On the role of morphological richness in the early development of noun and verb inflection

**Aris Xanthos[1], Sabine Laaha[2], Steven Gillis[3],
Ursula Stephany[4], Ayhan Aksu-Koç[5], Anastasia Christofidou[6],
Natalia Gagarina[7], Gordana Hrzica[8], F. Nihan Ketrez[9],
Marianne Kilani-Schoch[1], Katharina Korecky-Kröll[2],
Melita Kovačević[8], Klaus Laalo[10], Marijan Palmović[8],
Barbara Pfeiler[11], Maria D. Voeikova[12], and Wolfgang U. Dressler[2]**

Abstract

This study proposes a new methodology for determining the relationship between child-directed speech and child speech in early acquisition. It illustrates the use of this methodology in investigating the relationship between the morphological richness of child-directed speech and the speed of morphological development in child speech. Both variables are defined in terms of mean size of paradigm (MSP) and estimated in a set of longitudinal spontaneous speech corpora of nine children and their caretakers. The children are aged 1;3–3;0, acquiring nine different languages that vary in terms of morphological richness. The main result is that the degree of morphological richness in child-directed speech is positively related to the speed of development of noun and verb paradigms in child speech.

[1] University of Lausanne, Switzerland
[2] Austrian Academy of Sciences, Austria
[3] University of Antwerp, Belgium
[4] University of Cologne, Germany
[5] Yeditepe University and Boğaziçi University, Turkey
[6] Greek Academy of Sciences, Greece
[7] Centre for General Linguistics and Typology (ZAS), Germany
[8] University of Zagreb, Croatia
[9] Yale University, USA
[10] University of Tampere, Finland
[11] National Autonomous University of Mexico, Mexico
[12] Russian Academy of Sciences, Russian Federation

Corresponding author:
Aris Xanthos, Department of Computer Science and Mathematical Methods, University of Lausanne, Anthropole, CH-1015 Lausanne, Switzerland.
Email: Aris.Xanthos@unil.ch

Keywords
child-directed speech, language typology, mean size of paradigm, morphological development, morphological richness

Introduction

In this study we examine the role of morphological richness, as represented in the language addressed to young children, for children's early development of noun and verb morphology. The purpose of the present study is twofold. First, we introduce a new methodology to investigate the question of whether the richness of a morphological system, as represented in adult child-directed speech, is related to the speed at which this system develops in early childhood. Second, we apply this methodology to corpora of child-directed speech (CDS) and child speech (CS) in nine different languages to test our hypothesis regarding this relation.

The importance of variation in child-directed speech

In constructing and distinguishing lexical classes such as nouns and verbs, children must be attending to how members of the same class behave in the input with respect to combinatorial as well as to semantic properties. Caretakers display these properties in a language which – in contrast to the language addressed to adult speakers – is more clearly articulated, uses a reduced vocabulary, is in general syntactically less complex and consists of frequent repetitions and rephrasings (Aksu-Koç, 1998; Hoff, 2006; Hoff-Ginsberg, 1985; Pine, 1994; Snow, 1972, 1986, among others).

In studies of child-directed speech (CDS), four factors affecting children's early language development are frequently cited: frequency, utterance position (salience), morphological simplicity and pragmatic foregrounding (Choi, 2000; Goldfield, 1993; Tardif, Shatz, & Naigles, 1997; see also Gentner, 1982). For example, in their cross-linguistic study on lexical development in English, Italian and Mandarin, Tardif et al. (1997) argue that cross-linguistic differences in the predominance of nouns vs. verbs in early child speech can be explained by a combination of these four factors: English CDS tends to emphasize nouns (by placing them in utterance-final position, having fewer morphological variations on nouns and asking questions about objects), whereas Mandarin CDS tends to emphasize verbs (by producing them much more frequently than nouns, placing them in utterance-final position and having fewer morphological variations on verbs).

However, a number of studies on CDS stress the importance of *variation* for children's early language development. For example, Küntay and Slobin (1996) analysing CDS in Turkish, argue that the rate of repetition of verbs, which display a higher degree of inflectional variety, contributes to explain an early verb-learning bias in Turkish children. With these and further observations, Küntay and Slobin (1996, 2001) demonstrate that CDS – with its variation sets – provides the child with significant information about language structure. Naigles and Hoff-Ginsberg (1998), analysing order of acquisition in a set of 25 commonly used English verbs, observe that hearing particular verbs used more frequently and diversely leads children to a richer and more flexible understanding of those verbs. Similarly, Brodsky, Waterfall, and Edelman (2007) report a longitudinal

investigation of CDS in English where they found high correlations between children's production of a particular structure and parents' manipulation of that structure in variation sets. Tare, Shatz, and Gilbertson (2008) suggest that maternal use of English non-object terms in varied intentional and linguistic contexts helps the child to identify those terms. Wijnen, Kempen, and Gillis (2001) analysing CDS in Dutch, show that lexical variation (or informativeness) of verbs in infinitival form contributes to explain the root infinitive phenomenon in Dutch early child language.

The present study addresses another question in the same domain: is variation in CDS related to the rate of children's acquisition? In particular, our study proposes a comparison of the rate of noun and verb inflectional development in children acquiring languages which display different degrees of morphological richness. A similar issue is raised by Caselli, Casadio, and Bates (1999) in their comparative CDI (communicative development inventory) study of early lexical and grammatical development in English and Italian (see also Devescovi et al., 2005, p. 782–783; Laaha, 2004, p. 257; Stephany, 1997, p. 200):

> Italian children will have to acquire far more inflectional morphology than their English learning counterparts. . . . This problem can be resolved in one of two ways (with various points in between): (1) language learning may take much longer in Italian than it does in English, or (2) Italian children may keep pace with their English-speaking counterparts in the proportion of their target grammar that they are able to produce at any given point. (Caselli et al., 1999, p. 105)

Caselli and colleagues' results seem to support the hypothesis that morphological variation in CDS is positively correlated with the rate of morphological development in child speech. However, they note that 'much more evidence will be required to settle the issue, including evidence from free speech and structured elicitation' (Caselli et al., 1999, p. 105). In this article evidence from free speech in nine different languages is presented.

Morphological richness: Definitions and assumptions

Well-defined concepts of morphological richness or complexity have rarely been used in acquisition studies or even in language typology. In his discussion of grammatical complexity metrics, McWhorter (2005, p. 45) states: 'an area of grammar is more complex than the same area in another grammar to the extent that it encompasses more overt distinctions and/or rules than another grammar' (cf. the similar notion of structural complexity in Miestamo, Sinnemäki, & Karlsson, 2008). In the domain of inflectional morphology, this definition of richness needs to be further specified. Indeed, the morphological richness of an inflectional system can be divided into two distinct and interrelated components: syntagmatic and paradigmatic.

