ISSN 0776-3859

ANTWERP PAPERS IN LINGUISTICS

112

Typological perspectives on the acquisition of noun and verb morphology

Sabine Laaha & Steven Gillis (eds.)

2007

Contents

Preface

In this issue of *Antwerp Papers in Linguistics* we present recent results from the "Cross-linguistic Project on Pre- and Protomorphology in Language Acquisition". This project was founded in 1994 by Wolfgang U. Dressler (Austrian Academy of Sciences and University of Vienna) and aims at investigating the early phases of morphological development in a large sample of different languages that show important typological variation. The current volume is a collaborative effort of 21 project members studying 11 different languages:

Croatian:	G. Hrzica, M. Kovačević, M. Palmović
Dutch:	S. Gillis
Finnish:	K. Laalo
French:	M. Kilani-Schoch, A. Xanthos
German:	W. U. Dressler, K. Korecky-Kröll, S. Laaha
Greek:	A. Christofidou, U. Stephany, E. Thomadaki
Hungarian:	V. Barcza, P. Bodor
Lithuanian:	I. Savickienė
Russian:	N. Gagarina, M. D. Voeikova
Turkish:	A. Aksu-Koç, N. F. Ketrez
Yucatec Maya:	B. Pfeiler

This volume is divided into two parts: in the first part a quantitative picture is drawn of how fast nominal and verbal morphological structures are acquired, and in the second part the acquisition of diminutives is studied. A leading theme of the project that is developed in both parts of this volume, is how typological diversity influences children's acquisition of their mother tongue. A finding that has become increasingly prominent is that a system that looks quite intricate and complex at first sight, need not be hard or difficult to acquire, or need not take a lot of time to master. And vice versa: prima facie 'easy' morphological patterns are not necessarily acquired very early in development, and may take relatively more time to master.

In Part 1, the early development of nominal and verbal morphology in nine typologically different languages is systematically compared according to a common design. The main hypothesis is that "the richer noun or verb morphology is in the input, the more stimulated the child will be to develop noun or verb morphology rapidly" (Dressler, this volume, p. 8), i.e. that there exists a proportionality relationship between the degree of morphological richness of the input and the speed of development of morphological richness in child speech. In addition, factors such as transparency, uniformity and salience should show an impact on speed of development. Chapter 1 (by Dressler) provides the theoretical introduction. Chapter 2 (by Xanthos and Laaha) presents the method used in this research. Chapters 3 to 5 give detailed results for the three language groups studied: weakly inflecting languages (Chapter 3 by Laaha, Gillis, Kilani-Schoch, Korecky-Kröll, Xanthos and Dressler), strongly inflecting languages (Chapter 4 by Stephany, Voeikova, Christofidou, Gagarina, Kovačević, Palmović and Hrzica), and agglutinating languages (Chapter 5 by Aksu-Koç, Ketrez, Laalo and Pfeiler). Chapter 6 (by Xanthos) presents the general quantitative results. These are followed in Chapter 7 (by Dressler, Stephany, Aksu-Koç and Gillis) by a general discussion of the results and a conclusion of this research.

In Part 2, the paper by Savickienė, Dressler, Barcza, Bodor, Ketrez, Korecky-Kröll, Palmović, Stephany, and Thomadaki presents a cross-linguistic study of the development of diminutives in six different languages. The authors investigate the development of the semantics and pragmatics of diminutives and show to which extent the factors morphological productivity, transparency and salience favour the acquisition of diminutives.

An earlier version of the contributions to this volume was presented at the 10th International Congress for the Study of Child Language, which was held in July 2005, in Berlin. The work in Part 1 of this volume was presented in form of a symposium, the work in Part 2 was presented in form of an individual paper.

We would like to take this opportunity to thank Wolfgang U. Dressler (alias Ulli) for the enormous effort he put into this project over the years. We also thank all project members for their kind cooperation and for the amount of work that they invested into this publication.

We gratefully acknowledge the financial support of the Austrian Academy of Sciences – without this support the present publication would not have been possible.

Sabine Laaha and Steven Gillis

Vienna/Antwerp, March 2007

PART I

Early development of nominal and verbal morphology from a typological perspective

1. Introduction

Wolfgang U. Dressler[*]

In this joint research, the early development of nominal and verbal morphology in nine typologically different languages will be systematically compared according to a common design. This research is another result of the "Cross-linguistic Project on Pre- and Protomorphology in Language Acquisition" (see Dressler, 1997; Gillis, 1998; Voeikova & Dressler, 2002; Bittner et al., 2003). Work done within this project is not simply cross-linguistic, but at the same time typological, insofar language types are referred to as constellations of typologically relevant linguistic properties. The most important typological property of a language is its degree of (inflectional) morphological richness, defined as its number of morphological categories and patterns in inflection (Dressler, 2004). The present research focuses on morphological form, thus temporarily referring to morphological meaning and function only indirectly. The main hypothesis is that the richer the morphology of the language they are acquiring is, the faster children will develop morphology. The range of typological comparison is defined by the subsystems of noun and verb inflection, which demonstrate cross-linguistic correspondences between the structures of the languages compared and between the functions of items within these structures (see Jucquois, 1993), as they occur in the input. The development will not be pursued up to mastery, but mainly from emergence up to (and including) what Berman (2004: 13) has called the second, grammaticised phase, a phase which is structure-dependent and possibly rule-bound, but not yet fully conventionalised.

1.1. Three different approaches to language typology

Since we present a typological approach to language acquisition, a brief introduction to linguistic typology, as used here, is in order. In following Rudolf Carnap's general epistemological hierarchy, the founders of deductive-nomological epistemology, Hempel and Oppenheim (1936), have distinguished three epistemological levels of typology: classificatory, ordering and quantitative typology.

(I). The epistemologically lowest level is *classificatory typology*, which assigns each language to one specific linguistic type. As far as the sample of languages studied in the present research are concerned, this classification leads to the three classes of the agglutinating languages Turkish, Finnish, Yucatec, the strongly inflecting(-fusional) languages Russian, Croatian, Greek, and the weakly inflecting(-fusional) languages French, Dutch and German. Although this type of classical morphological typology dating from the 19th century is to be considered as presently outdated (see Lehmann, 1983; Ramat, 1995), these labels are maintained for convenience as a kind of shortcut and represent a first approximation. This holds both for the level of abstract systems of languages and, what is relevant for the present study, for the different types of inflectional morphology as these are represented in the input to young children.

(II). A second and higher level of linguistic typology, which is still relevant today, is represented by *ordering typology*. On this level, languages or rather linguistic subsystems are ordered according to the degree to which they approach the ideal constructs of morphological types (as defined by Skalička, 1979; see also Sgall, 1993, 1995 and further elaboration in Kilani-Schoch & Dressler, 2005). The three ideal types of interest in the present study are the

[*] I have to thank all project members for their suggestions, particularly Ursula Stephany.

agglutinating, the inflecting(-fusional) and the isolating type. Within the three agglutinating languages studied in this research, Turkish approaches the ideal agglutinating type most closely, whereas the two other languages also approximate the inflecting type. While Finnish is more agglutinating in the domain of noun inflection, Yucatec is so in verb inflection (see Jucquois, 1993). The ideal agglutinating type has a very rich morphology, richer than the ideal inflecting type. It is also more transparent than the latter and marks functions in a uniform, biunique way.

Skalička's (1979) view that the inflectional and word formation components of a language may behave differently, as far as typology is concerned, must be extended to the subcomponents or submodules of inflectional morphology itself. Thus, noun and verb inflection may have a different typological character within one and the same language and develop diachronically in typologically different directions (as has been the case for French, which developed from the mainly inflecting type of Latin towards the isolating type much more in noun than in verb morphology.[1] The interest apparent in studies of first language acquisition, which contrast the lexical and morphological development of nouns and verbs (see Lieven, 1997; Gentner & Boroditsky, 2001; Klampfer(-Laaha) & Korecky-Kröll, 2002; Pfeiler, 2002), is thus further justified by typological considerations.

Morphological richness in the inflecting type differs from the agglutinating type by cumulative marking, where one affix signifies several functions at the same time, as in the following examples from Turkish as opposed to Russian, which demonstrate the marking of the nominative, genitive and locative cases in both the singular and plural (Examples 1a and 1b):

(1)

a. Turkish

	NOM 'room'	GEN 'of a/the room'	LOC 'in a/the room'
SG	*oda*	*oda-(n)ın*	*oda-da*
PL	*oda-lar*	*oda-lar-ın*	*oda-lar-da*

b. Russian

	NOM 'room'	GEN 'of a/the room'	LOC 'in a/the room'
SG	*komnat-a*	*komnat-y*	*(v) komnat-e*
PL	*komnat-y*	*komnat*	*(v) komnat-ax*

While Turkish has no suffix for the semantically unmarked singular number and for the semantically unmarked nominative case, Russian marks the semantically unmarked Nom.Sg. by the suffix *-a* but leaves the semantically marked Gen.Pl. unmarked. Moreover, case and number marking in Turkish may be characterized in the following way: one and the same case suffix is attached to either singular or plural forms. The plural is always separately signalled by the same suffix. Since each suffix has just one meaning, more than one suffix is needed if a word is marked for both plural and case, as in the Gen. and Loc.Pl. (with the plural suffix preceding case). Russian never has more than one suffix, but each suffix is functionally cumulative in expressing number, case and grammatical gender at the same time; thus, the suffix *-a* in *komnata* signals nominative, singular and (by default) feminine gender at the same time.

The strongly inflecting languages of our sample approach the ideal inflecting type fairly well, Croatian more so than Russian and Greek. It is interesting to note that even the strongly inflecting languages of our sample may to a small extent carry isolating traits, e.g. Russian in

[1] see the analogous typological change of Bulgarian and Macedonian.

Example (1) above by using a function word, the preposition *v* 'in', for expressing locative meaning (analogously Croatian and Greek).

The ideal isolating type has no inflection and expresses grammatical functions and meanings by syntax and free function words such as articles, subject pronouns, auxiliary verbs, etc. As is to be expected, the weakly inflecting languages approach the isolating type more than the strongly inflecting ones do. Thus, the weakly inflecting languages of our sample (French, Dutch and German) have articles (of the strongly inflecting ones, only Greek), subject pronouns (whereas the strongly inflecting ones are pro-drop languages), and auxiliary verbs in past tense formation. This isolating tendency is strongest in French noun inflection, whereas Croatian is the most inflecting language. Rank orders can be seen in Table 1:

Table 1. Rank orders of ordering typology: noun vs. verb inflection

	Noun inflection	Verb inflection
Agglutinating languages	Turkish > Finnish > Yucatec	Turkish > Yucatec > Finnish
Strongly inflecting languages	Croatian > Russian > Greek	Croatian > Greek > Russian
Weakly inflecting languages	German > Dutch >>> French	German > Dutch > French

(III). The epistemologically highest level of typology is *quantitative typology*. This is a parameterized approach to typology in which, for each parameter, a given class of items is quantitatively compared across languages.

1.2. Typological parameters investigated

The typological parameters quantitatively studied in this research are paradigmatic and syntagmatic morphological richness, morphological transparency, uniformity and salience, which will mainly be illustrated with English examples for convenience (for methodological details, see Xanthos & Laaha, this volume).

A. Morphological richness

A1. Paradigmatic morphological richness

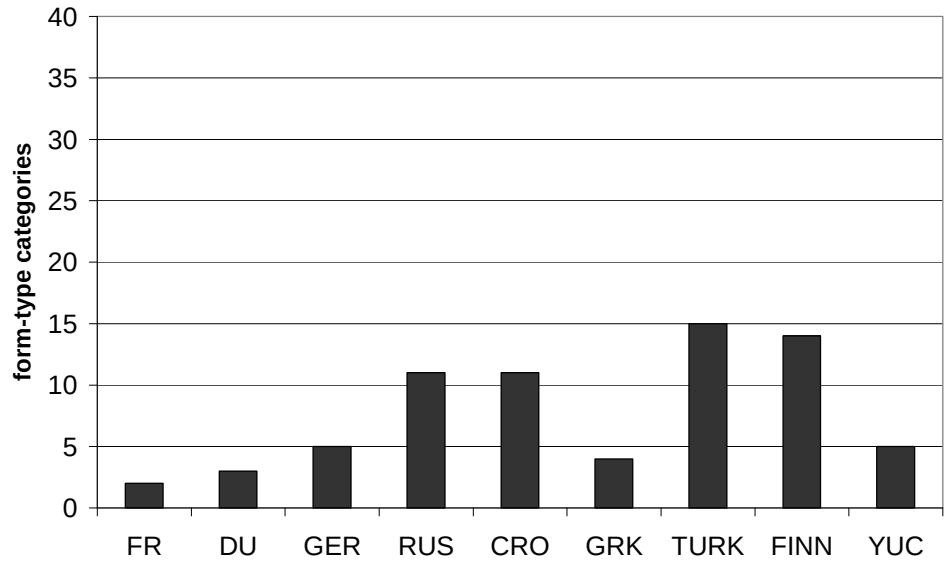

Figure 1a. Number of form-type categories in the input of the 9 languages: nouns

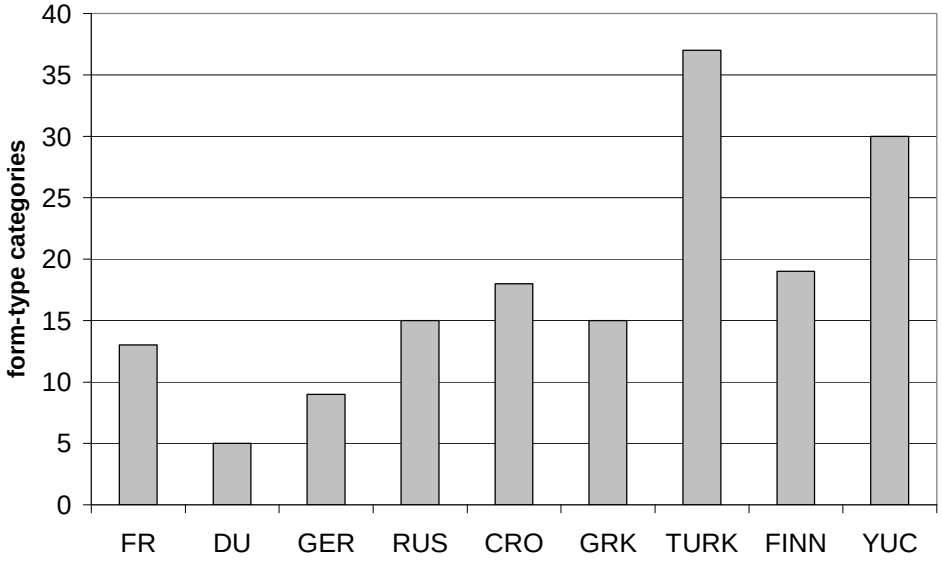

Figure 1b. Number of form-type categories in the input of the 9 languages: verbs

The parameter of paradigmatic morphological richness refers to the number of non-homophonous form-type categories on the paradigmatic axis as these occur in the input. Thus, English noun inflection for example has only two categories, the unmarked, suffixless singular and the marked plural. In English verb inflection, there are the unmarked, suffixless base-form, the 3.Sg.Pres., the Preterit, the Past Participle, and the Gerund, i.e. five categories, as in *sing, sing-s, sang, sung, sing-ing*.

Figures 1a and 1b present an overview of the different numbers of form-type categories for noun inflection and verb inflection in children's input of the nine languages (for details see Laaha et al., this volume; Stephany et al., this volume; Aksu-Koç et al., this volume):

Mean size of paradigm (MSP) represents a second way of measuring paradigmatic morphological richness in the input (see Xanthos & Laaha, this volume).

A2. Syntagmatic morphological richness

The parameter of syntagmatic morphological richness refers to the number of suffixes occurring one after the other in nouns and verbs input to children, starting with zero and reaching as many as five suffixes (Aksu-Koç et al., this volume). The following example taken from Turkish demonstrates how the role syntax plays in inflecting and isolating languages can be taken over by the rich morphology of an agglutinating language (Example 2):

(2) Turkish

Kayık	*-lar*	*-ımız*	*-da*	*-ki*	*-ler*	*-ı*	*-kurtar*	*-dı*	*-lar*
Boat	-PL	-POSS:PL	-LOC	-REL	-PL	-ACC	-save	-PRET	-PL

'They saved those who were in our boats'

Thus, the final three-morpheme verb *kurtar-dı-lar* corresponds to the two English words *they saved*, and the Turkish nominal direct object form, which comprises seven morphemes, *kayık-lar-ımız-da-ki-ler-ı* corresponds to the six English words *those who were in our boats*. Thus Turkish morphology is much richer on the syntagmatic axis than English: the noun has six suffixes, the verb two, in contrast to only one suffix in the English verb and noun each. This example demonstrates what is potential in this most agglutinating language. Clearly, Turkish caregivers avoid such long sequences in the input to children. The degrees of syntagmatic

morphological richness found in children's input of the nine languages are displayed in Figures 2a and 2b (for details see Laaha et al., this volume; Stephany et al., this volume; Aksu-Koç et al., this volume):

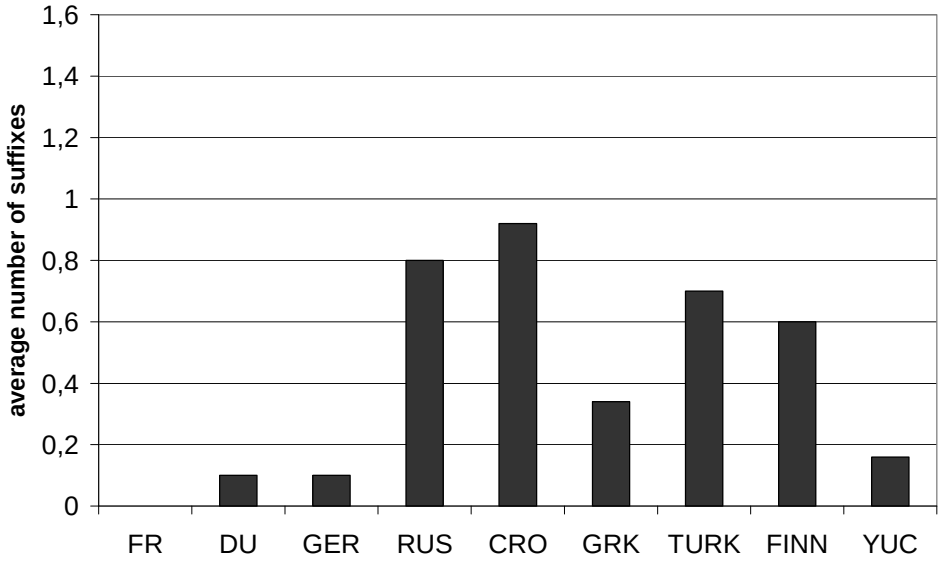

Figure 2a. Average number of suffixes in the input of the 9 languages: nouns

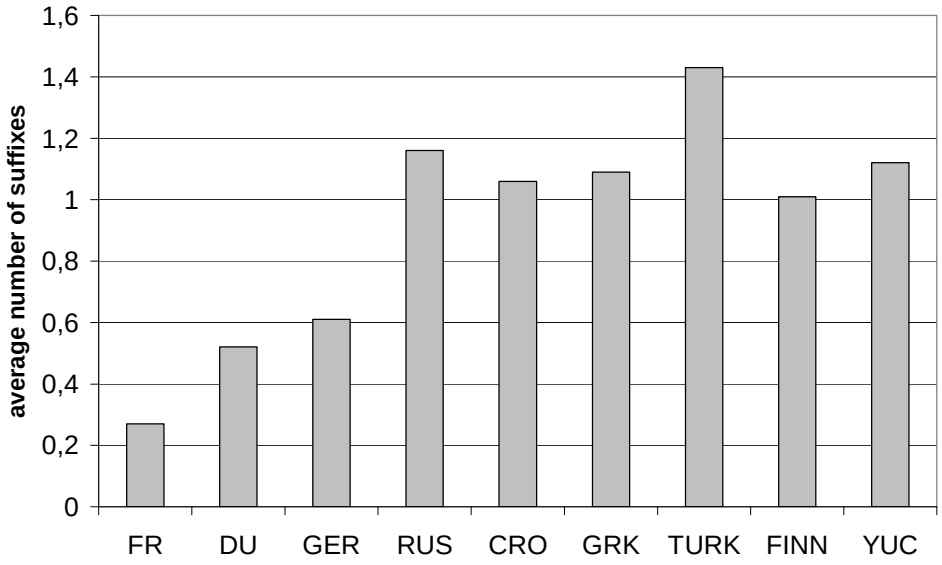

Figure 2b. Average number of suffixes in the input of the 9 languages: verbs

B. Transparency

The most important of the other typological parameters is morphological transparency of grammatical words, which is often referred to in studies of Natural Morphology (see Dressler et al., 1987; Kilani-Schoch & Dressler, 2005: 69-77,). Its sub-parameter of word transparency refers to the degree of transparency (or its opposite: opacity) of the whole inflectional paradigm of a noun or verb. Opacity refers to the fact that the shape of a base or a suffix is blurred. While the relation between the English singular form *fife* and its plural *fife-s* is transparent, the shape of the base *wife* is blurred in the plural form *wive-s*, where word-final

voiceless [f] is changed into voiced [v]. The plural form *children* (with [ɪ] instead of [aɪ] and irregular [r]) of the singular *child* is still more opaque.

