kanat+A3pl+Acc.
 %ENGMOR: onlar = this+plural (those) | kuşun = bird+genitive |
 kanatları = wing+plural+accusative
- (3b) *EXP: kim ütül+üyor?
 %ENG: who does the ironing?
 %PAR: kim ütüle+Prog1?
 %ENGMOR: kim = who | ütülüyor = iron+progressive

***CHI:** ben ütül+üyör+um.

%ENG: I do the ironing.

%PAR: ben ütüle+Prog1+A1sg.

%ENGMOR: ben = I | ütülüyörüm = iron+progressive + agreement1sg.

(3c) ***EXP:** yap+ar+mı+sın?

%ENG: (can) you do (it)?

%PAR: yap+Aor+mu+A2sg?

%ENGMOR: yaparmısın = do+aorist+question_morpheme+agreement2sg

***CHI:** yap+ar+sın (Ungrammatical: Agreement error).

%ENG: *you (can) do it (Intended meaning: I can do it).

%PAR: yap+Aor+A2sg

%ENGMOR: yaparsın = do+aorist+agreement2sg

References

- Acarlar, F., Johnston, J. (2006). Computer-based analysis of Turkish child language: Clinical and research applications. *Journal of Multilingual Communication Disorders*, 4, 78–94.
- Aksu-Koc, A. (1994). Development of linguistic forms: Turkish. In R. A. Berman & D. Slobin (Eds.), *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Aksu-Koc, A., & Slobin, D. (1985). The acquisition of Turkish. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition: Vol 1. The data*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bates, E., Appelbaum, M., Salcedo, J., Saygin, A. P., & Pizzamiglio, L. (2003). Quantifying dissociations in neuropsychological research. *Journal of Experimental and Clinical Neuropsychology*, 25, 1128–1153.
- Bates, E., Saygin, A. P., Moineau, S., Marangolo, P., & Pizzamiglio, L. (2005). Analyzing aphasia data in a multidimensional symptom space. *Brain and Language*, 92, 106–116.
- Behrens, H. (2006). The input-output relationship in first language acquisition. *Language and Cognitive Processes*, 21, 2–24.
- Bloom, L., Hood, L., & Lightbown, P. (1974). Imitation in language development: If, when, and why. *Cognitive Psychology*, 6, 380–420.
- Bloom, L., Lightbown, L., & Hood, L. (1975). Structure and variation in child language. *Monographs for the Society for Research in Child Development*, 40, Serial No. 160.
- Bloom, P., & Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of Child Language*, 24(3).
- Bohannon, J., & Stanowicz, L. (1989). Bidirectional effects of imitation and repetition in conversation: A synthesis within a cognitive model. In G. Speidel & K. Nelson (Eds.), *The many faces of imitation in language learning*. New York: Springer.
- Brodsky, P., Waterfall, H., & Edelman, S. (2007). *Characterizing motherese: On the computational structure of child-directed language*. (Paper presented at the Cognitive Science Society Conference, Nashville, TN, USA).
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, 27, 843–873.
- Clark, R. (1977). What's the use of imitation?. *Journal of Child Language*, 4, 341–358.
- Clark, E. V., & Bernicot, J. (2008). Repetition as ratification: How parents and children place information in common ground. *Journal of Child Language*, 35, 349–371.
- Ege, P., Acarlar, F., & Guleryuz, F. (1998). The relationship between age and mean length of utterance in acquisition of Turkish. *Turkish Journal of Psychology*, 13(41), 19–31.
- 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.
- Eryigit, G., Nivre, J., & Oflazer, K. (2008). Dependency parsing of Turkish. *Computational Linguistics*, 34, 357–389.