Syntagmatic richness refers to the capacity of a language to combine several inflectional affixes in a single word-form (Comrie, 1981; Greenberg, 1954/1960). This is what morphological richness consists of according to Hawkins (2004, p. 166). Thus, an English verb, which can only take a single tense or agreement marker (e.g. *walk-ed, walk-s*), is syntagmatically less rich than a Turkish verb, which may carry a number of suffixes

(e.g. *yürü-ye-mi-yecek-ti-m*, walk-ABIL-NEG-FUT-PAST-1SG, 'I was not going to be able to walk'). *Paradigmatic* richness, on the other hand, refers to the tendency of a language to have a large number of formally distinct inflected word-forms per lemma (Dressler, 2004). Thus, an English noun can only be inflected for number, as in *house* vs. *houses*, whereas Russian can distinguish six non-homophonous case forms in the singular and five in the plural. In the present article, we are specifically concerned with the *paradigmatic* richness[1] of inflectional morphology; this is what is meant here by the term morphological richness. We further restrict our attention to word-internal or synthetic morphology. In this context, *walked* counts as a form in the paradigm of *walk*, *walk-s*, *walk-ing*, whereas the periphrastic or analytic forms *is walking*, *have walked* do not add any further to the size of this paradigm.

When considering the degree of morphological richness of a given inflectional system, it is important to understand the difference between the grammatical knowledge that is available regarding the system in question, on the one hand, and the traces of the system as they show through the data, on the other hand. In our perspective, this is the basis of a distinction between *theoretical* and *observed* morphological richness. As a rule, only a reduced fraction of the theoretical morphological richness of a system will be observed in any given sample. The difference between theoretical and observed richness may vary considerably across different samples, in a way that crucially depends on sample size and that can be strongly affected by a number of situational and linguistic factors.

Among these factors, the present study is chiefly concerned with register and development. As a register, CDS is expected to display a relatively low morphological richness when compared to adult speech directed to adults. However, as simplified as CDS may be, the degree of morphological richness in samples of CDS tends to reflect the *theoretical* richness of the corresponding inflectional system. More precisely, samples of CDS will usually display a relatively higher richness in a 'theoretically' rich language than in a less rich one (Laaha & Gillis, 2007).

As regards development, morphological richness in samples of CS is expected to be globally increasing over time, as the child's productions display an increasing diversity of inflected word-forms for each lemma. In fact, it is hard to explain the emergence of adult language without assuming that morphological richness increases over the course of development both in CS and CDS. For the purpose of this research, however, we consider only the development of CS, and treat the morphological richness of CDS as a non-developing factor. Practically, this means that for each child–caretaker pair in our data, CS samples are monitored in a longitudinal, month-based fashion, while CDS samples are merged into a single dataset. This way, we attempt to focus on the relation between early development and the part of morphological richness in CDS that depends on the theoretical richness of the language – which is assumed not to vary at this time scale.

Mean size of paradigm

There is no widely accepted way to measure morphological richness (or complexity) on the basis of a sample (see Xanthos & Gillis, 2010, for a review of the literature). Arguably, the first quantitative index suitable for cross-linguistic acquisition studies was the *inflectional diversity* (ID) measure developed by Malvern, Richards, Chipere, and Durán (2004), based on the measure of lexical diversity D.

Xanthos and Gillis (2010) advocate an alternative approach starting out from an intuitive characterization of morphological richness in terms of an average number of distinct inflected word-forms per lemma. In its simplest version, *mean size of paradigm* (MSP) is defined as:

$$\text{MSP} := \frac{|F|}{|L|}$$

where $|F|$ stands for the number of distinct inflected word-forms in a sample and $|L|$ for the number of distinct lemmas. Thus, given the sample '*has, are, have, has, are*', containing 5 inflected English verb forms (tokens), one finds $|L| = 2$ (HAVE and BE), and $|F| = 3$ (*has, have* and *are*), so that MSP = 3/2 = 1.5 (for similar proposals, see Küntay & Slobin, 1996; Laaha, 2004, p. 188; Ogura, Dale, Yamashita, Murase, & Mahieu, 2006; Stephany, 1985, pp. 113–114).[2]

MSP ranges between 1 and $|F|$. Since the number $|F|$ of different word-forms in a sample cannot exceed the size (in tokens) of that sample, it follows that the maximum value of MSP is dependent on sample size. However, Xanthos and Gillis (2010) show that this dependence can be controlled for by applying a resampling procedure based on the work of Johnson (1944). The idea is to randomly construct a number of subsamples on the basis of the original corpus (say *B* subsamples), evaluate MSP on each subsample and finally report the average of these *B* MSP values. If *S* is the number of tokens per subsample (an arbitrary parameter), this average value is called the *normalized* MSP over *S* tokens, or MSP(*S*). This measure is the basis of our evaluation of morphological richness in both CDS and CS, and it is suitable for deriving the speed of development of morphological richness in CS.

Present study

The present study, which investigates the relationship between morphological richness in CDS and the speed of morphological development in CS, is concerned with the early phases of morphological acquisition, from emergence through what has been called the second, grammaticized phase by Berman (2004, p. 13).

In order to study the relationship between morphological richness in CDS and speed of development in CS, it is necessary to consider a number of cases with contrasting degrees of morphological richness in CDS. For this study, we have obtained data that display such properties by sampling a range of children acquiring typologically different languages. The language sample selected consists of six Indo-European languages (from four sub-families), one Finno-Ugric, one Turkic and one Mayan language. Typologically, all nine languages are suffixing languages.[3] However, among suffixing languages, they represent a great variety of morphological richness on the scale between the isolating language type (representing minimal morphological richness) and the agglutinating language type (representing maximal morphological richness (see Kilani-Schoch & Dressler, 2005; Sgall, 1999; Skalička, 1979): French, Dutch and German are weakly inflecting languages (with French showing the most isolating features); Russian, Croatian and Greek are strongly inflecting languages (with Russian showing the most inflecting-fusional features); Turkish, Finnish and Yucatec Maya are agglutinating languages (with Turkish showing the most agglutinating features).