The difference between word and form transparency may be illustrated with examples from the English verb. The paradigm of the verb *bring* has to be considered as non-transparent because of a single opaque member (*brought*), in spite of the fact that its other two members, *bring-s* and *bring-ing*, are in a transparent relationship with the base *bring*. What is important in the context of language acquisition is whether both transparent and opaque forms of a given lexeme are used in the input. This has been measured by the sub-parameter of form transparency: if *brought* does occur in the input, the verbal paradigm of *to bring* has been considered as opaque, if not, as transparent.

C. Uniformity

The parameter of uniformity of Natural Morphology (see Dressler et al., 1987; Kilani-Schoch & Dressler, 2005: 87, 187) is often referred to as unifunctionality in the acquisition literature (see Slobin, 1985: 1227-1228) and may lead to what is called inflectional imperialism (see Slobin, 1985: 1216). There is uniformity whenever a given slot in the paradigms of the members of a certain lexical class is filled by a single form throughout, as in the case of the English gerund, which has just one form, namely *-ing*. There is non-uniformity whenever several markers compete with each other, i.e. in the case of allomorphy, as in the English past participle suffix, which is either *-d* as in *work-ed* or *-en* as in *writt-en*.

D. Salience

The last parameter to be investigated is phonological salience (Peters, 1985; Gillis, 2003: 196-199). We have limited our investigation to two types of salience, segmental salience of vowels (ignoring consonants) and prosodic salience. Segmental salience is found if the vowels occurring in suffixes are full vowels, any reduced vowel being non-salient. Prosodic salience can at least be attributed to stressed suffixes in word-final position. Thus, suffixes in word-final vs. other positions will be contrasted.

1.3. Hypotheses

The foregoing leads to the following general hypothesis concerning the development of morphology in first language acquisition: children are sensitive to the typological properties of the language they are acquiring (Slobin, 1985; Peters, 1997; Bavin, 1998: 38, 52; Devescovi et al., 2005). In our case, this means that they are sensitive to the relative communicative importance and structure of morphology in their verbal interactions.

From this general hypothesis the following more specific main hypothesis can be derived: *The richer noun or verb morphology is in the input, the more stimulated the child will be to develop noun or verb morphology rapidly.*

The independent variables of this main hypothesis are morphological richness on the paradigmatic and the syntagmatic axes, on the one hand for nouns, on the other hand for verbs.

The question is how to choose the corresponding dependent variables? Although “early acquisition”, often referred to in the literature, might be a candidate, it is too vague a notion since it may refer to the early emergence of morphology (see Dressler et al., 2003), to the acquisition of sizable parts of morphology (but which ones?) or even to its entire acquisition or mastery (see Berman, 2004). Not to mention the disturbing variable of early vs. late talkers among children acquiring the same language.

Instead we have opted for a dependent variable which lends itself to operationalization, namely speed of early morphological development in nouns vs. verbs measured by growth of paradigms. We therefore reformulate our above main hypothesis in the following way: the degree of morphological richness of the input will have a positive impact on the speed of development of inflectional morphology in the child. It seems to follow from this that the richer morphology is in the input the more children will be stimulated to focus on this part of grammar and develop at least the core domain of morphology more rapidly. It must not be forgotten, however, that more complex morphology will take longer to be entirely mastered than poorer morphology.

The rationale behind our main hypothesis is that children acquiring a morphologically rich suffixing language will be more inclined to focus on differences in meaning expressed by variable endings than children learning a language with equally suffixing but poor inflectional morphology such as English. Morphologically rich languages are on the one hand agglutinating languages, such as Turkish and Finnish, where much of what is expressed by syntactic constructions in languages such as English, French, Dutch and German, is expressed by verbal or nominal inflectional morphology. On the other they include strongly inflecting languages, such as Russian and Croatian, where syntactic constructions are distinguished by inflectional endings rather than by word order as for example in English and French. We therefore predict that the speed of development of inflectional morphology should be higher in agglutinating and strongly inflecting languages than in weakly inflecting ones.

The dependent variable “speed of development” will be further explained and quantitatively related to the two independent variables of paradigmatic and syntagmatic morphological richness in the chapter “Method” (see Xanthos & Laaha, this volume).

Further independent variables the effects of which we have tested in our nine languages are morphological transparency, uniformity and salience:

The literature is full of claims or at least remarks that the morphology of Turkish and of other similar languages is acquired early, because it is very transparent (see Aksu-Koç & Slobin, 1985: 847; Slobin, 1985: 1216; Peters, 1997: 181; Savickiene, 2003). Similar arguments are made for the facilitating effect of uniformity and salience of marking (Slobin, 1973; Peters, 1985).

From these plausible claims in the literature we derived the hypothesis that the independent variables of morphological transparency, uniformity and salience should also influence speed of early development.

By testing these hypotheses across nine languages, we aim to gain further insights into factors affecting the course of early morphological development, and to explore the degree to which the three different approaches to language typology prove useful for dealing with the data.

2. Method

Aris Xanthos and Sabine Laaha[*]

2.1. Introduction

The aim of this chapter is to set out the method that was used to translate the hypotheses of this research into formal statements bearing on the relationship between quantitative variables, and to produce empirical values for those variables on the basis of a set of acquisition corpora. As stated by Dressler in the introduction (this volume), the main hypothesis of this research is that "the richer noun or verb morphology is in the input, the more stimulated the child will be to develop noun or verb morphology rapidly" (p. 8), i.e. that a proportionality relationship exists between the degree of morphological richness of the input (independent variable) and the speed of development of morphological richness in child speech (dependent variable). In addition, it is hypothesized that factors such as transparency, uniformity and salience should show an impact on speed of development.

Correlation coefficients are widely spread statistical tools for studying and testing hypothetical proportionality relationships. By contrast, there is not yet a consensus on how to quantify the degree of morphological richness of a language sample or the speed of its development. As regards the *dependent variable* involved in our hypothesis, we chose to focus on the paradigmatic aspect of morphological richness, i.e. the diversity of word-forms composing inflectional paradigms. In particular, we adopted the *mean size of paradigm* (*MSP*) measure (Xanthos & Gillis, in prep.) as a basis for the calculation of the speed of development of morphological richness.

Several *independent variables* were constructed and applied to the input portion of our corpora: they assess morphological richness (both from a paradigmatic and a syntagmatic perspective), aspects of transparency, uniformity and salience.

The rest of this chapter is organized as follows. Section 2.2 briefly describes the structure of the child corpora that were used to calculate the values of the dependent and independent variables. In section 2.3, we explain how MSP was used to measure the speed of development of morphological richness. Section 2.4 surveys the whole set of independent variables that were defined. Finally, section 2.5 reviews the material presented in the chapter and gives a succinct view of the score of each corpus for each variable.

2.2. Data

The data analyzed in this research are naturalistic, longitudinal production data, starting with the child's onset of speech until the age of about three years. The children investigated were recorded several times per month at their homes, in unstructured settings, interacting with their mother or other caretakers. The data were transcribed in CHAT format and coded morphologically according to the norms of CHILDES (MacWhinney, 2000).

As shown in Table 1, the data have a complex hierarchical structure. They come from nine different languages, clustered into three language groups. Each language is represented by one or more child corpora, divided into child versus input data. The child data are grouped into monthly

[*] This work was supported in part by a grant of the Swiss National Science Foundation to the first author. The authors are grateful to François Bavaud, Wolfgang Dressler, Steven Gillis, Marianne Kilani-Schoch and Gregor Laaha for stimulating discussions.

samples, whereas the input data are treated as a single sample. Both child and input data are in turn split into nominal lemmas (including proper nouns) versus verbal lemmas (including auxiliaries and modal verbs).[1] Ultimately, the data are token frequencies of the forms occurring with these lemmas, as measured by the CLAN program FREQ (MacWhinney, 2000).

The "language" and "language group" factors were used mainly as ways of structuring the qualitative analysis of the results. The corpora actually played the role of individuals (in statistical parlance), whose scores on the dependent and independent variables were measured and served as input to the statistical analysis.[2] In general, the whole range of data was used for the computation of independent variables, whereas an alignment procedure was applied prior to the calculation of the dependent variable (see section 2.3.4 below), hence the possibly differing starting and end points (SP and EP) reported in Table 1 under the labels *Original* and *Aligned*. Notice that the token counts reported under *Child data* refer to the counts resulting from the alignment procedure.

Table 1. Overview of the data

Language group	Language	Child	# tokens (nouns + verbs)		Age range (SP - EP)	
			Child data	Input data	Original	Aligned
Agglutinating	TURK	Deniz	1662 + 2230	973 + 1193	1;3 - 2;0	1;6 - 2;0
	FINN	Tuomas	1342 + 1286	3170 + 4484	1;7 - 2;3	1;7 - 2;3
		Tuulikki	1584 + 2750	2589 + 4329	1;7 - 3;0	1;7 - 2;10
	YUC	Armando	1439 + 1684	445 + 682	2;0 - 3;0	2;0 - 3;0
Strongly inflecting	RUS	Filipp	4292 + 2162	9316 + 7394	1;4 - 2;8	1;4 - 2;8
		Liza	1936 + 732	4977 + 4211	1;6 - 2;5	1;8 - 2;5
	CRO	Antonia	1762 + 4136	5505 + 10795	1;3 - 2;8	1;5 - 2;8
	GRK	Christos	3750 + 1620	733 + 927	1;7 - 2;6	1;8 - 2;6
Weakly inflecting	GER	Katharina	1138 + 1031	3568 + 5775	2;0 - 3;0	2;3 - 3;0
		Jan	4027 + 3121	8053 + 13984	1;3 - 2;6	1;5 - 2;6
	DU	Jolien	3021 + 1278	4247 + 4362	1;5 - 2;5	1;5 - 2;5
	FR	Sophie	6127 + 6655	7484 + 13774	1;6 - 3;0	1;6 - 3;0
		Emma	3154 + 4765	6158 + 8455	1;4 - 2;11	1;4 - 2;11

2.3. Dependent variable

The dependent variable involved in the main hypothesis of our research is the speed of development of morphological richness in child speech. In order to quantify this notion, we have used the *mean size of paradigm* (*MSP*) measure proposed by Xanthos and Gillis (in prep.). In this section, we will first recall the general definition of MSP. Then we will show how it was implemented in the developmental framework of this research; in particular, we will introduce the distinction between total and cumulative MSP and will explain how successive evaluations of MSP were used to calculate the speed of development of morphological richness. Then, we will consider the question of the dependence of MSP on sample size and describe the resampling procedure that was applied for size normalization. Finally, we will discuss the method that was used to set the starting and end point for each child corpus that was investigated.

[1] Diminutives/hypocoristics and compound nouns are counted as different noun lemmas; prefix verbs are counted as different verb lemmas (the only exception are Dutch and German separable particle verbs with the same base which are considered as one single verb lemma).

[2] An exception to this is the analysis of variance reported in the chapter “General results” (Xanthos, this volume), where individuals are monthly values of speed of development (see also fn. 4 below) and factors "language" and "language group" are briefly investigated.

2.3.1. Mean size of paradigm (MSP)

Despite frequent references to the intuitive notion of (paradigmatic) morphological richness, there is no widely accepted way to measure it. In fact, until recently, very few quantitative indices were designed to capture this aspect of a sample's linguistic richness (see Xanthos & Gillis, in prep., for a review). Arguably, the first proposal suitable for cross-linguistic acquisition studies was the *inflectional diversity* (*ID*) measure developed by Malvern et al. (2004). It is based on the measure of lexical diversity D, and defined as ID := $D^{\text{word-forms}}$ - D^{lemmas} (i.e. the difference between D as measured on inflected word-forms and the same measure for lemmas).

An alternative approach was proposed by Xanthos and Gillis (in prep.), where paradigmatic inflectional richness is characterized in terms of an average number of distinct (i.e. non-homophonous) inflected forms per lemma. In its simplest version, *mean size of paradigm* (*MSP*) is defined as

$$\text{MSP} := \frac{|F|}{|L|} \tag{1}$$

where $|F|$ stands for the number of different inflected forms in a sample, and $|L|$ stands for the number of different lemmas. For example, given the sample {*has*, *are*, *have*, *has*, *are*}, containing 5 inflected English verbs (tokens), one finds that $|L| = 2$ (*have* and *be*), and $|F| = 3$ (*has*, *have* and *are*), so that MSP = 3/2 = 1.5 (for a similar proposal, see Laaha, 2004: 188).

Like ID, MSP relies on the categorization of inflected word-forms into lemmas. On the other hand, it embodies a *multiplicative* conception of the relationship between morphological and lexical richness (as measured by $|F|$ and $|L|$): one finds indeed that $|F| = |L| \cdot \text{MSP}$. By contrast, Malvern and colleagues adopt an *additive* conception of the same relationship, where $D^{\text{word-forms}} = D^{\text{lemmas}} + \text{ID}$.

By construction, MSP ranges between 1 and $|F|$. It is minimal and equal to 1 if and only if $|L| = |F|$, i.e. if each lemma has exactly one (possibly repeated) inflected form in the sample. It is maximal and equal to $|F|$ if and only if $|L| = 1$, i.e. if all word-forms in the sample belong to a single lemma. 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. Xanthos and Gillis (in prep.) provide evidence that dependence on ample size is a characteristic feature of (raw) MSP; however, their results suggest that it can be undone by applying suitable normalization procedures (see section 2.3.3 below).

It is claimed that ID inherits from D the important property not to be a function of sample size (Malvern et al., 2004). Moreover, it can be assumed that it shares with it the property of representing the entirety of the relationship between morphological (respectively lexical) richness and sample size. For this research, however, we chose to use MSP for two main reasons. First, we believe that the characterization of morphological richness in terms of an (average) number of inflected forms is a particularly intuitive one; this will be an important asset when we will add an extra layer of complexity by considering the *speed of development* of morphological richness. Second, in contrast to ID, MSP does *not* rely on a specific assumption as to the nature of the relationship between morphological richness and sample size; to that extent, it may be rightly considered a more agnostic perspective.

2.3.2. Developmental perspective

Total vs. cumulative MSP

As indicated in section 2.2, the corpora that were used for this research are sequences of monthly samples. In order to account for this additional structure, we introduce a distinction between the *total* MSP, which is computed over a whole corpus, and the *cumulative* MSP (at month *m*), which is computed over the whole corpus *up to month m*. The latter is denoted by a subscript: MSP_m.

For instance, consider the following corpus, ranging over three months: month 1 {*has*}, month 2 {*have*, *having*} and month 3 {*am*, *are*}. Applying definition (1), one finds that the total MSP is equal to 5/2 = 2.5. As to the cumulative MSP, one finds $MSP_1 = 1/1 = 1$, $MSP_2 = 3/1 = 3$, and $MSP_3 = 5/2 = 2.5$ (by construction, MSP_3 is equal to the total MSP). This example shows that the cumulative MSP is not bound to increase over time. In particular, it increases between months *m* and *m′* if and only if the number of new word-forms is greater than the number of new lemmas times MSP_m.

The use of a cumulative (as opposed to *monthly*) definition of MSP is an important feature of our methodology. It conveys our intent to capture paradigmatic oppositions between word-forms whose occurrences may span different months. Consequently, morphological richness is in principle evaluated over an ever growing amount of data. Given the dependence between (raw) MSP and sample size (see previous section), this makes a strong case for carefully designing a normalization procedure. Section 2.3.3 below describes our efforts to do so.

Speed of development of MSP

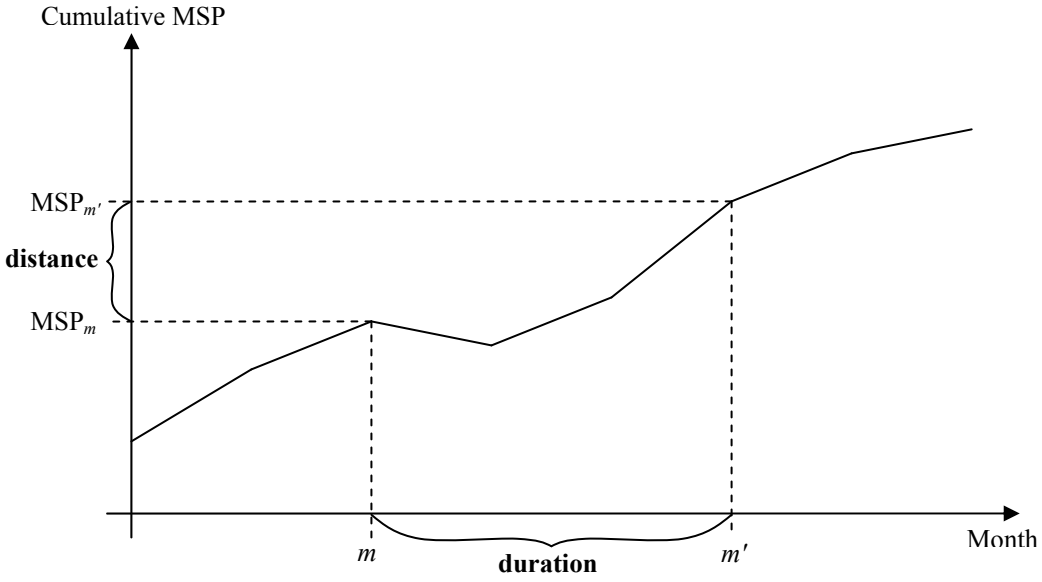

Figure 1. Speed of development of MSP is defined as the ratio of a "distance" to a duration

Cumulative MSP enables us to monitor the development of morphological richness over time. In analogy with physics, we may calculate the monthly *speed* of this development over a given time interval (say between months *m* and *m′*) as the ratio of a "distance" to a duration:

$$\text{speed}_{m,m'} := \frac{\text{MSP}_{m'} - \text{MSP}_{m}}{m' - m} \quad (2)$$

Here, distance is expressed as a difference of cumulative MSP (at months *m′* and *m*), and duration is simply the difference between *m′* and *m* (see Figure 1 above).