- Goldberg, A. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Science*, 7, 219–224.
- Hakkani-Tür, D., Oflazer, K., & Tür, G. (2002). Statistical morphological disambiguation for agglutinative languages. *Journal of Computers and Humanities*, 36, 381–410.
- Hoff-Ginsberg, E. (1985). Relations between discourse properties of mothers' speech and their children's syntactic growth. *Journal of Child Language*, 12, 367–385.
- Hoff-Ginsberg, E. (1986). Function and structure in maternal speech: Their relation to the child's development of syntax. *Developmental Psychology*, 4, 155–163.
- Hoff-Ginsberg, E. (1990). Maternal speech and the child's development of syntax: A further look. *Journal of Child Language*, 17, 85–99.
- Ingram, D. (1989). *First language acquisition: Method, description, and explanation*. Cambridge: Cambridge University Press.
- Keenan, E. (1977). Making it last: Uses of repetition in children's discourse. In S. Ervin-Tripp & C. Mitchell-Kernan (Eds.), *Child discourse*. New York, NY: Springer.
- Klee, T., & Fitzgerald, M. D. (1985). The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language*, 12, 251–269.
- Kornfilt, J. (1997). *Turkish*. London: Routledge.
- Kuntay, A., & Slobin, D. (1996). Listening to a Turkish mother: Some puzzles for acquisition. In D. Slobin, J. Gerhardt, A. Kyratzis, & J. Guo (Eds.), *Social interaction, social context, and language: Essays in honor of Susan Ervin-Tripp*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kuntay, A., & Slobin, D. (1999). The acquisition of Turkish as a native language: A research review. *Turkish Languages*, 3, 151–188.
- Kuntay, A., & Slobin, D. I. (2002). Putting interaction back into child language: Examples from Turkish. *Psychology of Language and Communication*, 6, 5–14.
- Lieven, E. V. M. (1984). Interaction style and children's language learning. *Topics in Language Disorders*, 4, 15–23.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk, 3rd Edn*. Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, B. (1999). The CHILDES system. In E. William, C. Ritchie, E., Tej K. Bhatia, et al. (Eds.), *Handbook of child language acquisition*, pp. 457–494. San Diego, CA, USA.
- Masur, E. (1995). Infants' early verbal imitation and their later lexical development. *Merrill-Palmer Quarterly*, 41, 286–306.
- Oflazer, K. (1994). Two-level description of Turkish morphology. *Literary and Linguistic Computing*, 9(2).
- Onnis, L., Waterfall, H., & Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition*, 109(3), 423–430.
- Parsis, C., & LeNormand, M.-T. (2000). How children build their morphosyntax: The case of French. *Journal of Child Language*, 27, 267–292.
- Parker, M. D., & Brorson, K. A. (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language*, 25, 365–376.
- Ramer, A. L. H. (1976). The function of imitation in child language. *Journal of Speech and Hearing Research*, 19, 700–717.
- Reger, Z. (1986). The functions of imitation in child language. *Applied Psycholinguistics*, 7, 323–352.
- Rogdon, M. M., & Kurdek, L. (1977). Vocal and gestural imitation in 8-, 14-, and 20-month old children. *The Journal of Genetic Psychology*, 131, 115–123.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Schnell de Acedo, B. (1994). Early morphological development: The acquisition of articles in Spanish. In E. Jeffrey, L. Sokolov, E. Catherine, E. Snow, et al. (Eds.), *Handbook of research in language development using CHILDES*, pp. 210–253. Hillsdale, NJ, USA.
- Saygin, A. P., & Wilson, S. M. (2002). *Paradigm reanalysis and the representation of morphologically complex words in Turkish*. (Paper presented at the 38th Annual Meeting of the Chicago Linguistic Society. University of Chicago, Illinois, USA).
- Sagae, K., MacWhinney, B., & Lavie, A. (2004). Automatic parsing of parent-child interactions. *Behavior Research Methods*, 36, 113–126.

- Slobin, D. (1982). Universal and particular in the acquisition of language. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 128–172). New York: Cambridge University Press.
- Slobin, D. (1995). Converbs in Turkish child language: The grammaticalization of event coherence. In M. Haspelmath & E. Koenig (Eds.), *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms - adverbial participles, gerunds*. Berlin/New York: Mouton de Gruyter.
- Slobin, D., & Bever, T. G. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition*, 12(3), 229–265.
- Snow, C. E. (1983). Saying it again: The role of expanded and deferred imitations in language acquisition. In K. E. Nelson (Ed.), *Children's language, Vol. 4* (pp. 29–58). New York: Gardner Press.
- Snow, C. E. (1994). Beginning from baby talk: Twenty years of research on input and interaction. In C. Galloway & B. Richards (Eds.), *Input and interaction in language acquisition* (pp. 3–12). London: Cambridge University Press.
- Snow, C. E., & Ferguson, C. A. (1977). *Talking to children: Language input and acquisition*. Cambridge: Cambridge University Press.
- Sokolov, J. L., & MacWhinney, B. (1990). The CHIP framework: Automatic coding and analysis of parent-child conversational interaction. *Behavior Research Methods, Instruments, & Computers*, 22(2), 151–161.
- Sokolov, J. L., & Moreton, J., et al. (1994). Individual differences in linguistic imitateness. In E. Jeffrey, L. Sokolov, E. Catherine, & E. Snow (Eds.), *Handbook of research in language development using CHILDES* (pp. 174–209). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sokolov, J. L., & Snow, C. E. (Eds.). (1994). *Handbook of research in language development using CHILDES*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Speidel, G., & Herreshoff, M. (1989). Imitation and the construction of long utterances. In G. Speidel & K. Nelson (Eds.), *The many faces of imitation in language learning*. New York: Springer.
- Stine, E., & Bohannon, J. (1983). Imitations, interactions, and language acquisition. *Journal of Child Language*, 10, 589–603.
- Taylan, E. E. (1984). *The function of word order in Turkish grammar*. Berkeley: University of California Press.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Underhill, R. (1976). *Turkish grammar*. Cambridge, MA: MIT Press.
- Valian, V. (1999). Input and language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of child language acquisition* (pp. 497–530). New York: Academic Press.
- Waterfall, H. R. (2006). *A little change is a good thing: Feature theory, language acquisition and variation sets*. Ph.D. thesis, University of Chicago.
- Whitehurst, G., & Vasta, R. (1975). Is language acquired through imitation? *Journal of Psycholinguistic Research*, 4.