Because of the prominent role played by nouns and verbs in early development (Bates et al., 1994; Bittner, Dressler, & Kilani-Schoch, 2003; Tomasello & Merriman, 1995; Voeikova & Dressler, 2002), we have restricted our investigation to these two categories. The inflectional categories of nouns and verbs are treated as separate *subsystems* for typological reasons: a single language can have a rather rich verb inflection but a poor noun inflection (Yucatec Maya, Greek and French are examples of this pattern), or the other way round, although the latter case does not occur in the languages of our sample (for further details see Dressler, 2005; Laaha & Gillis, 2007). A further justification for studying the development of noun and verb inflection separately lies in the debate of the noun bias in language acquisition. Based on the well-established fact that children's early lexicon is skewed towards nouns in several western languages (e.g. French: Bassano, Maillochon, & Eme, 1998; English: Bates et al., 1994; Italian: Caselli et al., 1995; Hebrew: Maital, Dromi, Sagi, & Bornstein, 2000; Dutch: Verlinden & Gillis, 1988), it has been hypothesized that a similar asymmetry between nouns and verbs might also be found in morphological development: children having at their disposal a larger 'critical mass' (Marchman & Bates, 1994; Marchman, Martinez-Sussmann, & Dale, 2004) of noun vocabulary than verb vocabulary should be likely to produce noun morphology earlier than verb morphology (e.g. Bassano, 2000; Tomasello, Akhtar, Dodson, & Rekau, 1997).

In line with the CDS studies presented in the section *The importance of variation in child-directed speech*, we argue that variation in CDS has a positive effect on children's early development of noun and verb inflection. Thus, we predict that morphological richness in CDS is positively correlated with speed of morphological development in CS. Put differently, in a language rich in morphology, what might at first appear as complexity could in fact be a facilitative factor due to its functionality.

Method

Participants

The participants of this study are nine children acquiring nine typologically different languages that vary in terms of morphological richness (see Table 1). The children were investigated from the child's onset of speech until the age of about 3 years. As the children represent nine different languages, they are members of different cultural-linguistic communities. Except for the Yucatec-speaking child, they all come from middle-class families where either one or both of the parents are highly educated professionals. Their conversations revolve around very similar activities in the here and now, playing with toys, looking at pictures in books, eating and carrying out other routine activities appropriate to the 2- to 3-year-old's world in western cultures. All children are monolingual and none of them shows developmental or linguistic problems.

Data collection

Each child was audio-recorded several times per month at his/her home, in unstructured settings, interacting with his/her mother or other caretakers. The data were transcribed in CHAT format and coded morphologically according to the norms of CHILDES (MacWhinney, 2000).

Table 1. Overview of the participants

Child	Origin	Language acquired	Sex	Birth order	SES	Age range
1	Zagreb, Croatia	Croatian (CRO)	Female	First-born	Middle	1;3−2;8
2	Antwerp, Flanders	Dutch (DU)	Female	First-born	Middle	1;5−2;5
3	Helsinki, Finland	Finnish (FINN)	Female	First-born	Middle	1;7−3;0
4	Lausanne, Switzerland	French (FR)	Female	Third-born	Middle	1;6−3;0
5	Vienna, Austria	German (GER)	Male	Second-born	Middle	1;3−2;6
6	Athens, Greece	Greek (GRK)	Male	First-born	Middle	1;7−2;6
7	St. Petersburg, Russia	Russian (RUS)	Male	First-born	Middle	1;4−2;8
8	Istanbul, Turkey	Turkish (TURK)	Female	First-born	Middle	1;3−2;0
9	Yalcobá, Mexico	Yucatec Maya (YUC)	Male	First-born	Low	2;0−3;0

Data sampling

Table 2 gives an overview of the data analysed in this study. For each of the nine child corpora, both CS and CDS data were analysed. The analysis of the CS data was conducted on monthly samples and an alignment procedure was applied prior to the analysis (hence the different age ranges reported in the columns *Original* and *Aligned* of Table 2).[4] The analysis of the CDS data was conducted on a single sample, by cumulating monthly samples. Finally, both CS and CDS data were split into nouns (including proper nouns) and verbs (including auxiliaries and modal verbs).[5]

Notice that the amount of CDS data analysed was not the same in the nine corpora: for three corpora (Greek, Turkish, Yucatec Maya), only reduced CDS samples of 900 input utterances each (300 at the beginning, 300 in the middle and 300 at the end of the observational period) were available. This is why the total number of tokens in CS exceeds the total number of tokens in CDS in these corpora. There was also considerable

Table 2. Overview of the data

Child	Lang.	Avg. # sessions per month	Avg. # minutes per month	# tokens				Age range	
				CDS		CS		Original	Aligned
				Nouns	Verbs	Nouns	Verbs		
1	CRO	3	135	5505	10,795	1799	4149	1;3−2;8	1;4−2;8
2	DU	1	60	4247	4362	3021	1278	1;5−2;5	1;5−2;5
3	FINN	2	40	2589	4329	1377	2420	1;7−3;0	1;7−2;9
4	FR	2	50	7484	13,774	6127	6655	1;6−3;0	1;6−3;0
5	GER	2	70	8053	13,984	4027	3121	1;3−2;6	1;5−2;6
6	GRK	4	80	733	927	3750	1620	1;7−2;6	1;8−2;6
7	RUS	2	120	9316	7394	4292	2162	1;4−2;8	1;4−2;8
8	TURK	2	25	973	1193	1162	1533	1;3−2;0	1;6−1;10
9	YUC	4	80	445	682	1439	1684	2;0−3;0	2;0−3;0

variation in the type of CDS across the nine corpora, for which, unfortunately, we could not control: in five corpora, it is exclusively the mother interacting with the child, in four corpora, the speech of other adult family members was analysed as well. In seven out of nine corpora, adult–adult interactions were included in the analysis, but they were very rare (except for the Yucatec data).

Measures

MSP in child-directed speech. In this study, the degree of morphological richness in CDS is defined as the normalized MSP over 1000 tokens in the CDS data of each corpus. This variable represents in effect the *mean number of inflected word-forms per lemma* (over 1000 tokens) in CDS. For some corpora, there were fewer than 1000 tokens of nouns (Greek, Turkish and Yucatec Maya) or verbs (Greek and Yucatec Maya) available in CDS. In these cases, we applied a logarithmic regression to compute the *expected* MSP for 1000 tokens. The decision to use a logarithmic model was based on the empirical observation by Xanthos and Gillis (2010) that the growth of MSP with regard to sample size is better represented by a logarithmic than a linear model or one based on a power transform.