For this research, we generally considered the case of successive months (i.e. $m' = m + 1$); this enables us to simplify equation (2) by defining the speed at month *m* (for $m > 1$) as:

$$\text{speed}_{m} := \text{MSP}_{m} - \text{MSP}_{m-1} \quad (3)$$

Using a corpus of *L* months, we may compute *L*-1 such values of monthly speed. These are in turn averaged to yield the monthly speed in the corpus as a whole. This average value is the reported score of each corpus on the dependent variable - with the additional specification that it is computed on the basis of *normalized* cumulative MSP, as indicated in the next section.[3]

For instance, consider the three months corpus defined above, and the corresponding values of cumulative MSP. The speed of development at month 2 is given by $\text{speed}_2 = \text{MSP}_2 - \text{MSP}_1 = 3 - 1 = 2$; likewise, one finds $\text{speed}_3 = \text{MSP}_3 - \text{MSP}_2 = 2.5 - 3 = -0.5$ (this negative speed accounts for the *regression* in cumulative MSP that was observed between months 2 and 3). The average monthly speed is equal to $(2 - 0.5) / 2 = 0.75$.

As mentioned in section 2.3.1 above, the intuitive interpretation of MSP extends directly to the speed of its development. Indeed, the unit of measurement of speed is an average number of word-forms per month. Thus an (average) monthly speed of 2 simply means that, on average, 2 new inflected forms of each lemma occur during a month.

2.3.3. MSP and sample size

The dependence between (raw) MSP and sample size is inherent in its definition, and empirical results indicate that it is prohibitively strong (Xanthos & Gillis, in prep.). This observation leads these authors to advocate the systematic application of some normalization procedure. This is an even more stringent requirement in view of our cumulative definition of MSP (see section 2.3.2 above).

Specifically, Xanthos and Gillis propose the use of the following resampling method, inspired by the *mean segmental type-token ratio* (Johnson, 1944). Let *N* denote the original sample size. The first step is to set the value of *S* (with $S \leq N$), which stands for the *expected number of tokens per subsample* (the only arbitrary parameter in the procedure). Then, $B := N/S$ subsamples are constructed by randomly picking tokens with replacement *between* but not *within* subsamples (conversely, each subsample can be viewed as the result of an independent, random process of deletion applied to the original sample). For each token, the probability of being selected when constructing each subsample is defined as $P := S/N$. Thus, each token is expected

[3] The reason why we calculated the monthly values before averaging them, rather than directly computing the overall monthly speed using formula (2), is that it enabled us to estimate the *variance* of speed for a given corpus, and perform an analysis of it (see Xanthos, this volume).

to occur in $B \cdot P = 1$ subsample on average[4]. Eventually, the MSP of each subsample is computed, and the average of these B values is reported as the *normalized* MSP (over S tokens) of the sample, or MSP(S).

Xanthos and Gillis report empirical evidence suggesting that, while the MSP(S) of a sample clearly depends on S, it is actually independent of the original sample size. We adopted their procedure, with a few modifications in order to fit the cumulative definition of MSP and the calculation of speed. By construction, cumulative MSP is evaluated on the basis of an ever increasing amount of data; our concern was to ensure that this increase remained comparable between corpora. Since our goal was to compute the *monthly speed* of development of MSP, we decided to set the parameter S to a possibly different value for each corpus, in such a way that the *expected number of tokens per month* would be the same for all of them.

Consider a corpus of N tokens ranging over L months. Let Z denote the expected number of tokens per month. Parameter S is now defined as $S := L \cdot Z$. The number B of subsamples and the probability P of selecting a token when constructing a subsample are then calculated as in the original procedure. For instance, in the case of a corpus of $N = 3200$ tokens ranging over $L = 14$ months and with an expected number of tokens per month arbitrarily set to $Z = 100$, we find that $S = 14 \cdot 100 = 1400$, so we construct $B = 3200/1400 \approx 2$ subsamples, where tokens are sampled with probability $P = 1400/3200 = 0.44$. Keeping Z constant, another corpus with a size of $N' = 6000$ tokens and a length of $L' = 22$ months would yield $S' = 2200$, $B' \approx 3$, and $P' = 0.37$.

An important feature of the modified resampling procedure is that it preserves the sequential structure of the original corpus: when constructing a given subsample, each selected token is assigned to the month where it actually occurred. Thus, each of the B subsamples is itself a sequence of L monthly (sub-)samples. We may then calculate the cumulative MSP for each month of each subsample (this corresponds to B series of L values), and the average cumulative MSP per month. We call this value the *normalized* cumulative MSP (over Z monthly tokens) at month m, and write it $\mathrm{MSP}_m(Z)$.

For this research, we arbitrarily set the number of expected tokens per month to $Z = 50$. The reported values of cumulative MSP were systematically normalized that way, and served as a basis for calculating the monthly values of speed of development as indicated in the previous section. These monthly values were eventually averaged over all months composing a given corpus, and the result was reported as the value of the dependent variable for that corpus.

2.3.4. Starting and end points

As mentioned in section 2.2, we did not use the full range of available data for the calculation of the dependent variable. Age-related differences in the original data provided for this research (e.g., age range 1;3-2;0 for the Turkish child Deniz, in contrast to 2;0-3;0 for the Austrian child Katharina, see also Table 1), as well as possible inter-individual differences between the children investigated led us to consider the question of a common *base-line* (including a common starting and end point) for this analysis.

We considered different possible candidates for a common base-line. Age was not qualified as a useful candidate, in view of the well-known inter-individual variation in onset time of language development (Bates et al., 1995: 97). We thought of using a common base-line expressing the MSP in child data as a proportion of the MSP in the input, but this approach proved impractical: for systems with extremely poor inflection, like French nouns, this proportion is close to 100% from the earliest age on, resulting in the elimination of such corpora as a whole. Mean length of utterance (MLU) was rejected as well, on the grounds that it is not a

[4] As noted by Xanthos and Gillis (in prep.), this number may actually diverge from 1, because the number B of subsamples must be rounded to the nearest integer.

satisfactory index for cross-linguistic comparison (for cross-linguistic differences between English and Italian MLUs, see Devescovi et al., 2005). As suggested by Devescovi et al. (2005), vocabulary measures constitute more robust indices than MLUs from a cross-linguistic point of view. We therefore settled for the *proportion of utterances with verbs (PUV)*.

The largest common interval of PUV between all corpora was found to be excessively small, so that we could not use this straightforward criterion for selecting months to be suppressed. Therefore, we applied a more complex method, based on the *cumulative distribution function* (*CDF*) of PUV. For a given sample, the CDF of a variable *X* is defined as the proportion of individuals having a score lower than or equal to *X*. In our case, the individuals are all the monthly samples in the whole set of corpora; the CDF of PUV is plotted in Figure 2 below.

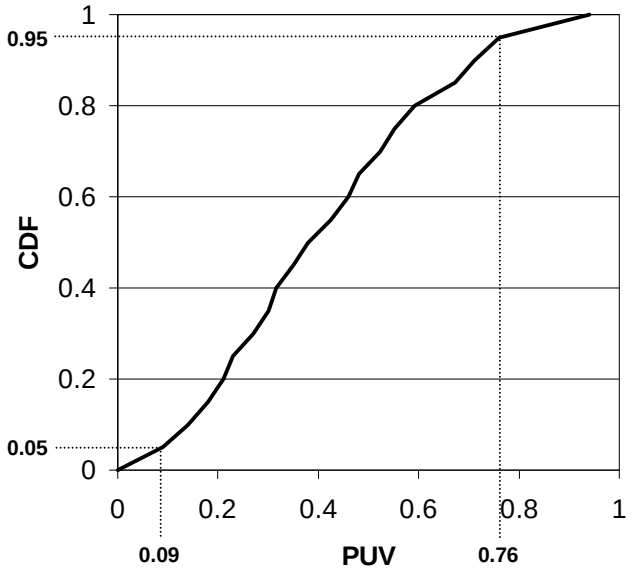

Figure 2. Cumulative distribution function of PUV; dotted lines indicate the value of the 5th and 95th percentiles

Having calculated this function for our data, we arbitrarily set to 10% the *a priori* proportion of data (in terms of monthly samples) that we were ready to give up for the sake of aligning corpora. This proportion was then evenly split between the two ends of the distribution of PUV. The CDF indicates that the 5th percentile of PUV, i.e. the value such that 5% of the monthly samples have a PUV lower than or equal to it, is 0.09 (see figure 2); the 95th percentile, i.e. the value such that 5% of the monthly samples have a PUV higher than it, is 0.76. The selection rules that we applied to each corpora were thus:

- Set the starting point to the first month such that PUV > 0.09.
- Set the end point to the last month such that PUV ≤ 0.76.

Notice that this formulation does not lead to the suppression of monthly samples exceeding these bounds when they occur between other samples that do not exceed them: it merely amounts to cropping the beginning and end of corpora when needed. Consequently, the proportion of months effectively suppressed is 7% rather than 10%. On average, the starting points are set one month later in the aligned corpora (1;8 instead of 1;7), and the end points one month earlier (2;8 instead of 2;9). As an estimate of the resulting reduction in variation, the standard deviation of PUV is reduced from 0.11 to 0.09 at starting point, and from 0.22 to 0.17 at end point.

2.4. Independent variables

The main independent variable involved in this research is the degree of morphological richness of the input (both from a paradigmatic and a syntagmatic perspective). In addition, several other independent variables were constructed which are susceptible to show an impact on speed of development of morphological richness in child speech, notably the transparency, uniformity and salience of word forms.

Input analyses were conducted for each child corpus separately; within each corpus, the input data are treated as a single sample (i.e. totals over all monthly samples; see also section 2.2 above). It has to be noted that the amount of input data being analyzed differs between the three language groups. For the weakly inflecting languages (see Laaha et al., this volume), the entire input samples (i.e. all input utterances) are analyzed, whereas for most strongly inflecting and agglutinating languages (see Stephany et al., this volume; Aksu-Koç et al., this volume), reduced input samples (i.e. 900 input utterances per corpus) are used.

In what follows, we will list the seven different independent variables constructed for this research (see also Dressler, this volume) and explain how they were calculated.

- *Variable A1: Paradigmatic morphological richness (total input MSP)*

For the main part, Variable A1 was calculated using the methodology described in section 2.3 above. However, in contrast to the dependent variable, it measures the *total* (as opposed to *cumulative*) MSP in the input, and it is considered from a *static* perspective (i.e. the step leading to the derivation of *speed* is not taken); thus A1 represents in effect the mean number of inflected forms per paradigm in the whole input of a given corpus.

With regard to the issue of sample size, we applied the original normalization procedure described by Xanthos and Gillis (in prep.) and summarized in section 2.3.3 above. In particular, we computed the normalized MSP over 1000 tokens (recall that in this version of the normalization procedure, the *total* size of the sample is normalized, as opposed to the *average* number of tokens *per month*). For samples where less than 1000 tokens were available (i.e. nouns and verbs in Greek and Yucatec, as well as Turkish nouns), we applied a logarithmic regression to compute the *expected* MSP for 1000 tokens.[5]

- *Variable A2: Syntagmatic morphological richness (average number of suffixes)*

Variable A2 is defined as the average number of inflectional suffixes in the input data. It was computed by dividing the total number of inflectional suffix tokens by the total number of word tokens. Nouns and verbs were treated separately.

- *Variable B1: Word transparency*

Variable B1 is defined as the proportion of transparent word types (or lemmas) in the input data. It was calculated in the following way: we used the CLAN program FREQ to establish a total word type list of the input data and classified these words (or rather: their inflectional paradigm) as either transparent or non-transparent (i.e. blurred by morphophonological alternations, such as stem vowel change, phonologically unpredictable mobile stress, consonant alternations, and suppletion). This means that any word belonging to a paradigm in which there is at least one

[5] This decision was based on the empirical observation by Xanthos and Gillis (in prep.) that the growth of MSP with regard to sample size is better represented by a logarithmic regression than by a linear regression or one based on a power transform. Clearly, this is at best a good approximation, which the availability of more data would enable us to discard.

non-transparent alternation in the target system, was counted as non-transparent, provided that the category which triggers such an alternation occurred in the input. Then, we divided the number of transparent word types (or lemmas) by the total number of word types. Nouns and verbs were treated separately.

- *Variable B2: Form transparency*

Variable B2 is defined as the proportion of transparent word forms (tokens) in the input data. We used FREQ to establish a total word form list of the input data and classified these forms as either transparent or non-transparent (i.e. blurred by morphophonological alternations, such as stem vowel change, phonologically unpredictable mobile stress, consonant alternations, and suppletion). Then, we divided the number of transparent word forms by the total number of word forms (tokens). Nouns and verbs were treated separately.

- *Variable C: Uniformity*

Variable C is defined as the proportion of uniform inflectional categories in the input data. We first classified inflectional noun and verb categories of the language being investigated as either uniform or non-uniform (i.e. absence vs. presence of allomorphy). Then, we used FREQ to calculate the token frequency of uniform categories in the input data and divided it by the total number of inflectional categories (tokens).

- *Variable D1: Phonological-segmental salience*

Variable D1 is defined as the proportion of phonologically and segmentally salient inflectional suffixes in the input data. We distinguished between a) languages which do not have inflectional suffixes with full vowel (0% salience); b) languages which have many inflectional suffixes with full vowel, but no inflectional suffixes with reduced vowel (100% salience); and c) languages which have inflectional suffixes with both full vowels and reduced vowels. If the language being investigated was of type (c), then we calculated the proportion of inflectional suffixes with full vowels out of all inflectional suffixes (tokens). Nouns and verbs were treated separately.

- *Variable D2: Prosodic salience*

Variable D2 is defined as the proportion of prosodically salient inflectional suffixes in the input data. We distinguished between a) languages in which inflectional suffixes are always unstressed (0% salience); b) languages in which final inflectional suffixes are always stressed, preceding suffixes unstressed (100% salience); and c) languages in which inflectional suffixes are both stressed and unstressed. If the language being investigated was of type (c), then we calculated the proportion of stressed inflectional suffixes out of all inflectional suffixes (tokens). Nouns and verbs were treated separately.

2.5. Summary

In this chapter, we presented the method used to investigate the hypotheses of this research. The main hypothesis is that there exists a proportionality relationship between the degree of morphological richness of the input and the speed of development of morphological richness in child speech. In addition, factors such as transparency, uniformity and salience should show an impact on speed of development (see Dressler, this volume).

After a brief description of the data investigated (section 2.2), we explained how the mean size of paradigm (MSP) measure proposed by Xanthos and Gillis (in prep.) was used to calculate the *dependent variable* of this research, the speed of development of morphological richness in child speech. We also considered the question of the dependence of MSP on sample size and discussed the method used to set a common base-line for this analysis (section 2.3). Then, we explained the calculation of the *independent variables* of this research: the main independent variables investigated concern the degree of morphological richness of the input, both from a paradigmatic and a syntagmatic perspective (Variables A1, A2); the other independent variables are word and form transparency (B1, B2), uniformity (C), phonological-segmental and prosodic salience (D1, D2) (section 2.4).

Tables 2a and 2b below summarize the values of each variable for each of the 13 child corpora investigated. Empty cells indicate that the value could not be calculated for the given child corpus. More details about the analysis of each language are given in the three language group chapters (see Laaha et al., this volume; Stephany et al., this volume; Aksu-Koç et al., this volume).

Table 2a. Values of the independent and dependent variables, by child corpus: nouns

Language	Child	Indep. V							Dep. V
		A1	A2	B1	B2	C	D1	D2	
CRO	Antonia	1,39	0,92	0,95	0,99	0,00	1,00	0,00	0,03
DU	Jolien	1,07	0,10	0,99	0,99		0,00	0,00	0,00
FINN	Tuomas	1,50	0,56	0,17	0,83	0,79	0,90	0,00	0,04
	Tuulikki	1,40	0,63	0,12	0,75	0,73	0,90	0,00	0,02
FR	Emma	1,01	0,00	1,00	1,00	1,00	1,00	1,00	0,00
	Sophie	1,01	0,00	1,00	1,00	1,00	1,00	1,00	0,00
GER	Jan	1,08	0,12	0,85	0,99		0,00	0,00	0,01
	Katharina	1,06	0,08	0,85	0,99		0,00	0,00	0,00
GRK	Christos	1,13	0,34	0,94	0,96	0,68	1,00		0,01
RUS	Filipp	1,38	0,79	0,63	0,97	0,01	0,85	0,08	0,01
	Liza	1,36	0,80	0,69	0,94	0,01	0,92	0,08	0,02
TURK	Deniz	1,91	0,70	0,88	0,89	1,00	1,00	0,92	0,04
YUC	Armando	1,37	0,16	0,97	0,98	0,96	0,90	1,00	0,02

Table 2b. Values of the independent and dependent variables, by child corpus: verbs

Language	Child	Indep. V							Dep. V
		A1	A2	B1	B2	C	D1	D2	
CRO	Antonia	1,91	1,06	0,50	0,43	0,81	1,00	0,00	0,05
DU	Jolien	1,82	0,52	0,50	0,56		0,00	0,00	0,06
FINN	Tuomas	2,06	0,99	0,14	0,51	0,15	0,90	0,00	0,09
	Tuulikki	2,11	1,02	0,12	0,56	0,25	0,90	0,00	0,05
FR	Emma	1,61	0,27	0,84	0,35	0,71	1,00	1,00	0,03
	Sophie	1,53	0,27	0,83	0,34	0,74	1,00	1,00	0,03
GER	Jan	1,83	0,63	0,71	0,55	0,52	0,00	0,00	0,05
	Katharina	1,99	0,59	0,66	0,61	0,57	0,00	0,00	0,07
GRK	Christos	2,41	1,09	0,17	0,65	0,00	1,00	0,11	0,12
RUS	Filipp	1,65	1,09	0,48	0,72	0,44	0,74	0,45	0,02
	Liza	1,65	1,24	0,48	0,62	0,46	0,73	0,29	0,06
TURK	Deniz	3,93	1,43	0,85	0,90	0,93	1,00	0,66	0,23
YUC	Armando	1,96	1,12	0,86	0,92	0,45	0,70	1,00	0,05

3. Weakly inflecting languages: French, Dutch, and German

Sabine Laaha, Steven Gillis, Marianne Kilani-Schoch, Katharina Korecky-Kröll, Aris Xanthos, and Wolfgang U. Dressler

3.1. Typological characteristics

French, Dutch and German are weakly inflecting languages. The two Germanic languages Dutch and German are very closely related genetically and typologically. In both languages, verb inflection is richer than noun inflection; both are SOV/V2 languages: the finite verb in main clauses must occupy a left-peripheral position, preceded by maximally one constituent; the non-finite verb occupies a sentence-final position (Jordens, 1990). The Romance language French is even less strongly inflecting than Dutch and German due to a number of characteristics of the isolating language type (Geckeler, 1984). Noun inflection is by far poorer than verb inflection; in contrast to Dutch and German, French is an SVO language.