Speed of development of MSP in child speech. In order to monitor the monthly development of morphological richness in the CS of a given corpus, we define the *cumulative MSP at month m* as the MSP over the whole corpus *up to month m*. Table 3 illustrates this definition on the basis of a corpus of three monthly samples. The use of a cumulative definition of MSP enables us to capture inflectional distinctions spanning different monthly samples. There are mainly two assumptions that underlie this approach: that the occurrence, in a child's productions, of a given form at some point in time counts as evidence for a change in the child's representations; and that the non-occurrence of that form at some later point should not be interpreted as evidence that the representations have reverted to an earlier state. In other words, cumulative MSP emphasizes an incremental view of the data, at the expense of disregarding their fluctuations over time.

  Cumulative MSP has the undesirable property of being evaluated over an ever increasing amount of data. In order to ensure that this increase remains comparable between corpora of different sizes (both in terms of number of tokens and in terms of number of months), the resampling procedure described in the section *Mean size of paradigm* was modified so that the *expected number of new tokens per month* is the same for all corpora (for details, see Xanthos & Laaha, 2007). For the present study, this

Table 3. The calculation of cumulative MSP

Month	Word-forms	Cumulative MSP
1	*has*	1/1 = 1
2	*have, having*	3/1 = 3
3	*am, are*	5/2 = 2.5

number was arbitrarily set to 50 tokens, and we call the resulting measure the *normalized cumulative MSP over 50 tokens*, or cumulative MSP(50).

The next step was to calculate the *speed of development* of cumulative MSP. Since this calculation is performed on a monthly basis, the speed of development at a given month is simply the difference between the cumulative MSP at this month and the cumulative MSP at the previous month. In the case of the data given in Table 2, the speed at month 2 is $3 - 1 = 2$, and the speed at month 3 is $2.5 - 3 = -0.5$. The negative value reflects the *regression* in cumulative MSP at month 3; this is due to the fact that there is an increase in *lexical* richness at this point (namely the occurrence of the new lemma BE) which is not matched by a corresponding increase in morphological richness (the new lemma has only 2 inflected word-forms, as against 3 for HAVE).

For each corpus, we have calculated the speed of development of cumulative MSP(50) in CS for each month, as well as the average of this value over all months. The latter represents our overall measure of speed of development of morphological richness in CS, i.e. the *mean number of new inflected word-forms per lemma and per month* (over 50 new tokens).

Results

MSP in child-directed speech

The values of the variable MSP(1000) in child-directed speech (CDS) for nouns and verbs are represented in Figure 1. A first general remark is that in CDS the MSP of verbs is higher than that of nouns for all the corpora of our sample. The results also make it
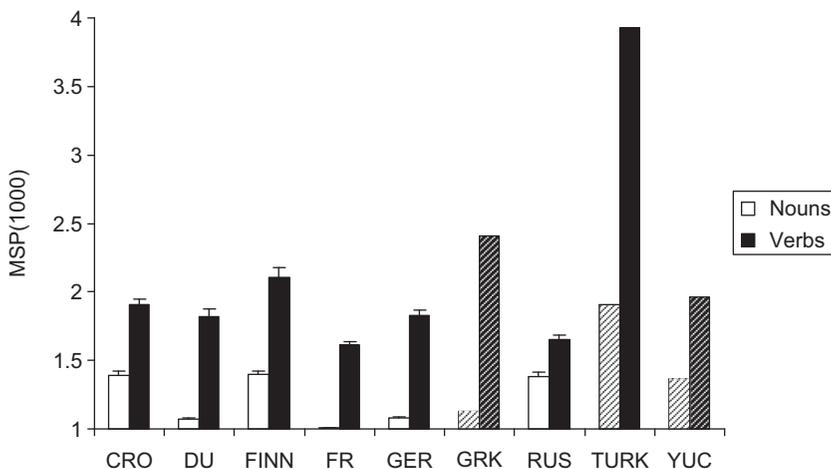
Figure 1. MSP(1000) in child-directed speech. Hatched bars denote values predicted by a logarithmic regression (see text). Error bars represent 90% confidence intervals, which cannot be estimated for predicted values; in the case of Turkish verbs, they are too small to be visible.

clear that caregivers are far from exhausting the inflectional potential of nouns and verbs in their speech addressed to young children. For instance, in Croatian, a verb can in principle distinguish between as many as 18 distinct forms (Stephany et al., 2007). Yet, in our data, the MSP for this subsystem is only 1.91. This discrepancy between theoretical and observed morphological richness can only be explained by the fact that most lemmas generally occur in only one or two inflected forms in CDS, with other forms being only marginally represented.

The corpora of our sample appear to fall essentially into three groups, based on their MSP values for nouns. The Turkish corpus exhibits the highest MSP (1.91) and therefore belongs in a group of its own. The Croatian, Finnish, Russian and Yucatec Mayan corpora form a second group with an MSP of about 1.4. The third group consists of the Dutch, French, German and Greek corpora, which display little or no morphological richness as measured by MSP (ranging between 1.01 and 1.13).

The MSP values for verb inflection in CDS induce a different ordering of the corpora. The Turkish corpus again stands out with the highest MSP value (3.93), followed by the Greek and Finnish corpora, similarly isolated in the second and third position (2.41 and 2.11, respectively). The Croatian, Dutch, German and Yucatec Mayan corpora fall into a single group, with MSP values ranging between 1.82 and 1.96. Finally, the French and Russian corpora constitute the group with the lowest MSP (about 1.6).

Speed of development of MSP in child speech

The development of cumulative MSP(50) in child speech (CS) for nouns and verbs is plotted in Figure 2. As predicted by the hypothesis that morphological richness in CDS is positively correlated with the speed of development of morphological richness in CS, the curves of MSP in CS rise more steeply for verbs than for nouns in all the corpora of our sample.

Note that in spite of this visibly higher speed of development of verb inflection as compared to noun inflection, the MSP in the CS of each corpus (but Russian) gets closer to the corresponding MSP in CDS for nouns than it does for verbs: on average, the ratio of the MSP in the last sample of CS to the MSP in CDS is 0.94 for nouns and 0.88 for verbs.

The mean speed of development of cumulative MSP(50) in CS is represented in Figure 3. These values are in remarkable correspondence with the respective values of MSP(1000) in CDS (see Figure 1). For nouns, we find again a clustering into three groups: the speed of development for the Dutch-, French- and German-speaking children is lower than 0.01; it ranges between 0.01 and 0.02 for the Croatian-, Finnish-, Greek-, Russian- and Yucatec-speaking children; finally, the Turkish child reaches an unequalled rate of about 0.07.