In German, nouns can receive number and case marking. Plural is marked by suffixation and/or stem vowel change (Umlaut). There are four cases: nominative, accusative, dative and genitive. Case is marked on the noun less distinctively than on the article; in colloquial Austrian German, case distinctions are reduced with respect to Northern German varieties (Korecky-Kröll & Dressler, in press). German verbs encode the grammatical categories of person, number, tense, mood and voice. There is no grammatical category of aspect. Non-finite categories are the infinitive (*spiel-en* 'to play'), the past participle (*ge-spiel-t* 'played') and the marginally used present participle (*spiel-end* 'playing'). Person (1^{st}, 2^{nd}, 3^{rd}) and number (Sg., Pl.) are expressed synthetically by verbal suffixes. Within the category tense, colloquial Austrian German distinguishes between present, future, perfect, and pluperfect. The present is formed synthetically (*ich spiel-e* 'I play', *er spiel-t* 'he plays'). Perfect, pluperfect and future tense are expressed by periphrastic constructions, i.e. by a combination of Aux + PP / INF. In contrast to Northern German varieties, the synthetic preterite occurs very rarely in colloquial Austrian German (except for the copula *sein* 'to be' and modal verbs) (Laaha, 2004: 58). The synthetic conjunctive preterite, however, is commonly used in some Austrian dialects, in competition with a compound form (preterite conjunctive Aux *würde* + INF).

In Dutch, nouns can receive number marking. This is accomplished by suffixation: the suffixes *-en* and *-s* are productive, while the suffix *-eren* is not productive (occurs with a limited number of words). Case marking is non existent, except for some fossilized phrases. As in German, Dutch verbs encode the grammatical categories person, number, tense, mood and voice. The categories person (1^{st}, 2^{nd}, 3^{rd}) and number (Sg., Pl.) are expressed synthetically by verbal suffixes, whereby there is no person distinction in the plural, which is formally indistinguishable from the infinitive. The non-finite verb forms include the infinitive (*werk-en* 'to work'), the past participle (*ge-werk-t*, 'worked'), and the present participle (*werk-end* 'working'). The category tense is expressed by the present (*ik werk* 'I work', *hij werk-t* 'he works') and the imperfectum (*ik werk-te*, *hij werk-te* 'I/he worked). These are the only synthetic forms. The perfectum, plusquamperfectum, futurum, and futurum exactum are periphrastic. For instance, the perfectum and plusquamperfectum are formed by a combination of the present, resp. the imperfectum of the auxiliaries *hebben* ('have') and *zijn* ('be') plus the past participle of the main verb. The imperative is formally (almost) indistinguishable from the present, and the conjunctive has disappeared from colloquial speech.

In French, the system of the noun is predominantly isolating, i.e. the noun tends to be invariable. There are no cases and number is expressed by determiners (Sg. *le pas* /lə pa/ 'the step', Pl. *les pas* [le pa] 'the steps'), apart from a few unproductive inflectional relics with stem modification e.g., Sg. *cheval* /ʃəval/ 'horse', Pl. *chevaux* /ʃəvo/ 'horses'. In contrast, the verb system is typologically more mixed. Grammatical categories of the French verb are person (1st, 2nd, 3rd), number (Sg., Pl.), tense, mood and voice. Aspect is not encoded separately from tense. In the spoken language these grammatical categories are expressed by proclitic pronouns (*je, tu, il, elle, ils, elles, on parle(s/nt)*[1] /parl/ 'I, you, he, she, we speak'), stem modifications depending on the verb classes, and auxiliaries. Non-finite categories are infinitive, present participle (occurring only as gerundive in the spoken language) and past participle (PP). The infinitive is used in periphrastic constructions such as compound future and modal constructions. PP is part of the compound forms. Within the category tense, spoken French has four compound forms (compound past, compound future, pluperfect, and past future), and two synthetic forms : imperfect (*parl-ais* imperfective aspect) and simple future (*parl(e)-ra*). The simple past (*parl-a* perfective aspect) is used only in fairy tales. Within the category mood, subjunctive and conditional have a synthetic form and a compound form (past subjunctive, past conditional). The French verb system has several inflectional microclasses and paradigms (Kilani-Schoch & Dressler, 2005). They can be divided into two macroclasses like in the weakly inflecting-fusional languages German and Dutch. But the number of the microclasses is higher in French than in the two other languages. To that extent the French verb system can be considered as more inflecting-fusional than the verb system of German and Dutch (Dressler et al., 2006).

One important characteristic of the two Germanic languages, which distinguishes them from the other languages studied in this research, are separable particle verbs such as German *aufmachen* 'to open'; if the particle verb occurs in a V2 position, verb base and particle have to be separated, e.g., *er macht auf* 'he opens', *mach auf!* 'open!' (Bennis et al., 1995). Thus, young Dutch and German children are confronted both with the separated and non-separated variant of these verbs. For this reason, particle verbs with the same base were considered as one single verb type (or lemma) for the analyses presented in this chapter. For one analysis (total input MSP), we calculated two different versions: a version in which particle verbs with the same base were merged, and a version in which they were not merged (see section 3.3.A1 below).

3.2. The data

The data analyzed in this chapter stem from two Swiss children acquiring French, one Flemish child acquiring Dutch and two Austrian children acquiring German as their first languages; all children are monolingual, with no developmental or linguistic problems.

The French-speaking children Emma and Sophie are born in Lausanne (Switzerland) in upper class families. The parents recorded the children at home, in situations of free play or while looking at picture books.[2]

The Dutch-speaking child Jolien was recorded at home interacting with one of her parents in free play interactions, picture book reading, during meals, etc. Both her parents hold a university degree. The recordings were made by a research assistant, who also

[1] *On parle* instead of *nous parlons*.

[2] Thanks are due to the parents of Emma and Sophie for their help in collecting the data and checking the transcription. The data of Emma are more limited than the data of Sophie. Emma was recorded generally only twice a month and some of the recordings are very short (e.g. 1;6, 1;7, 2;0; at 1;7 diary notes are used to complement the recordings). This diversity in the data of Emma is probably responsible for the greater heterogeneity of some of the findings on her language development.

transcribed the recording. Part of speech tagging, lemmatization and morphological decomposition were accomplished by the full-form version of CLAN's MOR (Gillis, 1999). Data from six additional children come from the two Dutch-speaking triplets (Gijs, Joost, Katelijne, and Arnold, Diederik, Maria) available in the CHILDES database (http://www.cnts.ua.ac.be/childes/data/Germanic/Dutch/Schaerlaekens). More information concerning these data can be found in the database manual of CHILDES: http://www.cnts.ua.ac.be/childes/manuals/Germanic. Unfortunately parent language addressed to the two triplets is not available.

The German-speaking children Jan and Katharina come from high-to-middle class families in Vienna, Austria. The children were recorded at home, 2-3 times a month, in interaction with their mothers (and occasionally the father, the observer or the brother/sisters). Recording situations vary between free/toy play, everyday situations (e.g., eating, washing) and picture book sessions. The data were transcribed in CHAT format and coded morphologically using the full-form version of the CLAN program MOR (Gillis, 1999; MacWhinney, 2000; Laaha, 2004).[3]

Tables 1 and 1' present the age ranges investigated, as well as the number of noun and verb tokens occurring in the children's data.

Table 1. Characteristics of the child data

Child	Age range		# Child tokens (nouns + verbs)
	Original	Aligned	
Emma (FR)	1;4 - 2;11	1;4 - 2;11	3154 + 4765
Sophie (FR)	1;6 - 3;0	1;6 - 3;0	6127 + 6655
Jolien (DU)	1;5 - 2;5	1;5 - 2;5	3021 + 1278
Jan (GER)	1;3 - 2;6	1;5 - 2;6	4027 + 3121
Katharina (GER)	2;0 - 3;0	2;3 - 3;0	1138 + 1031

Table 1'. Additional Dutch child data

Child	Age range		# Child tokens (nouns + verbs)
	Original	Aligned	
Gijs (DU)	1;8 - 2;10	1;8 - 2;2	1648 + 1082
Joost (DU)	1;8 - 2;10	1;8 - 2;8	1286 + 797
Katelijne (DU)	1;8 - 2;10	1;8 - 2;8	2219 + 1441
Arnold (DU)	1;10 - 3;1	1;10 - 3;1	2200 + 1318
Diederik (DU)	1;10 - 3;1	1;10 - 3;1	1852 + 1127
Maria (DU)	1;10 - 3;1	1;10 - 2;10	1688 + 1093

Table 2 shows the number of noun and verb tokens in the input. For each of the three languages, the input analysis is based on the entire maternal input sample, i.e. on all maternal utterances in the children's corpora in this age period.

[3] The German data were collected and transcribed by K. Korecky-Kröll and by B. Müller; morphological coding of the data was done by S. Laaha.

Table 2. Characteristics of the input data

Child	Age range	# Input tokens (nouns + verbs) *entire sample*
Emma (FR)	1;4 - 2;11	6158 + 8455
Sophie (FR)	1;6 - 3;0	7484 + 13774
Jolien (DU)	1;5 - 2;5	4247 + 4362
Jan (GER)	1;3 - 2;6	8053 + 13984
Katharina (GER)	2;0 - 3;0	3568 + 5775

3.3. Independent variables of the input

In what follows, we will present the input analyses for French, Dutch and German. They constitute the independent variables of our research.

A. Morphological richness

The first input variable investigated is morphological richness. It was examined both from a paradigmatic and a syntagmatic perspective.

A1. Paradigmatic morphological richness

In order to give a first overview of paradigmatic morphological richness in the input, we made a list of all non-homophonous form-type categories in nouns and verbs occurring in the input of the three languages (Table 3). As can be seen, in nouns, the number of categories is very low in all three languages, with the lowest number in French (only 2 categories). In verbs, the number of categories is higher than in nouns in all three languages; French has more verb categories than the two Germanic languages (13 categories, as compared to Dutch 5 and German 9 categories).

Table 3. Form-type categories in French, Dutch and German

	Nouns	Verbs
FR	*Sg, Pl*	*Sg, 1Sg, 2+3Sg, 1Pl, 2Pl, INF, Imperfect, Future, Conditional, Gerundive, Past participle, Past participle:fem, Simple past*
DU	*Sg, Pl, Gen*	*1Sg/2Sg(inv.)/IMP:Sg, 2Sg/3Sg, Pl/INF/IMP:Pl, Past participle, Past*
GER	*Sg, Pl, Gen Sg, Oblique, Dat Pl*	*1Sg, 2Sg, 3Sg, 1+3Pl/INF, 2Pl, IMP:Sg, Past participle, Preterite, Subjunctive preterite*

Paradigmatic morphological richness was examined quantitatively in terms of total input mean size of paradigm (Variable A1, see Xanthos & Laaha, this volume). The values of total input MSP for each of the three languages are given in Table 4.

Table 4. Total input MSP (mean size of paradigm) in French, Dutch and German

	Nouns	Verbs (+/- merging of particle verbs)	
FR	1.01 *(Input Emm 1.01, Input Sop 1.01)*	1.57 *(Input Emm 1.61, Input Sop 1.53)*	
DU	1.07 *(Input Jol)*	1.82 *(Input Jol)*	1.70 *(Input Jol)*
GER	1.07 *(Input Jan 1.08, Input Kat 1.06)*	1.91 *(Input Jan 1.83, Input Kat 1.99)*	1.57 *(Input Jan 1.51, Input Kat 1.62)*

As can be seen, in nouns, total input MSP is very low in all three languages, and in French even lower than in Dutch and in German; the vast majority of nouns in the input occurs in only one single form. In verbs, total input MSP is higher than in nouns in all three languages. For Dutch and German verbs, we calculated two different versions of total input MSP: a version in which particle verbs with the same base were merged, and a version in which they were not merged. In the case of merging, total input MSP in verbs is clearly higher in the two Germanic languages than in French; in the case of non-merging, only the Dutch input shows higher values than French. Note also that there is a considerable difference between the two Austrian input data as far as verbs are concerned: in Jan's input, total input MSP is at 1.83, whereas in Katharina's input, it is at 1.99. Further investigations are necessary to explain this difference.

A graphical summary of total input MSP in the three languages is presented in the two figures below (Figures 1a, 1b).

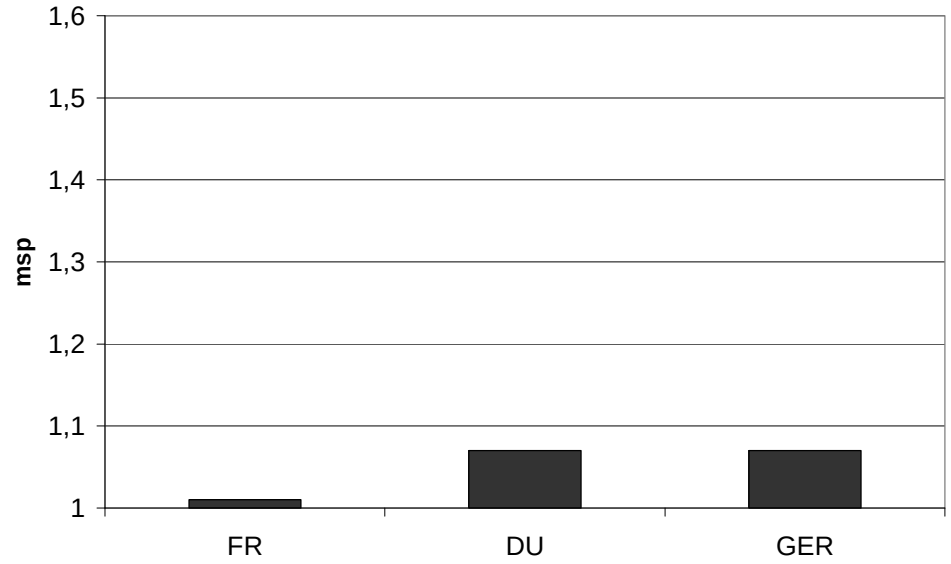

Figure 1a. Total input MSP (mean size of paradigm) in nouns for French, Dutch and German

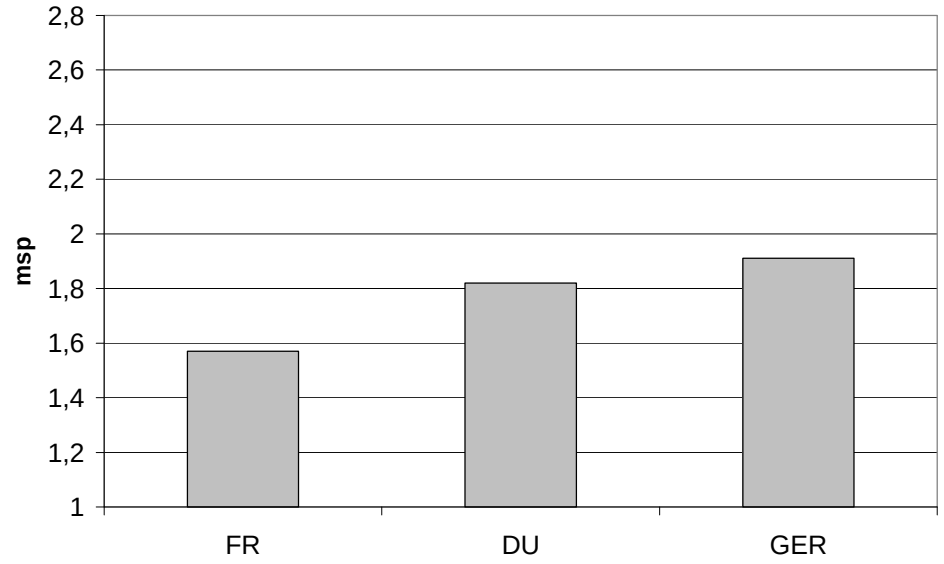

Figure 1b. Total input MSP (mean size of paradigm) in verbs for French, Dutch and German (particle verbs merged)

A2. Syntagmatic morphological richness

We now turn to syntagmatic morphological richness in the input. Figure 2a shows the distribution of inflectional suffixes in nouns, for the three languages. In the morphologically "poorest" language French, nouns never take a suffix; i.e. all nouns in the input are suffixless base forms. In Dutch and in German, 9-10% of all nouns take one suffix, e.g., Dutch Pl. *tafel-s* 'tables'.

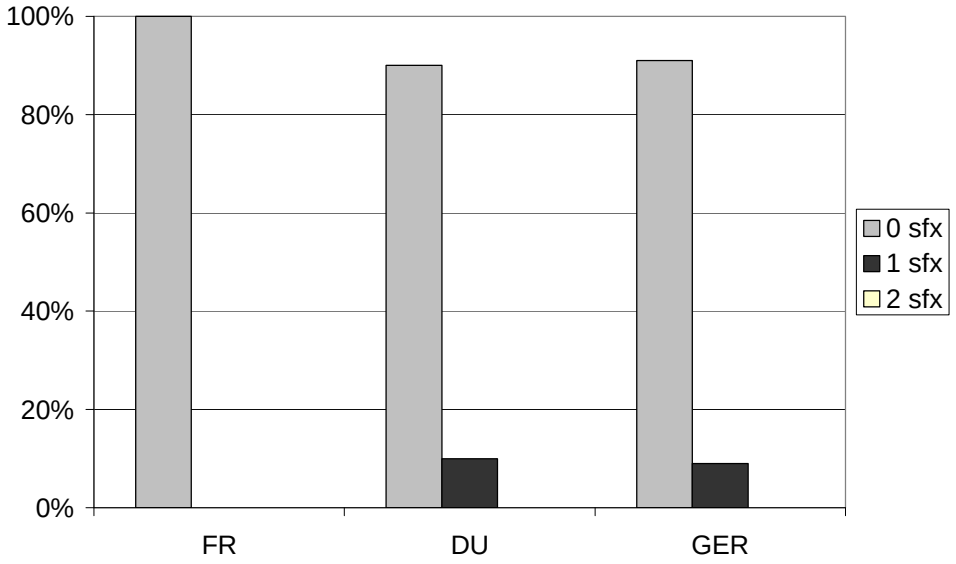

Figure 2a. Distribution of suffixes in nouns for French, Dutch and German

Examples (1)-(3) below illustrate the different form types for each of the three languages (the number of suffixes is given in parentheses).