For verbs, we find again that the Turkish- (0.31) and Greek-speaking children (0.11) are isolated in the first and second position, as predicted by the corresponding MSP in CDS. Speed of development for the Finnish child scores a bit lower than expected; together with the Croatian-, Dutch-, German- and Yucatec-speaking children, she belongs to a group reaching an approximate speed of 0.05. The French- and Russian-speaking children follow with a speed of about 0.03.
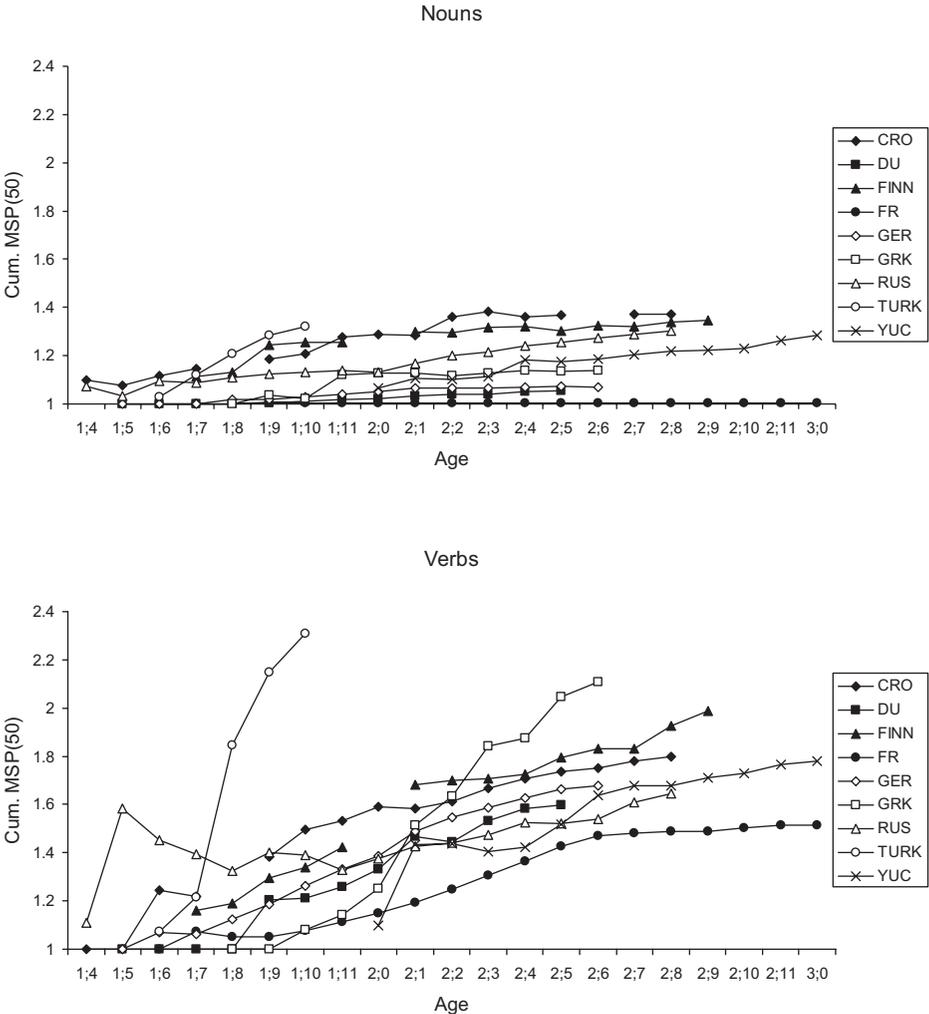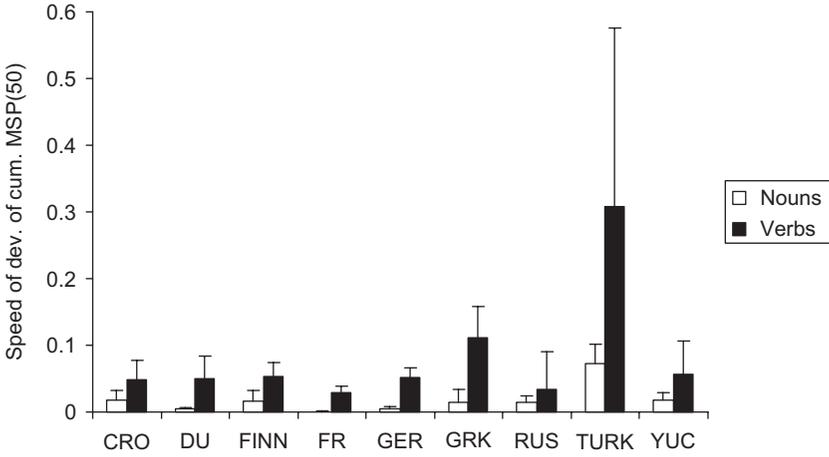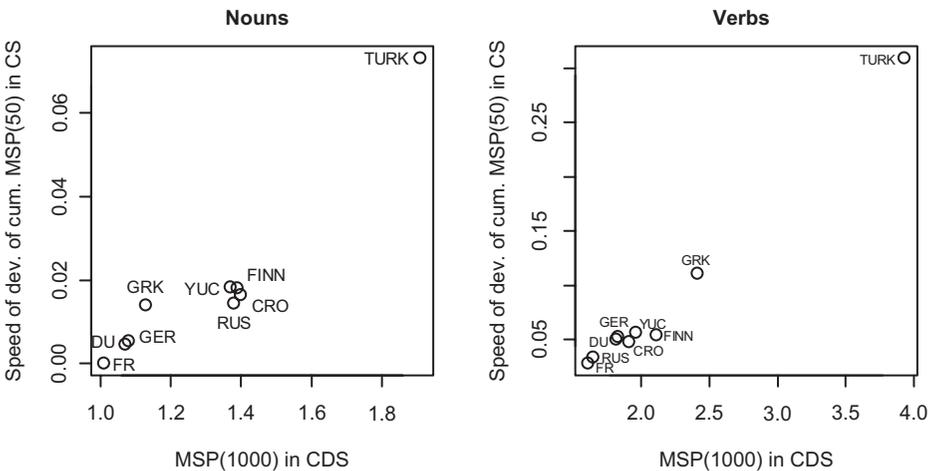
Nouns

Verbs

Figure 2. Development of cumulative MSP(50) in child speech for nouns and for verbs

Correlation of MSP in child-directed speech with speed of development of MSP in child speech