(1) French
Sg. *fleur* /flœr/ ‘flower’ (0)
Pl. *fleurs* /flœr/ ‘flowers’ (0)

(2) Dutch
Sg. *tafel* ‘table’ (0)
Pl. *tafel-s* ‘tables’ (1)

(3) German
Sg. *Stift* ‘pencil’ (0)
Pl. *Stift-e* ‘pencils’ (1)
Dat. Pl. *mit Stift-e-n* ‘with pencils’ (2)

Also in the distribution of inflectional suffixes in verbs, there are clear differences between French and the two Germanic languages (Figure 2b). In French, 75% of the verb forms are suffixless; in Dutch, base forms and 1 suffix have about the same frequency (52% and 48% respectively); in German, verbs with 1 suffix are more frequent than suffixless ones (60% vs. 40%). This difference between French and the two Germanic languages is mainly due to the fact that in French, all persons in the singular are expressed by the base form, e.g., PRES:Sg *joue(s)* /ʒu/ ‘I/you/he/she play(s)’, whereas in German and in Dutch, 2nd and 3rd person singular, i.e. verb forms which occur very frequently in the input, take 1 suffix, e.g., Dutch PRES:2/3Sg *werk-t* ‘you/he/she work(s)’.

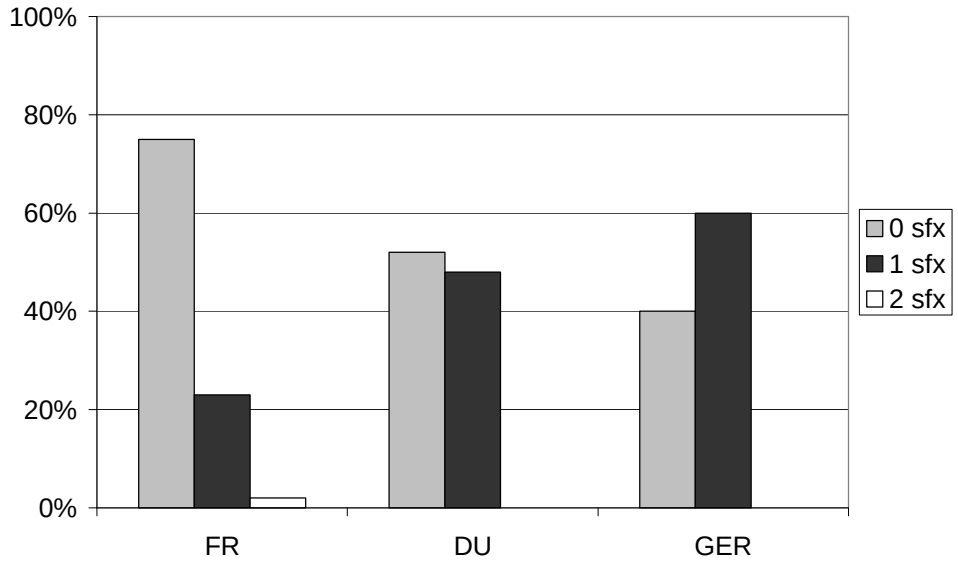

Figure 2b. Distribution of suffixes in verbs for French, Dutch and German

Examples (4)-(6) below illustrate the different form types for each of the three languages (the number of suffixes is given in parentheses).

(4) French
PRES:Sg *joue(s)* /ʒu/ ‘I/you/he/she play(s)’ (0)
INF *jouer* /ʒu-e/ ‘to play’ (1)
FUT:2/3Sg *jouera(s)* /ʒu-r-a/ ‘you/he/she will play’ (2)

(5) Dutch
PRES:1Sg *werk* 'I work' (0)
PRES:2/3Sg *werk-t* 'you/he/she work(s)' (1)
INF *werk-en* 'to work' (1)

(6) German
IMP:Sg *lauf!* 'run!' (0)
PRES:3Sg *läuf-t* 'he/she runs' (1)
INF *lauf-en* 'to run' (1)

These differences in syntagmatic morphological richness observed between French and the two Germanic languages can also be expressed in terms of the average number of inflectional suffixes in the input (Variable A2). The average number of suffixes for each of the three languages is given in Table 5. As can be seen, both in nouns and in verbs, the average number of suffixes is lower in French than in the two Germanic languages.

Table 5. Average number of suffixes in French, Dutch and German

	Nouns	Verbs
FR	0.00 *(Input Emm 0.00, Input Sop 0.00)*	0.27 *(Input Emm 0.27, Input Sop 0.27)*
DU	0.10 *(Input Jol)*	0.52 *(Input Jol)*
GER	0.10 *(Input Jan 0.12, Input Kat 0.08)*	0.61 *(Input Jan 0.63, Input Kat 0.59)*

B. Transparency

The next input variable is transparency. We investigated two different types of input transparency: word transparency (Variable B1) and form transparency (Variable B2).

Examples (7)-(9) below illustrate transparent and opaque noun and verb forms for each of the three languages. Note that only substantial stem changes which involve a stem vowel change and/or a consonant alternation (including suppletion) were qualified as opaque; slight stem changes which involve a change in word-final devoicing and/or a vowel lengthening were qualified as transparent.

(7) French
a. transparent: Noun: Sg. *arbre* /arbr/ – Pl. *arbres* /arbr/ 'trees'
Verb: INF *parler* /parl-e/ 'to speak' – PRES:Sg *parle(s)* /parl/ 'speak(s)'
INF *li-re* /lir/ 'to read' – PRES:3Pl *li-sent* /liz/ 'read'
b. opaque: Noun: Sg. *animal* /animal/ – Pl. *animaux* /animo/ 'animals'
Verb: INF *venir* /vən-ir/ 'to come' – PRES:Sg *viens/vient* /vjɛ̃/ 'come(s)'
INF *fai-re* /fɛr/ 'to do' – PRES:3Pl *font* /fɔ̃/ 'do'

(8) Dutch
a. transparent: Noun: Sg. *spin* /spɪn/ – Pl. *spinn-en* /spɪn-ə(n)/ 'spiders'; *bed* /bɛt/ –
Pl. *bedd-en* /bɛd-ə/ 'beds'; Sg. *hof* /hɔf/ - Pl. *hov-en* /ho:v-ə/ 'gardens'
Verb: INF spelen /spel-ən/ 'to play' – PRES:3Sg *speel-t* /spel-t/ 'plays' –
PP *ge-speel-d* /g-speel-t/ 'played'

b. opaque: Noun: Sg. *schip* /sxɪp/ – Pl. *schep-en* /sxep-ə/ ‘ships’
Verb: INF *slap-en* /slap-n/ ‘to sleep’ – PAST:3Sg *slip* /slip/ ‘slept’;
INF *drinken* /drink-n/ ‘to drink’ – PP *gedronken* /g-trɔnk-n/ ‘drunk’

(9) German
a. transparent: Noun: Sg. *Bus* /bʊs/ – Pl. *Buss-e* /bʊs-ə/ ‘buses’; Sg. *Hund* /hʊnt/–
Pl. *Hund-e* /hʊnd-ə/ ‘dogs’; Sg. *Bub* /bʊp/ – Pl. *Bub-en* /bu:b-ən/ ‘boys’
Verb: INF *spiel-en* /ʃpi:l-ən/ ‘to play’ – PRES:3Sg *spiel-t* /ʃpi:l-t/ ‘plays’ –
PP *ge-spiel-t* /gə-ʃpi:l-t/ ‘played’
b. opaque: Noun: Sg. *Hut* /hu:t/ – Pl. *Hüt-e* /hy:t-ə/ ‘hats’; Sg. *Zug* /tsu:k/ –
Pl. *Züg-e* /tsy:g-ə/ ‘trains’
Verb: INF *schlaf-en* /ʃla:f-ən/ ‘to sleep’ – PRES:3Sg *schläf-t* /ʃle:f-t/ ‘sleeps’;
INF *trinken* /trɪnk-ən/ ‘to drink’ – PP *getrunken* /gə-trʊnk-ən/ ‘drunk’

As shown in Figure 3a, French and Dutch nouns are very little affected by morphophonological processes that reduce transparency (word and form transparency is at 99-100% in both languages). In German nouns, word transparency is lower (85%); this is due to the German plural system in which certain plural classes show an opacifying stem vowel change (Umlaut), e.g., German Sg. *Hut* /hu:t/ – Pl. *Hüt-e* /hy:t-ə/ ‘hats’; form transparency however rates as high as in the other two languages (99%). This result suggests that nouns tend to show up in the German input in a transparent form, although some of the nouns belong to opaque plural classes.

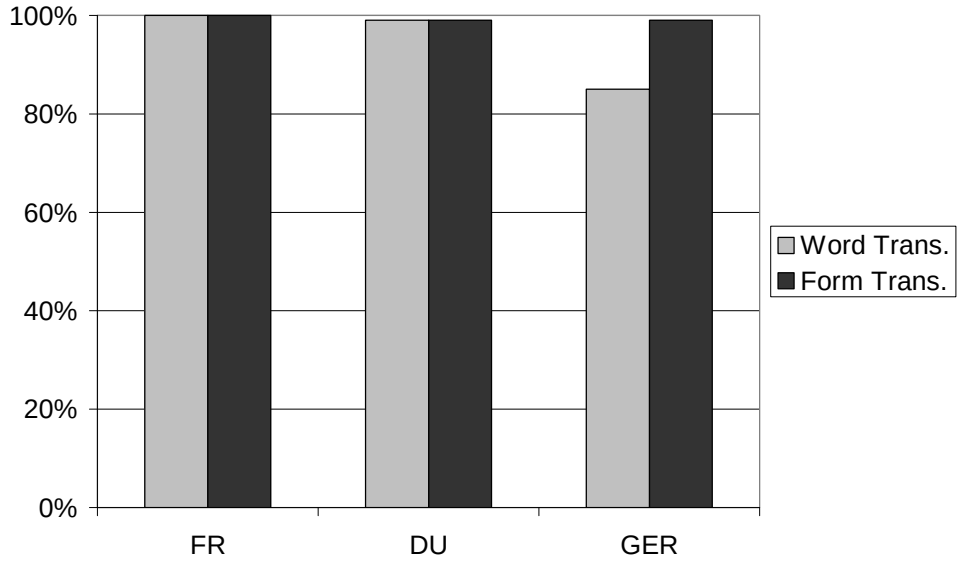

Figure 3a. Word and form transparency in nouns for French, Dutch and German

Figure 3b shows word and form transparency in verbs, for the three languages. For French and German, we conducted two different analyses of form transparency: form transparency for all verbs (V) and form transparency for lexical verbs only (VLEX).

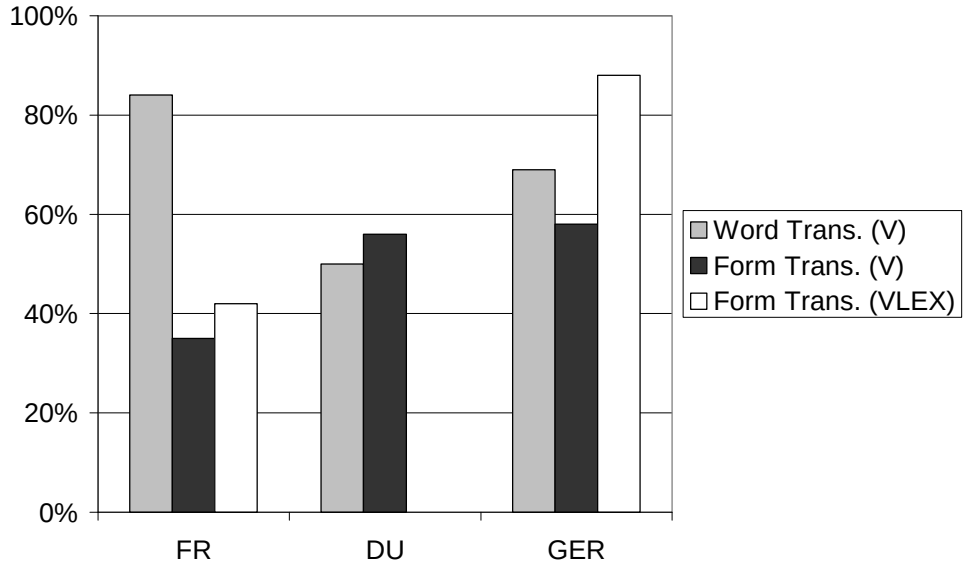

Figure 3b. Word and form transparency in verbs for French, Dutch and German

As can be seen, in French verbs, the difference between word and form transparency is very high: 84% of the French verbs are transparent; however, only 35% of the verb forms actually used in the input are transparent. This difference is related to the frequency difference between 1st and 2nd/3rd conjugation class (macroclass I and II) verbs in French. Macroclass I (e.g., *jouer* 'to play') has the largest number of verbs and most of these verbs are transparent; macroclass II (e.g., *finir* 'to finish'; *prendre* 'to take') is smaller, but it contains verbs which occur very frequently in adult speech, and also in child-directed speech; the great majority of them are opaque (more opaque than Dutch and German strong verbs). In Dutch and German, word and form transparency are at about the same level (Dutch: 50% vs. 56%; German: 69% vs. 58%). Both in French and in German, form transparency is higher if one considers lexical verbs only; this difference is due to the high token frequency of opaque auxiliaries and modal verbs in the two languages.

C. Uniformity

Uniformity (Variable C) was calculated for French and German verbs. The verb categories qualified as uniform in French are Sg. (= zero/base form), 1Sg., 2+3Sg., 1Pl., 2Pl., imperfect, future, conditional, gerundive. German has a smaller number of uniform verb categories (1Sg., 2Sg., 1+3Pl., 2Pl., INF, IMP). This difference is also visible in the input data. As shown in Table 6, the proportion of uniform verb categories is higher in the French input than in the German one.

Table 6. Uniform categories in French and German verbs

	Verbs
FR	72.5% *(Input Emm 71%, Input Sop 74%)*
GER	54.5% *(Input Jan 52%, Input Kat 57%)*

D. Salience

The last input variable investigated is salience. We examined two different types of input salience: phonological-segmental salience (Variable D1) and prosodic salience (Variable D2).

As shown in Table 7, French and the two Germanic languages differ with respect to phonological salience. Both nouns and verbs are phonologically salient in French, because French has no inflectional suffixes with a final reduced vowel. By contrast, they are non-salient in Dutch and German: both languages have no inflectional suffixes with a final full vowel.

Table 7. Phonological-segmental salience of inflectional suffixes in French, Dutch and German

	Nouns	Verbs
FR	100.0% *(Input Emm 100%, Input Sop 100%)*	100.0% *(Input Emm 100%, Input Sop 100%)*
DU	0.0% *(Input Jol 0%)*	0.0% *(Input Jol 0%)*
GER	0.0% *(Input Jan 0%, Input Kat 0%)*	0.0% *(Input Jan 0%, Input Kat 0%)*

The same holds true for prosodic salience (Table 8). Both nouns and verbs are prosodically salient in French, because word-final inflectional suffixes are always stressed in French. By contrast, they are non-salient in Dutch and German where word-final inflectional suffixes are always unstressed.

Table 8. Prosodic salience of inflectional suffixes in French, Dutch and German

	Nouns	Verbs
FR	100.0% *(Input Emm 100%, Input Sop 100%)*	100.0% *(Input Emm 100%, Input Sop 100%)*
DU	0.0% *(Input Jol 0%)*	0.0% *(Input Jol 0%)*
GER	0.0% *(Input Jan 0%, Input Kat 0%)*	0.0% *(Input Jan 0%, Input Kat 0%)*

3.4. Development of mean size of paradigm in child speech

We now turn to the development of mean size of paradigm (MSP) in child speech, which constitutes the dependent variable of our research (see Xanthos & Laaha, this volume).

Figure 4a shows the development of MSP in nouns for the three languages. As can be seen, in the French-speaking children, MSP in nouns shows no increase at all; in the Dutch-speaking child, it shows a very slight increase; and in the German-speaking children, an increase, similar to that in the Dutch-speaking child.

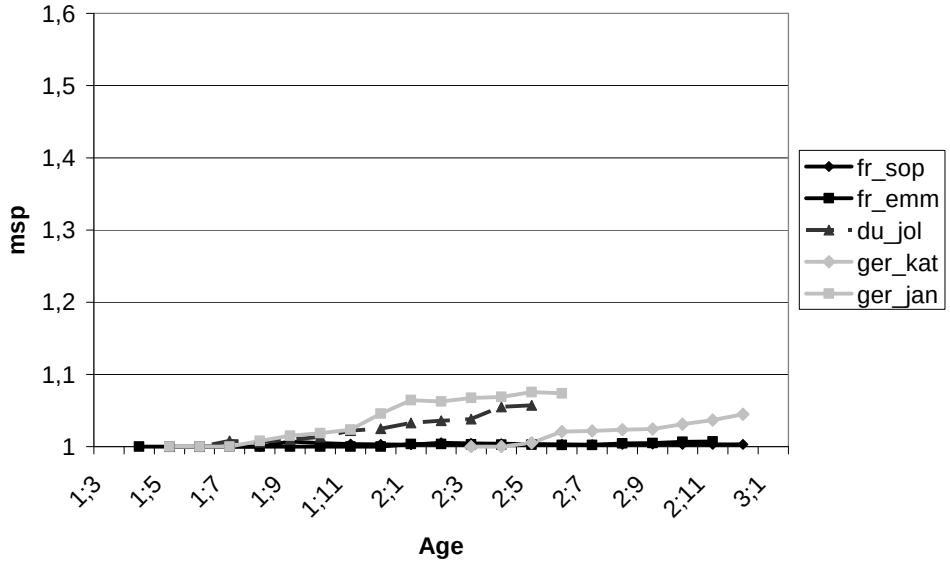

Figure 4a. Development of mean size of paradigm in nouns for French, Dutch and German

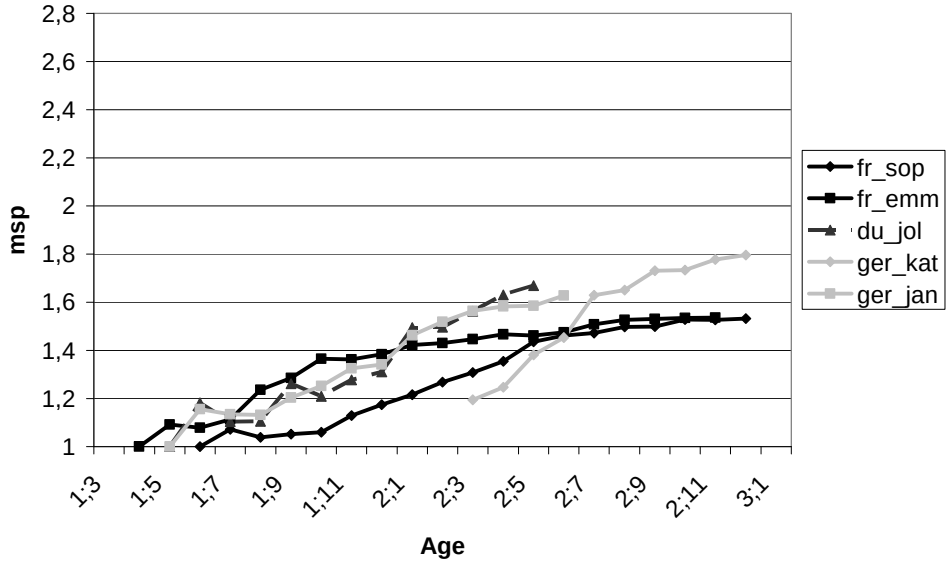

Figure 4b. Development of mean size of paradigm in verbs for French, Dutch and German

Figure 4b shows the development of MSP in verbs for the three languages. As can be seen, MSP shows a steeper rise (i.e. a faster development) in verbs than in nouns in all three languages. In the French-speaking children, MSP in verbs shows a certain increase with age; in the Dutch-speaking child, it shows a slightly higher increase; and in the German-speaking children an increase similar to that in the Dutch-speaking child.