The scatter plots in Figure 4 show the relationship between MSP in child-directed speech and speed of development of MSP in child speech, for nouns and for verbs. In general, speed of development is remarkably well predicted by MSP in CDS; the only exceptions are Greek nouns, for which speed of development is higher than predicted by the MSP in CDS, and Finnish verbs, for which it is lower. Second, the plots clearly point out the exceptional character of the Turkish corpus in our sample, insofar as the values for both variables in noun as well as verb inflection are of another order of magnitude than those

Figure 3. Mean speed of development of cumulative MSP(50) in child speech for nouns and verbs. Error bars represent 90% confidence intervals

Figure 4. MSP(1000) in child-directed speech vs. speed of development of cumulative MSP(50) in child speech for nouns and verbs

of the other corpora. For verbs, this observation must be qualified by reference to the massive confidence interval for this case; however, it can be partly explained by the spurt observed in this child's data, which entails large differences in speed of development over consecutive months (see Figure 2).

In the presence of such an outlier, Pearson's correlation coefficient is likely to be overestimated. Therefore, we have used Spearman's *rank* correlation to assess the strength of the relationship between MSP in CDS and speed of development of MSP in

CS. In spite of the fact that the analysis bears only on 9 data points, the results are unequivocal: there is a very high positive correlation both for nouns ($\rho = .88$, $p = .002$) and for verbs ($\rho = .93$, $p < .001$).

Discussion

The aim of the present article was to investigate the relationship between the morphological richness of child-directed speech and the speed of morphological development in children's language. More specifically, we set out to study the morphological diversity of inflectional systems. For this purpose, a new metric for measuring inflectional diversity was used, viz. normalized *mean size of paradigm* (MSP). In essence, MSP measures the average number of inflected forms per lemma attested in a language sample. As such, MSP was used to assess the inflectional richness of speech addressed to children acquiring nine different languages, and to assess those children's speed in acquiring the inflectional system of their language.

Our results show that the morphological richness of an inflectional system, as represented in child-directed speech (CDS), has a strong positive correlation with the speed of development of morphological richness in child speech (CS). Correlations were tested separately for noun and verb inflection in children acquiring nine languages representing different degrees of morphological richness. They were found to be significant in both cases. This result is in line with other studies which stress the importance of *variation* in CDS for children's early language development (see *Introduction*). Although early exposure to a variety of inflectional forms may seem to complicate the learning task for the child, it may help children exposed to a richly inflected input to focus more on different forms and on differences in meaning expressed by inflectional means than children exposed to a less richly inflected input.

At a more general level, this study is another contribution to the growing body of research showing children's sensitivity to the typological properties of the language they acquire as reflected in CDS (e.g. Aksu-Koç & Slobin, 1985; Bavin, 1998; Bowerman & Choi, 2001; Gillis & Ravid, 2006; Morgan, Barrière, & Woll, 2006; Gathercole, 2006; Peters, 1997; Shirai, 1998; Slobin, 1997, 2001; Wittek & Tomasello, 2005). The particularity of our approach has been to examine the relationship between a specific dimension of linguistic *complexity*, namely paradigmatic morphological richness, and the *speed of development* along this dimension.

The separate treatment of noun and verb inflection in this study was justified by typological considerations: these subsystems can have very different degrees of morphological richness within the same language. This methodological design reveals an interesting developmental contrast: in all the corpora of our sample, the speed of development of morphological richness in CS is higher for verbs than for nouns, in accordance with the systematically higher degree of morphological richness in CDS for verbs than for nouns. This sheds new light on the question of whether the 'noun bias' in lexical acquisition extends to morphological acquisition. Our results suggest that, as far as *speed of development* of noun and verb inflection is concerned, the determining factor is the morphological richness of the corresponding subsystem in CDS; in the nine corpora of our sample, this translates into a *grammatical verb bias* (as opposed to a *lexical noun bias*) in early language development.

This study was based on a quantitative typological framework. Following Hempel and Oppenheim (1936), we think that such models are at an advantage compared to classificatory approaches, which divide languages into agglutinating, inflecting, etc. In our model, the languages of our sample are mapped onto a single dimension with Turkish, the strongest agglutinating language, at one end, and weakly inflecting languages at the other end. This configuration, which is observed both in CDS and in the development of CS, is compatible with the classificatory view of typology. Near the middle of the scale, however, the quantitative and classificatory accounts differ: the former finds that strongly inflecting languages can have an equal or even higher degree of morphological richness than certain agglutinating languages. Thus, we have noted that the Croatian and Yucatec corpora (as well as the Dutch and German ones) have comparable MSP values for verbs in CDS. By contrast, in the typological model of Skalička (1979), verb inflection approaches the ideal inflecting-fusional type in Croatian and the ideal agglutinating type in Yucatec. This discrepancy can be partly explained by our limitation to the *paradigmatic* dimension of morphological richness. To some extent, the distinction between these types pertains to the *syntagmatic* dimension of morphological richness. Indeed, the possibility of accumulating a larger number of inflectional affixes in a single word-form is characteristic of the agglutinating as compared to the inflecting-fusional type. This illustrates the need for simultaneous investigations of several morphological characteristics, including not only paradigmatic and syntagmatic morphological richness, but also transparency, salience, etc. (see Laaha & Gillis, 2007).

In future work, we plan to use the methodology developed in this article to study the correlation between syntagmatic morphological richness in CDS and speed of development of syntagmatic richness in CS, as well as the relation between syntagmatic and paradigmatic morphological richness. The proposed methodology could also be extended to the study of caretakers' fine-tuning of CDS to the child's course of linguistic development (see Ketrez, 2003; Snow, 1989, 1995; Snow & Ferguson, 1977). The extent to which this phenomenon can be observed in the domain of inflectional morphology could be assessed by correlating the cumulative MSP in CDS and CS. Another issue left for future research is the degree of inter-individual variation of MSP both in CS and CDS. Laaha and Gillis's (2007) findings suggest that while there is some variation in speed of development of MSP among children learning the same language, MSP in CDS is fairly stable. Assessing the generality of our typological considerations will require the study of more cases per language, as well as the inclusion of more diverse languages, such as prefixing and transfixing ones (which would imply to control for the intervening variable of positional salience).

Our interpretation of the results of this study is that the major or minor role played by noun and verb inflection in CDS induces the child to put more or less emphasis on the acquisition of noun vs. verb inflection. Although this kind of sensitivity to typological features of the input has been established before, the present study has advanced our knowledge about this issue by quantifying paradigmatic morphological richness in CDS as well as its speed of development in CS based on longitudinal corpora in nine typologically different languages.