A comparison of the development of MSP in the seven Dutch children suggests that individual differences of speed of development within one and the same language are quite restricted (see Figures 5a and 5b).

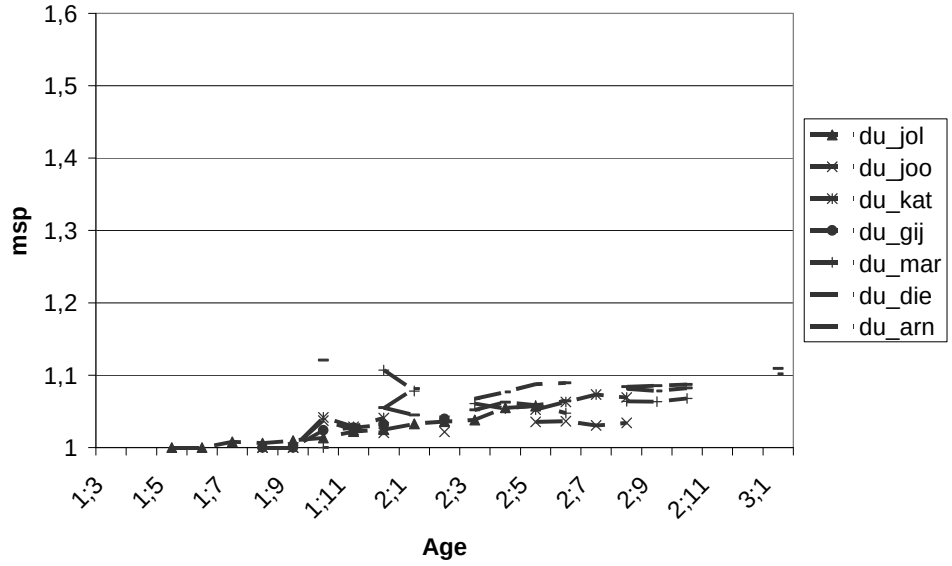

Figure 5a. Development of mean size of paradigm in nouns for the seven Dutch corpora

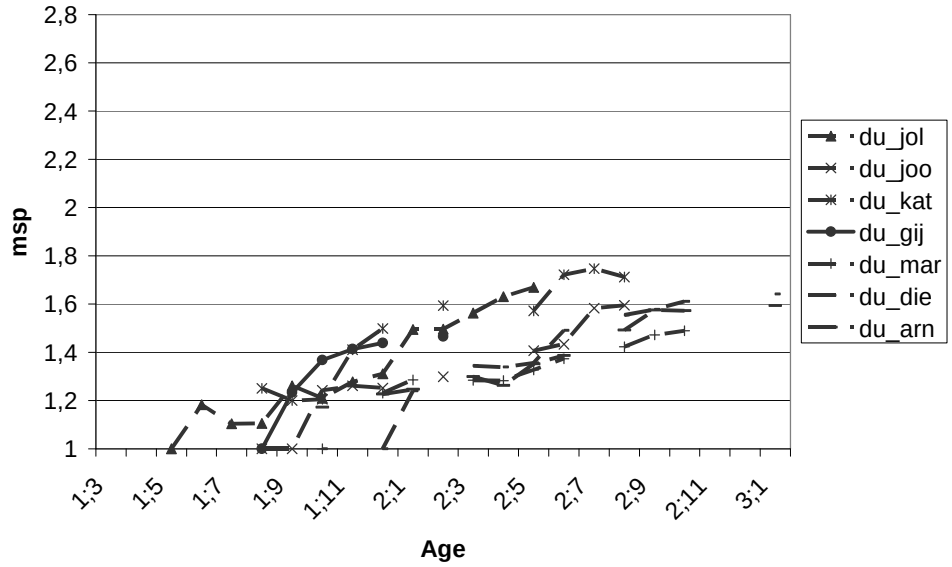

Figure 5b. Development of mean size of paradigm in verbs for the seven Dutch corpora

To summarize, our results suggest that the French-, Dutch- and German-speaking children investigated show a faster development of MSP in verbs than in nouns. Furthermore, the children acquiring the typologically closely related languages Dutch and German show a faster development than the children acquiring French. These differences in speed of development seem to fit to differences in morphological richness in the input, as presented in section 3.3.A above. Whether this impression can be confirmed statistically, will be reported in the chapter "General results" (Xanthos, this volume).

4. Strongly inflecting languages: Russian, Croatian, and Greek

Ursula Stephany, Maria D. Voeikova, Anastasia Christofidou, Natalia Gagarina, Melita Kovačević, Marijan Palmović, and Gordana Hrzica

4.1. Typological characteristics

The languages to be dealt with in this section are three strongly inflecting Indo-European languages, namely West-Slavic Russian, South-Slavic Croatian, and Greek. In the three languages, verb morphology is richer than noun morphology, although the difference is much bigger in Greek than in Russian and Croatian.

While the case-number system of the two Slavic languages is very rich, that of Greek is relatively poor. Greek has three grammatical cases (Nom, Gen, Acc), whereas Russian and Croatian possess four core cases (Nom, Gen, Dat, Acc) and two peripheral ones (Instr, Prep). Besides, Greek and Croatian, but not Russian, distinguish a vocative form of certain nouns.[1] Since Russian and Croatian have no article, the grammatical categories of the noun are foremost marked synthetically, while, in Greek, they are expressed both synthetically (on the noun) and periphrastically (on the determiner) (see Voeikova, 2002).

As far as language acquisition is concerned, the degree of syncretism of case forms is important. With the exception of the class of masculine nouns ending in *-os*, Greek nouns distinguish only two case forms in the singular as well as the plural (for classification of the Greek noun see also Christofidou, 2003). In the singular, 'diptota' nouns either contrast an unmarked Nom/Acc form with a marked genitive form (feminine and neuter nouns) or an unmarked Acc/Gen form with a marked nominative (masculine nouns not ending in *-os*). The 'triptota' nouns ending in *-os* show a threefold distinction between the nominative, accusative and genitive (besides the vocative). There is much less syncretism in the two Slavic languages as compared to Greek. While, in Russian, productive nouns distinguish five case forms to express six case categories in the singular as well as the plural, unproductive ones may be limited to three different case forms in each number. In the singular, feminine nouns use the same form for the dative and the 'prepositional' case, whereas inanimate masculine nouns and neuter ones have the same form for the nominative and the accusative. Animate masculine nouns, however, do not distinguish between the accusative and the genitive singular. In the plural, animate nouns show syncretism of the genitive and the accusative, while inanimate ones do this for the nominative and the accusative. In contrast to nouns, verb forms have no syncretism at all in Russian and very little in Greek and Croatian.

Whereas the inflectional system of the noun is very similar in the two Slavic languages, there is an important difference in the inflection of the verb. Croatian comprises both synthetic and periphrastic past tense forms, while Russian is limited to synthetic past forms. Besides, both Slavic languages possess (imperfective) present and perfective as well as imperfective (synthetic) past forms.[2] The perfective future is synthetically formed in both of these languages, whereas the imperfective future is expressed periphrastically. The distinction between the perfective and the imperfective aspect is expressed by the choice of a perfective vs. an imperfective verb in the past and the future. Greek differs from the Slavic languages by its strong grammaticization of verbal aspect. Nearly all Greek verbs oppose a perfective to an imperfective grammatical stem form. Greek also has a rich temporal system comprising both

[1] In Russian, there is a so-called 'colloquial vocative' the use of which is optional and limited to the *a*-class of feminine and masculine nouns.

[2] Although, in Croatian, the past is most commonly expressed by the perfect in everyday speech, this tense is not taken into consideration in the present study because it is periphrastically formed.

synthetically and periphrastically expressed tenses. While the (imperfective) present and the perfective and imperfective past is expressed by synthetic verb forms, the future tense and the subjunctive mood are marked by future vs. modal particles added to the perfective or imperfective verb forms carrying a non-past ending. Thus, while Greek combines a strongly grammaticized aspectual category with a rich tense system, Croatian possesses a rich tense system but less strongly grammaticized aspectual distinctions. Russian resembles Croatian as far as the grammatical status of aspect is concerned, but has a poorer temporal system.

In order to oppose the perfective and the imperfective aspect in the past, Greek uses two grammatical forms of the same lexical item (Example 1a) whereas, in Russian, this distinction is expressed by the choice of a perfective vs. an imperfective verb (Example 1b). In Croatian, it is rare that an imperfective verb may also form a perfective past (aorist) (Example 1c). Normally, perfective or imperfective verbs are used for these two forms, respectively.[3] According to the foregoing, it is to be expected that the number of verb lemmas in the input will be higher in Russian and Croatian than in Greek.

(1)
a. Greek: *ipj-a* (drink:PFV-PAST:1S) 'I drank'
e-pin-a (AUGM-drink:IPF-PAST:1S) 'I drank'
b. Russian: *pi-l-a* (drink-PAST-FEM:SG) 'I/you/she drank'
vy-pi-l-a (PFV-drink-PAST-FEM:SG) 'I/you/she drank up'
c. Croatian: *pi-jah* (drink-IPF:PAST:1S) 'I was drinking'
pi-h (drink-PFV:PAST:1S) 'I drank'

In the three languages, the category of aspect dominates tense since it not only occurs in different moods but also narrows down the choice of tense in the indicative mood: Whereas perfective verbs (RUS, CRO) or perfective verb forms (GRK) cannot be used in the present tense, past and future forms (as well as subjunctive forms in GRK) are preferentially perfective in early child-directed speech as well as in child language.

4.2. The data

Two children have been investigated for Russian and one child each for Greek and Croatian (see Table 1). As with the other children of the Project, tape recordings started with the onset of verbal production and lasted until about the second half of their third year.[4] While the two Russian children and the Greek boy use twice as many nouns than verbs overall (N/V: Liza 2.645, Filipp 1.985, Christos 2.315), the Croatian girl uses twice as many verbs than nouns (V/N: Antonia 2.404). This V/N ratio corresponds to that of one of her caregivers.

Table 1. Characteristics of the child data

Child	Age range		# Child tokens (nouns + verbs)
	Original	Aligned	
Liza (RUS)	1;6 - 2;5	1;8 - 2;5	1936 + 732
Filipp (RUS)	1;4 - 2;8	1;4 - 2;8	4292 + 2162
Antonia (CRO)	1;3 - 2;8	1;5 - 2;8	1762 + 4136
Christos (GRK)	1;7 - 2;6	1;8 - 2;6	3750 + 1620

[3] The only synthetic past forms, imperfect (imperfective past) and aorist (perfective past) in modern Croatian are rarely used, item-based and restricted to a small number of situations. In everyday speech they are mostly replaced by the periphrastic perfect.
[4] As in the entire Project, the present data was transcribed and coded in the framework of the CHILDES Project (MacWhinney, 2000).

Since the input has not yet been entirely coded for the Greek boy, the analyses of the Greek data are based on merely 900 utterances each (Table 2). Three hundred utterances each were selected from the first, middle and last month of the observational period. Since the data of the two Russian children and the Croatian girl have been entirely coded, it was possible to include all of it into the computation of mean size of paradigm (see section 4.3.A1 below). Computation of the other variables is also based on a sample of 900 utterances.

Table 2. Characteristics of the input data

Child	Age range	# Input tokens (nouns + verbs) *entire sample*	# Input tokens (nouns + verbs) *reduced sample*
Liza (RUS)	1;6 - 2;5	4977 + 4211	553 + 500
Filipp (RUS)	1;4 - 2;8	9316 + 7394	771 + 676
Antonia (CRO)	1;3 - 2;8	5505 + 10795	447 + 803
Christos (GRK)	1;7 - 2;6		733 + 927

Both Russian children are growing up in upper middle-class families living in St. Petersburg. Although standard Russian is spoken in both families, the mothers' linguistic behavior towards their children is quite different, at least as far as this is evident from the tape recordings. While Filipp's mother tries to elicit speech from her only son by encouraging him to repeat her own utterances, Liza's mother concentrates more on explaining things to her daughter (a second child) and discussing them with her. Liza and Filipp use rather different acquisition strategies. While Liza is a typical "referential" child with a relatively low MLU and correct morphological marking of her utterances, Filipp uses a repetitional strategy with longer but ungrammatical utterances.[5] Although we expected the variables studied to differ in the input of the two mothers, what we found were strikingly similar input features instead.

The Croatian girl, an only child at the time of recording, is growing up in an upper middle class family in Zagreb speaking the Zagreb Stokavian dialect, one of the two major dialects of the Croatian capital. While Antonia's mother speaks the Zagreb Stokavian dialect to her daughter, in the child's interactions with her grandparents living in the same household there are also strong elements of the Zagreb Kajkavian dialect. Antonia was recorded in her home during spontaneous interactions with her parents and grandparents. There were about three recording sessions per month lasting about 45 min. each.

The Greek boy Christos was the only child of a Greek upper middle class family at the time of observation growing up monolingually in Athens. He has been observed from the age of 1;7.11 to the age of 3;0 at (almost) weekly intervals. Each of the recordings of the boy's interaction with (mainly) his mother or grandmother while he was playing or looking at picture books is about 20 min. long. The data analyzed for the present study consists of approximately 50 such recordings from 1;7 to 2;6.

4.3. Independent variables of the input

A. Morphological richness

Morphological richness is studied both on the paradigmatic and the syntagmatic axis. Paradigmatic richness is determined by the number of form-type categories of nouns and verbs in each of the three languages as well as by mean size of a paradigm calculated

[5] Direct repetitions, citations, frozen forms, utterances consisting of *yes* or *no* were excluded from the analysis.

separately for nouns and verbs. Syntagmatic richness is demonstrated by the average number of suffixes of nouns and verbs.

A1. Paradigmatic morphological richness

The paradigmatic richness of nouns in both the standard language and child-directed speech is much higher in the two Slavic languages than in Greek. The number of grammatical categories occurring in child-directed speech tends to be lower than that of the standard languages. Thus, standard Russian and Croatian possess 12 and 14 case-number forms of nouns, respectively, while only 11 of these are found in the input (Table 3). Certain Greek nouns distinguish up to 7 case-number forms, but only 4 of these occur in the input. The reason is that, in the three languages, the case distinctions of the plural are not fully represented in the input. Besides, in Greek, the inflectionally richest nouns have a minor role to play in the input. Thus, as far as the noun is concerned, the difference between Greek and the two Slavic languages is even greater in the input than in the standard languages.

Table 3. Form-type categories in Russian, Croatian and Greek[6]

	Nouns	Verbs
RUS	*Sg: Nom, Gen; Dat; Acc; Inst; Loc* *Pl: Nom, Gen, Acc, Inst, Loc*	*(Imperf.)Pres/Perf. Fut: 1S, 2S, 3S, 1P, 2P (only Pres.), 3P* *Imperf./Perf. Past: fem:sg, masc:sg, neut:sg, pl* *Infinitive, Imp:Sg* *Passive past participles: fem:sg, masc:sg, pl*
CRO	*Sg: Nom, Gen, Dat, Acc, Loc, Instr, Voc* *Pl: Nom, Gen, Acc, Instr*	*(Imperf.) Pres: 1S, 2S, 3S, 1P, 2P, 3P* *Perf. Past (Aorist): 1S, 2/3S, 1P, 2P* *Infinitive, Imp:Sg* *Past Participles: fem:sg, masc:sg, neut:sg, fem:pl, masc:pl, neut:pl*
GRK	*Sg: Nom, Gen, Acc* *Pl: Nom/Acc*	*Imperf./Perf. Nonpast:1S, 2S, 3S, 1P, 2P, 3P* *Med.pass. 3S, 3P* *Imperf./Perf. Past:2S, 3S, 1P, 2P, 3P* *Imp:Sg; Perfective*

In the three languages, the number of verbal form-type categories is larger than that of the noun. This is especially true of Greek. As with nouns, the number of verbal form-type categories occurring in the input is much lower than that of the three standard languages. The high number of Croatian verb forms as compared to both Russian and Greek is due to the numerous participles that have gender distinctions in both the singular and plural while Russian distinguishes gender only in the singular.

The following continua of paradigmatic morphological richness for nouns and verbs can be established for the input of the three strongly inflecting languages:

Nouns: Croatian → Russian → Greek
Verbs: Croatian → Russian/Greek

Mean size of paradigm (MSP) represents a second way of measuring morphological richness (Variable A1, see Xanthos & Laaha, this volume). It shows that the caregivers are far from exhausting the inflectional potential of nouns and verbs as evidenced by the input for all lexical items (Table 4). Many nouns occur in a single form only. Thus, in Russian, names of food or drinks either occur in the accusative or the genitive (partitive) and in Greek, many nouns are only used in a kind of all-purpose unmarked nominative-accusative singular (or,

[6] The table is limited to synthetic verb forms. Thus, Russian imperfective future and Croatian perfect have been omitted in spite of their important role in child-directed speech (see fn. 2).

more rarely, plural) form. Even for verbs, the average number of inflectional forms per lemma is slightly below two for the Slavic languages and only between 2 and 3 for Greek.

Table 4. Total input MSP (mean size of paradigm) in Russian, Croatian and Greek

	Nouns	Verbs
RUS	1.37 *(Input Fil 1.38, Input Liz 1.36)*	1.65 *(Input Fil 1.65, Input Liz 1.65)*
CRO	1.39 *(Input Ant)*	1.91 *(Input Ant)*
GRK	1.13 *(Input Chr)*	2.41 *(Input Chr)*

The difference between the 15 to 18 verb form categories which could be distinguished in the Russian and Croatian input and the low value of less than two different forms per verb on average is especially striking. In both Slavic languages, the MSP value is even lower for nouns and is at the same level as in the Germanic languages (see Laaha et al., this volume). This means that in both language types, most nouns occur in only one or two different forms in spite of the fact that many more case-number forms are distinguished in the Russian and Croatian input (11 case-number forms each) than in the Germanic languages. Syncretism of nominal inflection typical of both Slavic languages cannot explain this state of affairs since, otherwise, the MSP values of verbs (without such syncretism) should be significantly higher than that of nouns. In Figures 1a and 1b the values indicated in Table 4 are presented in the form of charts.

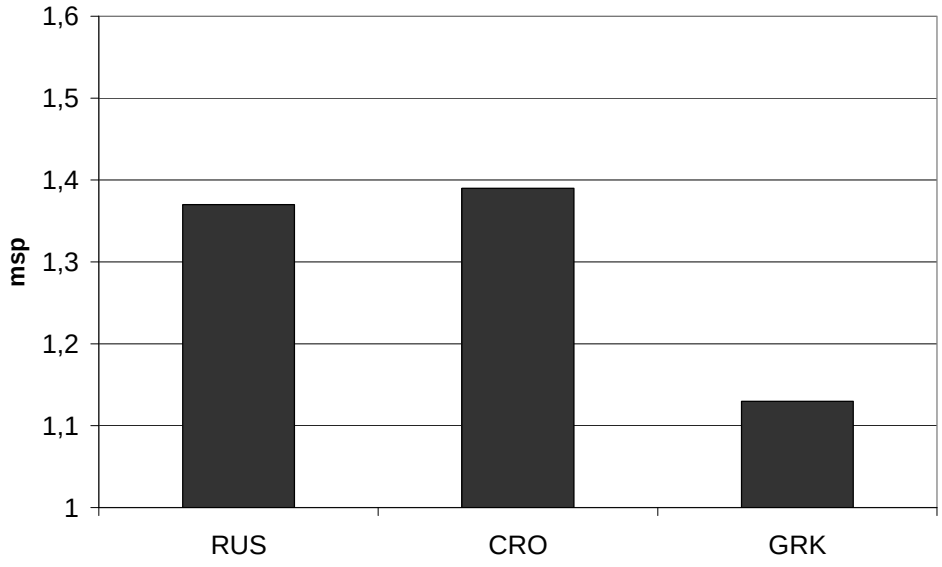

Figure 1a. Total input MSP (mean size of paradigm) in nouns for Russian, Croatian and Greek

Figure 1a demonstrates that, in spite of a big difference between the number of noun form-type categories in the two Slavic languages (11 types) on the one hand and Greek (4 types) on the other, the difference in MSP is very small. An explanation coming to mind is the high degree of syncretism of case-number forms in the three languages which has been mentioned above.

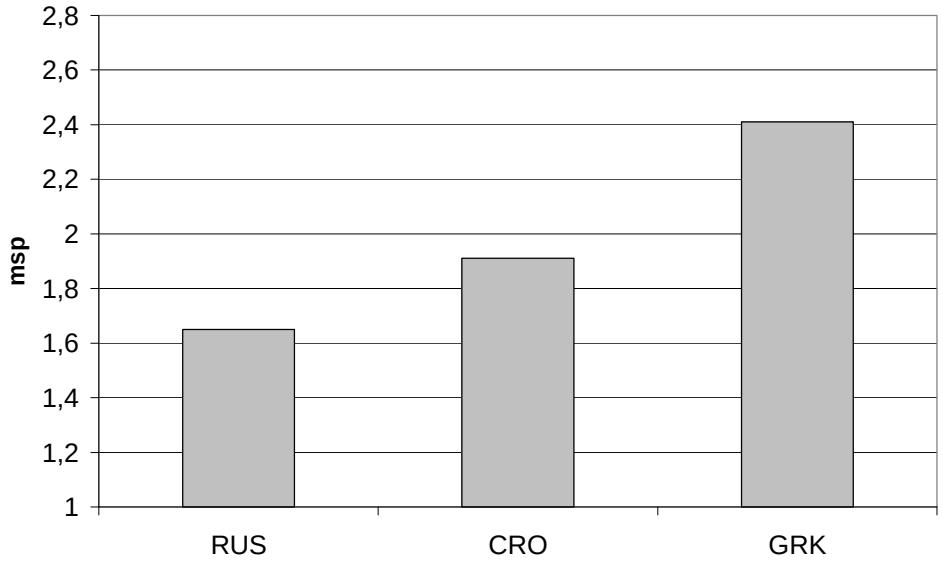

Figure 1b. Total input MSP (mean size of paradigm) in verbs for Russian, Croatian and Greek

In verbs, differences in the MSP do not directly reflect numbers of form-type categories found in the input. Although the number of verbal form-type categories of the two Slavic languages is equal to or higher than that of Greek, MSP is much lower in Russian and Croatian than in Greek (Figure 1b). Hybrid forms such as past participles functioning as adjectives, which do not entirely belong to the verbal paradigm, do influence the total number of form-type categories occurring in the input but have a minor role to play in mean size of paradigm.

A2. Syntagmatic morphological richness

Turning from the paradigmatic to the syntagmatic axis of morphological richness (Variable A2), the verb again rates higher in all three languages than the noun. While in the two Slavic languages nearly every noun carries a suffix, in Greek two thirds of noun tokens either end in a thematic vowel, not having been counted as an inflectional marker in the present study, or are uninflected foreign names (Table 5). The average number of noun suffixes is less than 1 in the Slavic languages because there are masculine nouns like *Sankt Peterburg* or *Zagreb* with a zero suffix in the nominative/accusative. The corresponding value for verb forms is nearly the same in the three languages, with the exception of the Russian girl Liza.

Table 5. Average number of suffixes in Russian, Croatian and Greek

	Nouns	Verbs
RUS	0.80 *(Input Fil 0.79, Input Liz 0.80)*	1.16 *(Input Fil 1.09, Input Liz 1.24)*
CRO	0.92 *(Input Ant)*	1.06 *(Input Ant)*
GRK	0.34 *(Input Chr)*	1.09 *(Input Chr)*

In the three languages, nouns carry at most a single suffix. While inflectionally unmarked noun forms by far prevail in Greek (partly for methodological reasons), most noun tokens have one suffix in both Russian and Croatian (Figure 2a).

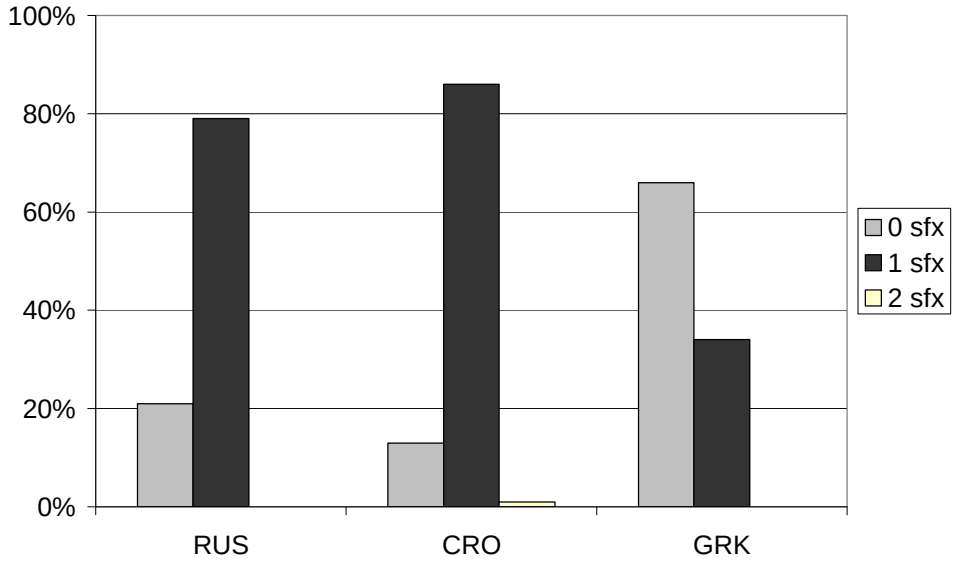

Figure 2a. Distribution of suffixes in nouns for Russian, Croatian and Greek

Although verb forms may carry two suffixes in all three languages, and in Russian even up to three (although these very rarely occur in the input), verb forms with one suffix by far prevail (Figure 2b). The impression that, in Greek, suffixless verb forms are more numerous than those with two suffixes is due to the high number of tokens of the copula (*ine* 'is') which has been counted as a suffixless form in the present study. The higher number of zero endings in Croatian as compared to Russian is explained by the fact that, in Croatian, the third person singular consists of a bare stem.

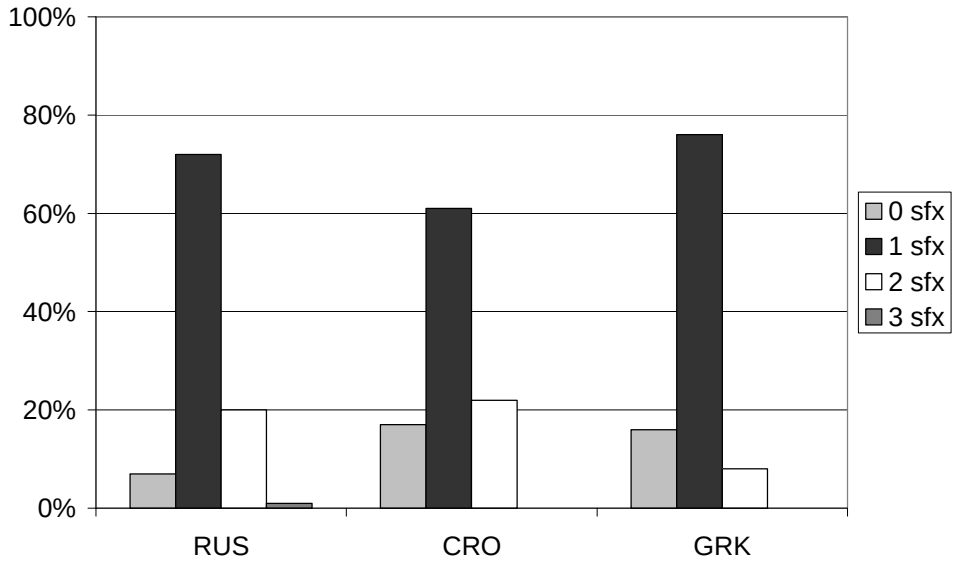

Figure 2b. Distribution of suffixes in verbs for Russian, Croatian and Greek

B. Transparency

Transparency has been calculated on the basis of a reduced sample of 900 utterances of the input. Word transparency (Variable B1) in nouns is very high in both Croatian and Greek but less so in Russian (Figure 3a). This is due to stress shift accompanying noun suffixation on

the one hand and deletion of the last stem vowel (*o* or *e*) of masculine nouns on the other (e.g. *petush-ok-0* 'cock-DIM-NOM:SG', *petush-k-a* 'cock-DIM-GEN:SG'). Form transparency (Variable B2) is even higher than word transparency in the three languages, which shows that caretakers seem to prefer transparent forms even when the lemma as a whole is non-transparent.

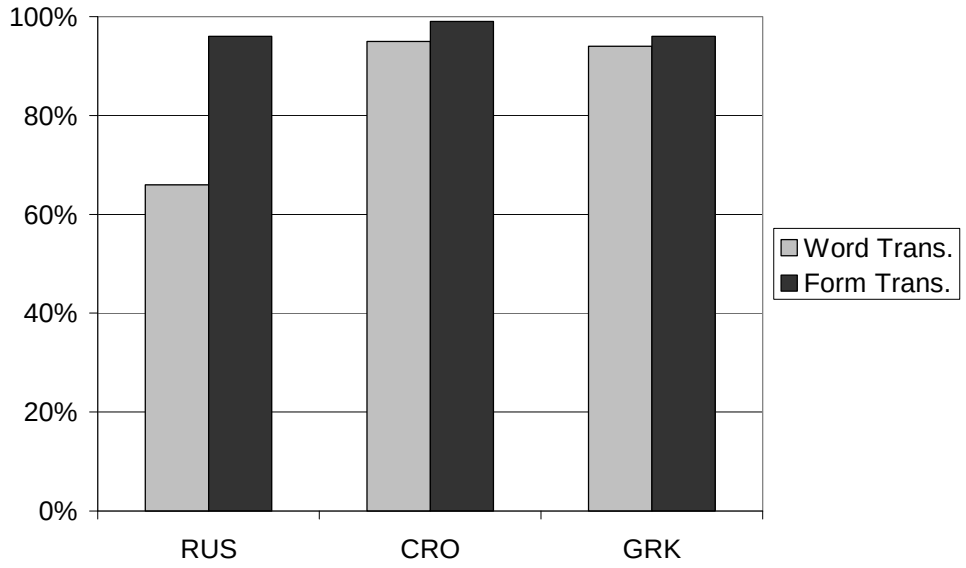

Figure 3a. Word and form transparency in nouns for Russian, Croatian and Greek

Comparing word transparency of nouns to that of verbs, the typological characteristics of inflecting-fusional languages with their high degree of opacity show more clearly in the children's input (Figure 3b). Word transparency of the verb only reaches about 50% in the two Slavic languages (as compared to more than 60% for Russian nouns and more than 90% for Croatian ones) and is much lower in Greek (18% for verbs as compared to more than 90% for nouns). As with nouns, form transparency of verbs is higher than word transparency in Russian and Greek, but not in Croatian, where non-transparent verb forms prevail. If only lexical verbs (VLEX) are taken into consideration, with the highly irregular auxiliaries being excluded, form transparency increases slightly in the three languages.

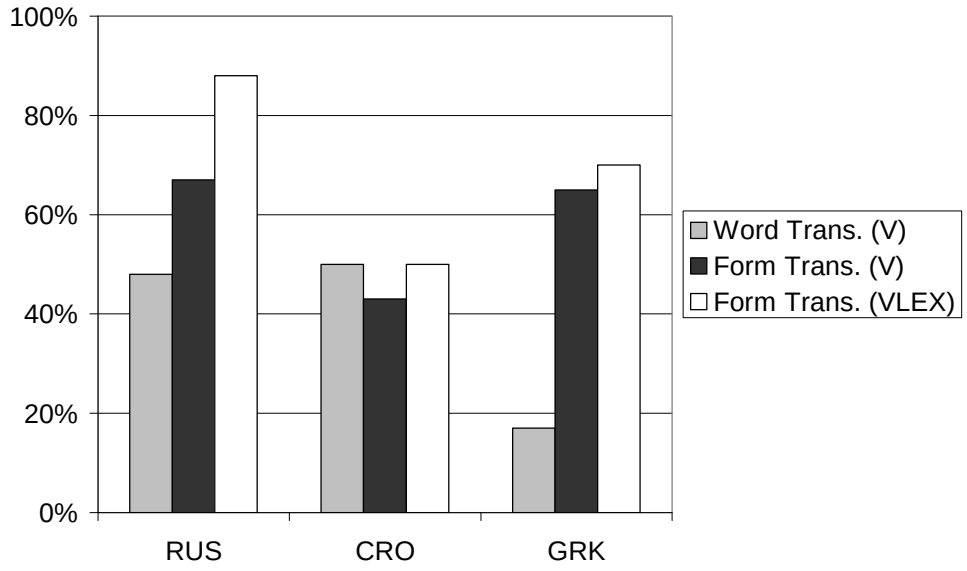

Figure 3b. Word and form transparency in verbs for Russian, Croatian and Greek

C. Uniformity

Uniformly inflected forms (Variable C) have one and the same marker in all inflectional classes. This characteristic is untypical of inflecting languages. Thus, in both Slavic languages, the percentage of uniformly marked grammatical categories in nouns is zero or close to zero (Table 6). The reason is that the only uniform categories (e.g. Russian Dat:Pl and Inst:Pl) very rarely occur in the input. As far as the two Slavic languages are concerned, a reverse tendency to use uniform or transparent forms can be observed in the input: While the latter are preferred, the former occur more rarely than in the standard language. In Greek, nouns show a big percentage of uniformity because most inflected forms are suffixless (in the strict sense of the term, i.e. leaving the thematic vowel out of consideration), see Figure 2a above.

Table 6. Uniform categories in Russian, Croatian and Greek

	Nouns	Verbs
RUS	1.0% *(Input Fil 1%, Input Liz 1%)*	45.0% *(Input Fil 44%, Input Liz 46%)*
CRO	0.0% *(Input Ant)*	81.0% *(Input Ant)*
GRK	68.0% *(Input Chr)*	0.0% *(Input Chr)*

In contrast to nouns, nearly half of Russian verb forms but more than 80% of the respective Croatian forms are uniformly marked. For determining uniformity entire suffixes or suffix sequences representing bundles of grammatical categories were taken into consideration (e.g. *-la* in Russian *bra-l-a* 'take-PAST-FEM:SG', *pe-l-a* 'sing-PAST-FEM:SG'). In this approach, the only uniform categories occurring in the Russian input are the infinitive, all past forms and the present/future first person singular. The amount of uniform verb forms would be higher if partial uniformity typical of the other present/future forms was taken into consideration. Thus, all pres/fut:2S forms end in *-š'* although this consonant is not a morpheme (e.g. *ber'-oš'* 'take-PRES:2S' vs. *bež-iš'* 'run-PRES:2S'). In Greek, entire verb forms comprising an aspect marker and an inflectional ending expressing tense, mood, person and number were taken into consideration. Since aspect markers vary to a much higher degree than inflectional endings, there were no uniform category bundles in the verbs of this language. Matters would change if the inflectional endings were looked at separately (e.g. *pe-s-i* 'fall-PFV-NONPAST:3S', *aghap-is-i* 'love-PFV-NONPAST:3S').

D. Salience

To determine phonological-segmental and prosodic salience (Variables D1 and D2) only forms with inflectional endings containing a vowel have been taken into consideration. Such forms constitute very different subtotals of nouns and verbs in Greek as compared to Russian. While, in Greek, salience concerns only a small class of nouns[7] but a big class of verbs, the reverse is true for Russian. In Russian, the half-open vowels /e/ and /o/ are reduced or shifted to high /i/ and low /a/, respectively, in unstressed syllables. These vowels have been considered to be non-salient irrespective of the inflectional endings being stressed or unstressed. In Croatian and Greek, all vowels occurring in inflectional suffixes are unreduced. This leads to the values of phonological-segmental salience of inflectional suffixes represented in Table 7.

[7] The reason is that in the present study nouns ending in a thematic vowel have been disconsidered.

Table 7. Phonological-segmental salience of inflectional suffixes in Russian, Croatian and Greek

	Nouns	Verbs
RUS	88.5% *(Input Fil 85%, Input Liz 92%)*	73.5% *(Input Fil 74%, Input Liz 73%)*
CRO	100.0% *(Input Ant)*	100.0% *(Input Ant)*
GRK	100.0% *(Input Chr)*	100.0% *(Input Chr)*

Prosodically salient, i.e. stressed inflectional endings do not exist in Croatian and are rather infrequent in Russian and Greek (Table 8). Thus, in Croatian, all inflectional endings are unstressed but salient. The relatively high value of 27% for Greek nominal suffixes containing a vowel is based on only 19% of all nouns, however (see fn. 7). Therefore, this value was excluded from the general statistic analyses of our research (see Xanthos & Laaha, this volume, Table 2a).Unlike the other independent variables, there is a remarkable difference in prosodic salience of inflectional endings of verbs in the input of the two Russian children. However, more endings are stressed in verbs than in nouns in Russian. It must be kept in mind that the seemingly opposite situation in Greek is based on very different subtotals of nouns and verbs.