Acknowledgements

Funding

Notes

1   See Laaha and Gillis (2007) for a fuller account including syntagmatic richness as well as other typological variables such as transparency and salience.
2   Note that we make no distinction between *have* as a 1st person singular present and as an infinitive, for instance; they are counted as instances of a single word-form. On the other hand, in cases where homophonous word-forms correspond to distinct lemmas (e.g. *a drink* vs. *I drink*), their occurrences are counted separately; thus, in this example, they would contribute either to the inflectional diversity of the noun lemma or to the diversity of the verb lemma.
3   Note that it is not straightforward to compare the morphological richness of suffixing languages with prefixing languages and transfixing ones (e.g. Arabic *katab-a* 'to write', *kita:b* 'book', *ka:tib* 'writer'), because the positional salience of suffixes, prefixes and transfixes (or vowel patterns transfixing the consonantal root of Arabic or Hebrew) is different and needs to be controlled for.
4   The alignment procedure was based on the *proportion of utterances with verbs* (PUV). The principle was to remove monthly samples at the beginning (respectively at the end) of each corpus when their PUV was lower than (respectively exceeded) a given threshold. The lower and higher thresholds were set to be the 5th and 95th percentiles of PUV, as calculated over all monthly samples (details can be found in Xanthos & Laaha, 2007).
5.   Since we have restricted our investigation to synthetic morphology, periphrastic perfects such as (*they*) *have walk-ed* are treated as independent instances of the auxiliary *have* and the perfect participle *walked*. Diminutives/hypocoristics and compound nouns are counted as distinct noun lemmas. Prefix verbs are also counted as distinct verb lemmas, the only exceptions being Dutch and German separable particle verbs sharing the same base (e.g. Ger. *auf-mach-en* 'to open', *zu-mach-en* 'to close' vs. 3SG present *mach-t auf*, *mach-t zu*), which are treated as a single verb lemma.

References

Aksu-Koç, A. (1998). The role of input vs. universal predispositions in the emergence of tense-aspect morphology: Evidence from Turkish. *First Language, 18*, 255–280.

Aksu-Koç, A., & Slobin, D. (1985). The acquisition of Turkish. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 1, pp. 839–878). Hillsdale, NJ: Lawrence Erlbaum.

Bassano, D. (2000). Early development of nouns and verbs in French: Exploring the interface between lexicon and grammar. *Journal of Child Language, 27*, 521–559.

Bassano, D., Maillochon, I., & Eme, E. (1998). Developmental changes and variability in the early lexicon: A study of French children's naturalistic productions. *Journal of Child Language, 25*, 493–531.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J., et al. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language, 21*, 85–123.

Bavin, E. (1998). Factors of typology on language acquisition: Some examples from Warlpiri. In A. Siwierska & J. Song (Eds.), *Case, typology and grammar* (pp. 37–56). Amsterdam: John Benjamins.

Berman, R. (2004). Between emergence and mastery: The long developmental route of language acquisition. In R. Berman (Ed.), *Language development across childhood and adolescence: Psycholinguistic and crosslinguistic perspectives* (pp. 9–34). Amsterdam: John Benjamins.

Bittner, D., Dressler, W., & Kilani-Schoch, M. (2003). *Development of verb inflection in first language acquisition: A cross-linguistic perspective*. Berlin: Mouton de Gruyter.

Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475–511). Cambridge: Cambridge University Press.

Brodsky, P., Waterfall, H., & Edelman, S. (2007). Characterizing motherese: On the computational structure of child-directed language. In D. S. McNamara & J. G. Trafton (Eds), *Proceedings of the 29th Meeting of the Cognitive Science Society* (pp. 833–838). Austin, TX: Cognitive Science Society.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., et al. (1995). A crosslinguistic study of early lexical development. *Cognitive Development, 10*, 159–199.

Caselli, M. C., Casadio P., & Bates E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language, 26*, 69–111.

Choi, S. (2000). Caregiver input in English and Korean: Use of nouns and verbs in book-reading and toy play contexts. *Journal of Child Language, 27*, 69–111.

Comrie, B. (1981). *Language universals and linguistic typology: Syntax and morphology*. Oxford: Blackwell.

Devescovi, A., Caselli, M. C., Marchioni, D., Pasqualetti, P., Reilly, J., & Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language, 32*, 759–786.

Dressler, W. (2004). Degrees of grammatical productivity in inflectional morphology. *Italian Journal of Linguistics, 15*, 31–62.

Dressler, W. (2005). Morphological typology and first language acquisition: Some mutual challenges. In G. Booij, E. Guevara, A. Ralli, S. Sgroi, & S. Scalise (Eds.), *Morphology and linguistic typology* (pp. 7–20). Bologna: Università degli studi di Bologna.

Gathercole, V. C. M (2006). Introduction to the special issue: Language-specific influences on acquisition and cognition. *First Language, 26*, 5–17.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.), *Language development, Vol. II: Language, thought and culture* (pp. 301–334). Hillsdale, NJ: Lawrence Erlbaum.

Gillis, S., & Ravid, D. (2006). Typological effects on spelling development: A crosslinguistic study of Hebrew and Dutch. *Journal of Child Language, 33*, 621–659.

Goldfield, B. (1993). Noun bias in maternal speech to one-year-olds. *Journal of Child Language, 20*, 85–99.

Greenberg, J. (1954/1960). A quantitative approach to the morphological typology of language. *IJAL, 26*, 178–194.

Hawkins, J. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hempel, C., & Oppenheim, P. (1936). *Der Typenbegriff im Lichte der neuen Logik*. Leiden: Sijthoff.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*, 55–88.

Hoff-Ginsberg, E. (1985). Relations between discourse properties of mothers' speech and their children's syntactic growth. *Journal of Child Language, 12*, 367–385.

Johnson, W. (1944). Studies in language behaviour: I. A program of research. *Psychological Monographs, 56*, 1–15.

Ketrez, N. (2003). Variation in Turkish mothers' style. In A. Özsoy, D. Akar, M. Nakipoğlu-Demiralp, E. Erguvanlı-Taylan, & A. Aksu-Koç (Eds.), *Studies in Turkish linguistics* (pp. 447–452). Istanbul: Boğaziçi University Press.

Kilani-Schoch, M., & Dressler, W. (2005). *Morphologie naturelle et flexion du verbe français*. Tübingen: Narr.

Küntay, 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* (pp. 265–286). Mahwah, NJ: Lawrence Erlbaum.

Küntay, A., & Slobin, D. (2001). Discourse behavior of lexical categories in Turkish child-directed speech: Nouns vs. verbs. In M. Almgren, A. Barreña, M. Ezeizabarrena, I. Idiazabal, & B. MacWhinney (Eds.), *Research on child language acquisition: Proceedings for the 8th Conference of the International Association for the Study of Child Language* (pp. 928–946). Somerville, MA: Cascadilla Press.

Laaha, S. (2004). *Développement précoce de la morphologie verbale: une étude comparative sur l'acquisition de l'allemand autrichien et du français.* Unpublished PhD, University of Vienna – University Paris 5.

Laaha, S., & Gillis, S. (Eds.). (2007). *Typological perspectives on the acquisition of noun and verb morphology.* Antwerp Papers in Linguistics 112. Antwerp: University of Antwerp.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (2 vols). Mahwah, NJ: Lawrence Erlbaum.

Maital, S., Dromi, E., Sagi, A., & Bornstein, M. (2000). The Hebrew communicative development inventory: Language-specific properties and cross-linguistic generalizations. *Journal of Child Language, 27*, 43–67.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke: Palgrave Macmillan.

Marchman, V., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21*, 339–366.

Marchman, V., Martinez-Sussman, C., & Dale, P. (2004). The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental Science, 7*, 212–224.

McWhorter, J. (2005). *Defining creole*. Oxford: Oxford University Press.

Miestamo, M., Sinnemäki, K., & Karlsson, F. (Eds.). (2008). *Language complexity: Typology, contact, change*. Amsterdam: John Benjamins.

Morgan, G., Barrière, I., & Woll, B. (2006). The influence of typology and modality on the acquisition of verb agreement morphology in British Sign Language. *First Language, 26*, 19–43.

Naigles, L., & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *Journal of Child Language, 25*, 95–120.

Ogura, T., Dale, P., Yamashita, Y., Murase, T., & Mahieu, A. (2006). The use of nouns and verbs by Japanese children and their caregivers in book-reading and toy-playing contexts. *Journal of Child Language, 33*, 1–29.

Peters, A. (1997). Language typology, prosody, and the acquisition of grammatical morphemes. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 5, pp. 135–197). Mahwah, NJ: Lawrence Erlbaum.

Pine, J. (1994). The language of primary caregivers. In C. Gallaway & B. Richards (Eds.), *Input and interaction in language acquisition* (pp. 15–37). Cambridge: Cambridge University Press.

Sgall, P. (1999). Prague school typology. In M. Shibatani & T. Bynon (Eds.), *Approaches to language typology* (pp. 49–84). Oxford: Oxford University Press.

Shirai, Y. (1998). The emergence of tense-aspect morphology in Japanese: Universal predisposition? *First Language, 18*, 281–309.

Skalička, V. (1979). *Typologische Studien*. Braunschweig: Vieweg.

Slobin, D. (1997). The universal, the typological, and the particular in acquisition. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 5, pp. 1–39). Mahwah, NJ: Lawrence Erlbaum.

Slobin, D. (2001). Form–function relations: How do children find out what they are? In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 406–449). Cambridge: Cambridge University Press.

Snow, C. (1972). Mother's speech to children learning language. *Child Development, 43*, 549–565.

Snow, C. (1986). Conversations with children. In P. Fletcher & M. Garman (Eds.), *Language acquisition: Studies in first language development* (pp. 69–89). New York: Cambridge University Press.

Snow, C. (1989). Understanding social interaction and language acquisition: Sentences are not enough. In M. Bornstein & J. Bruner (Eds.), *Interaction in human development* (pp. 83–103). Hillsdale, NJ: Lawrence Erlbaum.

Snow, C. (1995). Issues in the study of input: Finetuning, universality, individual and developmental differences, and necessary causes. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 180–193). Oxford: Blackwell.

Snow, C., & Ferguson, C. (Eds.). (1977). *Talking to children: Language input and acquisition*. Cambridge: Cambridge University Press.

Stephany, U. (1985). *Aspekt, Tempus und Modalität: zur Entwicklung der Verbalgrammatik in der neugriechischen Kinderprache*. Tübingen: Narr.

Stephany, U. (1997). The acquisition of Greek. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 4, pp. 183–333). Mahwah, NJ: Lawrence Erlbaum.

Stephany, U., Voeikova, M., Christofidou, A., Gagarina, N., Kovačević, M., Palmović, M., et al. (2007). Strongly inflecting languages: Russian, Croatian, and Greek. In S. Laaha & S. Gillis (Eds.), *Typological perspectives on the acquisition of noun and verb morphology*. Antwerp Papers in Linguistics 112 (pp. 35–46). Antwerp: University of Antwerp.

Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns vs. verbs: A comparison of English, Italian and Mandarin. *Journal of Child Language, 24*, 535–565.

Tare, M., Shatz, M., & Gilbertson, L. (2008). Maternal uses of non-object terms in child-directed speech: Color, number and time. *First Language, 28*, 87–100.

Tomasello, M., Akhtar, N., Dodson, K., & Rekau, L. (1997). Differential productivity in young children's use of nouns and verbs. *Journal of Child Language, 24*, 373–387.

Tomasello, M., & Merriman, W. (Eds.). (1995). *Beyond names for things: Young children's acquisition of verbs*. Hillsdale, NJ: Lawrence Erlbaum.

Verlinden, A., & Gillis, S. (1988). Nouns and verbs in the input: Gentner (1982) reconsidered. *ABLA Papers, 12*, 163–187.

Voeikova, M., & Dressler, W. (2002). *Pre- and protomorphology: Early phases of morphological development in nouns and verbs*. Munich: Lincom.

Wijnen, F., Kempen, M., & Gillis, S. (2001). Bare infinitives in Dutch early child language: An effect of input? *Journal of Child Language, 28*, 629–660.

Wittek, A., & Tomasello, M. (2005). German-speaking children's productivity with syntactic constructions and case morphology: Local cues act locally. *First Language, 25*, 103–125.

Xanthos, A., & Gillis, S. (2010). Quantifying the development of inflectional diversity. *First Language, 30*, 175–198.

Xanthos, A., & Laaha, S. (2007). Method. In S. Laaha & S. Gillis (Eds.), *Typological perspectives on the acquisition of noun and verb morphology*. Antwerp Papers in Linguistics 112. (pp. 11–20). Antwerp: University of Antwerp.