Table 8. Prosodic salience of inflectional suffixes in Russian, Croatian and Greek

	Nouns	Verbs
RUS	8.0% *(Input Fil 8%, Input Liz 8%)*	37.0% *(Input Fil 45%, Input Liz 29%)*
CRO	0.0% *(Input Ant)*	0.0% *(Input Ant)*
GRK	[27.0%] *[(Input Chr)]*	11.0% *(Input Chr)*

There is some evidence that vowel reduction may slow down the acquisition of case forms in Russian as compared to Polish (Smoczyńska, 1985; Slobin, 1997) and Lithuanian (Voeikova & Savickienė, 2001). Russian inflectional endings with a 'weak' vowel (/e/ or /o/) have low 'cue validity' (Kempe & MacWhinney, 1998). However, the functional distinction between syncretistic case forms resulting from vowel reduction do not play an important role in some of their uses at least. Thus, the distinction between the dative expressing benefactive and the genitive expressing possession is functionally not important in examples such as *kniga teti* (book aunt-GEN/DAT:SG) 'the aunt's book' or 'the book for the aunt'. In other cases, the functions of syncretistic forms may be disambiguated by prepositions (e.g. *iti k tet-i* 'go to aunt-DAT:SG' vs. *byt' u tet-i* 'be at aunt-GEN:SG'). Thus, a low percentage of salient case forms may be quite sufficient from a functional point of view.

The inflectional endings of Russian verbs are prosodically more salient than those of nouns (37% vs. 8%). Since consonants are more important than vowels for distinguishing Russian verb forms we may predict that prosodic salience will be of minor relevance for the development of verb paradigms.

4.4. Development of mean size of paradigm in child speech

In this section, children's development of mean size of inflectional paradigms (MSP) in nouns and verbs will be related to the variables of child-directed speech studied in the preceding section. The question is in how far differences and similarities of the input variables may

influence the Slavic and Greek children's inflectional development, i.e. their development of verb and noun paradigms.

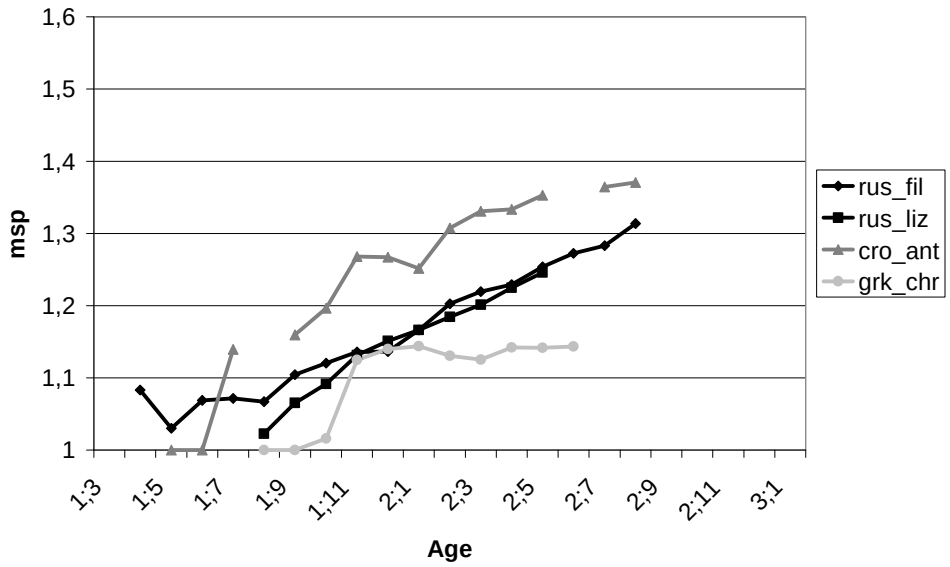

Figure 4a. Development of mean size of paradigm in nouns for Russian, Croatian and Greek

The development of MSP for nouns ranges from 1 to little more than 1.1 for the Greek child and to a value between 1.3 and 1.4 for the two Slavic children who were observed beyond the first half of their third year (Figure 4a). Except for the Russian boy Filipp, the curves rise steeply until an MSP of 1.1 is reached. After that point, development becomes more gradual, except for the Croatian child. The largest difference exists between the Greek boy and the Slavic children, since he stays at an MSP of less than 1.2, while the Slavic children develop further. The Russian boy Filipp crosses the 1.3 line of MSP only by 2;9, but the Croatian girl does so already by 2;3. It is important to note that the Russian boy and the Croatian girl reach the MSP of the input shortly before 2;9, while the Greek boy attains the input level already before the end of his second year. This is not surprising, since so little inflectional diversity of nouns exists in the Greek input, namely 4 form-type categories as opposed to 11 in the two Slavic languages. The endpoints of the developmental curves of the three children who were investigated through the second half of their third year roughly coincide with the different MSPs of the input.

While the difference between the development of noun inflection of the Greek child in comparison to the Slavic children may be attributed to differences in the MSP of the input (GRK 1.13, RUS 1.37, CRO 1.39), the reason for the difference between the Croatian child and the Russian ones must be sought elsewhere since the two values for Russian and Croatian are nearly the same. As we saw earlier, the average number of suffixes per noun is higher in Croatian than in Russian and may attract the child's attention to differences between noun endings. Also, word transparency is much higher in Croatian and there is no vowel reduction. This may eventually explain the Croatian girl's faster development of noun inflection.

The general conclusion to be drawn from a comparison of the three strongly inflecting languages as far as noun inflection is concerned seems to be the following: whether there is less or more to acquire, children all get off the ground of one form per lemma equally quickly, although they start at different points in time. It is only after the 1.1 limit has been crossed that development begins to differ depending on the morphological richness of the language being acquired and, maybe, on transparency.

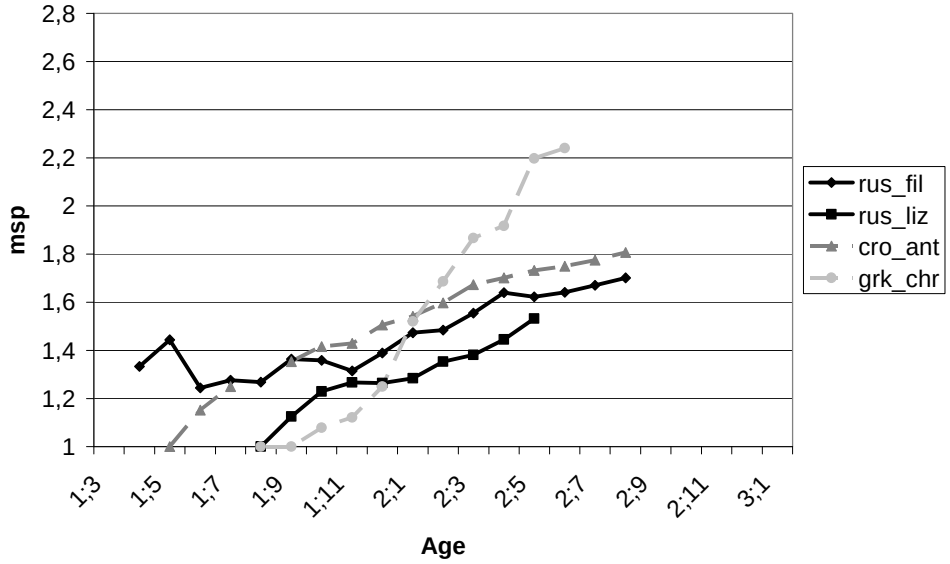

Figure 4b. Development of mean size of paradigm in verbs for Russian, Croatian and Greek

The developmental curves of MSP for verbs (Figure 4b) show both parallels and differences in comparison to those for nouns. All children reach higher MSP values earlier for verbs than for nouns. Again, the Greek boy differs from the group of children acquiring Slavic languages. But this time, it is his MSP which develops most and also more quickly than that of the Slavic children. As with nouns, the endpoints of the developmental curves more or less coincide with the MSP of the input, except for the Russian girl Liza, who was not observed beyond 2;5.

Why should the Greek boy enlarge his MSP with verbs more quickly than the Slavic children? None of the independent variables except MSP can furnish an explanation. MSP is much higher in the Greek input (2.41) than in Croatian (1.91) and Russian (1.65).[8]

Summarizing what we found, it seems that morphological richness of the input rather than any of the other independent variables is crucial not only for the value of MSP attained before the end of the third year but also for speed of inflectional development. Thus, in the domain of the noun, Croatian ranks before Russian, and Greek is last, both as far as the number of form-type categories and MSP are concerned. However, in the domain of the verb, Greek only leads for MSP, followed by Croatian and Russian.

[8] On the early development of the Greek verb see Christofidou & Stephany (2003).

5. Agglutinating languages: Turkish, Finnish, and Yucatec Maya

Ayhan Aksu-Koç[*]*, F. Nihan Ketrez, Klaus Laalo, and Barbara Pfeiler*

5.1. Typological characteristics

Turkish, Finnish and Yucatec Maya are agglutinating languages. Turkish has the purest agglutinating morphology whereas Finnish and Yucatec Maya have some inflecting-fusional properties; Yucatec Maya is also polysynthetic to some extent. Word order is variable on pragmatic basis in all three languages but the basic order is different in each: SOV in Turkish, SVO in Finnish and VOS in Yucatec Maya. Turkish and Finnish are nominative-accusative languages whereas Yucatec displays split ergativity. Turkish and Finnish allow for subject and object ellipsis, in Yucatec the corresponding pronouns usually are not omitted. Turkish and Finnish are characterized by vowel harmony, in Yucatec, however, vowel harmony is limited to certain noun and verb categories.

In Turkish, nominals can receive case, number and possessive marking. In predicate position they also get tense-aspect-mood and subject-verb agreement markers. There is no grammatical gender. Case marking is synthetic and stem based. There are seven cases: nominative, accusative, dative, genitive, locative, ablative, and commutative/instrumental. Except for the nominative which has no phonological realization, cases receive distinct morphological marking on nouns, question words, pronouns, and nominalized forms of the verb or the adjective. The case system consists of a single paradigm that is fully productive and regular. Phonetically homophonous forms in the inflectional paradigm are rare, but exist. The accusative and the 3rd person singular possessive have the same phonological shape (*-i*) on nouns ending in consonants, for example. The inflectional verbal affixes mark negation, tense-aspect-mood, number and person. Turkish derivational morphology is also very rich; voice particles (such as the causative or the passive), when present, are interposed between the verb and the tense-aspect-mood morphemes. Each morpheme is syllabic, the typical stress pattern is word-final though there are exceptions to this pattern in the environment of certain suffixes (Lees, 1961; Sezer, 1983 among others).

Finnish can be typologically characterized as an agglutinating language. Finnish is also synthetic: derivational and inflectional suffixes are attached to the word stem, e.g. *käde+ssä+si* (hand + inessive + 2SG possessive suffix) 'in your hand', *kirja+sto+sta* (book + derivative element *sto* + elative) 'from the library'. However, Finnish also displays some inflecting-fusional properties: there are a few portmanteau-morphemes (e.g. the past participles, both active and passive) and there occur various morphophonological changes when certain suffixes are attached to certain word stems, e.g.

vesi 'water' :	partitive *vet+tä* :	genitive *vede+n* :	illative *vete+en*
veli 'brother' :	partitive *velje+ä*	: genitive *velje+n* :	illative *velje+en*
rikas 'rich' :	partitive *rikas+ta* :	genitive *rikkaa+n* :	illative *rikkaa+seen*
sisar 'sister' :	partitive *sisar+ta* :	genitive *sisare+n* :	illative *sisare+en*
isoin 'biggest' :	partitive *isoin+ta* :	genitive *isoimma+n*:	illative *isoimpa+an*

Colloquial Finnish is less synthetic than standardized written Finnish. For example, possessive pronouns are often used in spoken Finnish instead of the possessive suffixes: the synthetic forms of the written language *kirja+nne* (book + 2PL possessive suffix) 'your book'

[*] The Turkish data were collected for the project "A longitudinal study of the acquisition of Turkish" (project no: 96S0017), that was supported by a grant to A. Aksu-Koç, from the Boğaziçi University Research Fund.

and *talo+mme* (house + 1PL possessive suffix) 'our house' have as their colloquial counterparts the analytical structures *teidän kirja* (2PL pronoun in genitive + book) 'your book' and *meidän talo* (1PL pronoun in genitive + house) 'our house'. The derivational morphology of Finnish is also rich, e.g. the causative suffix *-tta* is productive: *katoaa* 'disappears' > *kadottaa* 'loses, makes disappear', *putoaa* 'falls' > *pudottaa* 'drops, makes fall'.

Yucatec has agglutinating, mildly polysynthetic, but also inflectional-fusional properties. Yucatec is a head-marking language: the verbal complex can function on its own as a complete sentential proposition (Lucy, 1994: 627). Verbs are distinguished between transitive and intransitive verbs according to their argument marking properties. In the verb complex, person, mood, and aspect are represented by sets of inflectional affixes; non-emphatic personal pronouns do not occur in surface structure, and the lexical arguments can be dropped in the adult language. Yucatec has two sets of pronominal affixes, generally known as Set "A" (ergative pronouns) and Set "B" (absolutive pronouns). Both occur within nominal and verbal constructions. Set "A" is characterized by prefixes which mark agent with transitive verbs and possession with nouns, while Set "B" is composed of suffixes which mark the patient with transitive verbs, the subject with intransitive verbs in incompletive aspect, and form equational constructions with nouns. In Yucatec, nouns can receive number and possessive marking. Plural marking is optional; the singular is marked by numerals with classifiers. There is no grammatical gender. The noun phrase can be modified by demonstratives or by enumeration with numeral classifiers. The numeral "one" along with the appropriate classifier can be used to indicate indefinite reference. In predicate position nouns get tense-aspect-modality and optionally subject-verb agreement markers.

5.2. The data

The data analyzed are taken from one child acquiring Turkish, two children acquiring Finnish and one child acquiring Yucatec Maya as their first language.

The Turkish data come from a monolingual Turkish child of middle class background, Deniz, recorded longitudinally between the ages 1;3.3 and 2;0.4. Recordings were done approximately twice a month and each of them was about 20 minutes long. During the sessions, the child was engaged in various natural everyday activities.

The Finnish data come from two monolingual Finnish children of middle class background, one boy (Tuomas) and one girl (Tuulikki). Recordings have been made once or twice a month and each of them was about 30 minutes. During the sessions, the child was engaged in everyday activities.

The Yucatec data come from Armando, a monolingual Yucatec child who is raised in an indigenous peasant family in Yalcobá, a small village in the East of the state of Yucatán in Mexico. The child was recorded at home, twice a week, in interaction with his mother, grandmother, aunt, the observer and other children. Recording situations are characterized by everyday situations and free/toy play.

All the child utterances were transcribed and coded morphologically according to the CHAT conventions of CHILDES (MacWhinney, 2000). For the input, three sessions corresponding to three developmental points were sampled.

Table 1 presents the age ranges of the children studied, as well as the number of noun and verb tokens in the output.

Table 1. Characteristics of the child data

Child	Age range		# Child tokens (nouns + verbs)
	Original	Aligned	
Deniz (TURK)	1;3 - 2;0	1;6 - 2;0	1662 + 2230
Tuomas (FINN)	1;7 - 2;3	1;7 - 2;3	1342 + 1286
Tuulikki (FINN)	1;7 - 3;0	1;7 - 2;10	1584 + 2750
Armando (YUC)	2;0 - 3;0	2;0 - 3;0	1439 + 1684

Table 2 shows the number of noun and verb tokens in the input. For each of the three languages, the input analysis was carried out on a reduced sample of maternal speech consisting of three sessions corresponding to three developmental points in the child's speech, the beginning, the point where there is a burst in morphological marking, and the end. Furthermore, since younger children are often taken care of by their older siblings, the Yucatec input data consist of few adult expressions as well as input of siblings older than 12 years.

Table 2. Characteristics of the input data

Child	Age range	# Input tokens (nouns + verbs) *reduced sample*
Deniz (TURK)	1;3 – 2;0	973 + 1193
Tuomas (FINN)	1;7 – 2;3	3170 + 4484
Tuulikki (FINN)	1;7 – 3;0	2589 + 4329
Armando (YUC)	2;0 – 3;0	445 + 682

5.3. Independent variables of the input

A. Morphological richness

A1. Paradigmatic morphological richness

In terms of the paradigmatic morphological richness of the input data, the three languages differ with respect to the number of non-homophonous form-type categories. In nouns, Yucatec has the least number of categories, only 5. Turkish and Finnish are about equally rich with 15 and 14 noun categories, respectively. In verbs, there are more inflectional categories than in nouns in each of the three languages. Turkish has the highest number of categories, 37. Finnish has the least, 19, and Yucatec is in between with 30 categories. The categories delineated in each language are presented in Table 3 below.

Table 3. Form-type categories in Turkish, Finnish and Yucatec

	Nouns	Verbs
TURK	*Sg(=0), Pl; Possessive (1Sg, 2Sg, 3Sg, 1Pl, 2Pl, 3Pl); Case: Nominative (=0), Accusative, Dative, Genitive, Locative, Ablative, Commutative-Instrumental; Pronominalizer (ki); Compound marker*	*Imperative, Aorist, Present/Prog, Future, Past-direct, Past-evidential, Optative, Abilitative, Conditional, Infinitive; agreement: 1Sg, 2Sg, 1Pl, 2Pl, 3Pl, Opt&1Pl, Imp&2Sg; neg, neg&aorist; question; passive, reflexive; copula, gerunds (3 forms); relativizers (3 forms); nominalizers (2 forms); cases (2 forms) Possessive: 1Sg, 2Sg, 3Sg, 1Pl*

Table 3 (continued). Form-type categories in Turkish, Finnish and Yucatec

	Nouns	Verbs
FINN	*Sg, Pl and the cases Nominative, Partitive, Accusative, Genitive, Translative, Essive, Adessive, Allative, Ablative, Inessive, Illative, Elative*	*Imp, Present, Past /Preterite, neg, Conditional, 1Sg, 2Sg, 3Sg, 2Pl, 3Pl, passive, Inf1, Inf2Inessive, Inf3Illative, Inf3Inessive, Inf3Adessive, PresParticAct, PastParticAct, PastParticPass*
YUC	*Sg, Pl, Possessive (associative/inalienable/ownership)*	*Transitive verbs: Incompletive, Completive, Subjunctive/Imperative, Perfect, Passive incompletive (2 forms), Completive (2 forms), Participle;* *Intransitive verbs: Incompletive (3 forms), Completive (3 forms); Subjunctive (4 forms), Imperative (2 forms); Future participle; reflexive 2Sg; Absolutive:1Sg, 2Sg, 3Sg (2 forms), 1Pl , 2Pl, 3Pl.*

The paradigmatic morphological richness was examined also in terms of total input mean size of paradigm (MSP) (Variable A1, see Xanthos & Laaha, this volume). The values of total input MSP for each of the three languages are given in Table 4.

Table 4. Total input MSP (mean size of paradigm) in Turkish, Finnish and Yucatec

	Nouns	Verbs
TURK	1.91 *(Input Den)*	3.93 *(Input Den)*
FINN	1.45 *(Input Tuo 1.50, Input Tuu 1.40)*	2.09 *(Input Tuo 2.06, Input Tuu 2.11)*
YUC	1.37 *(Input Arm)*	1.96 *(Input Arm)*

As can be observed, total input MSP is highest in Turkish, both for nouns and for verbs, and lowest in Yucatec. It is larger for verbs than for nouns in all the three languages.

These findings are presented graphically in the following figures (Figures 1a, 1b).

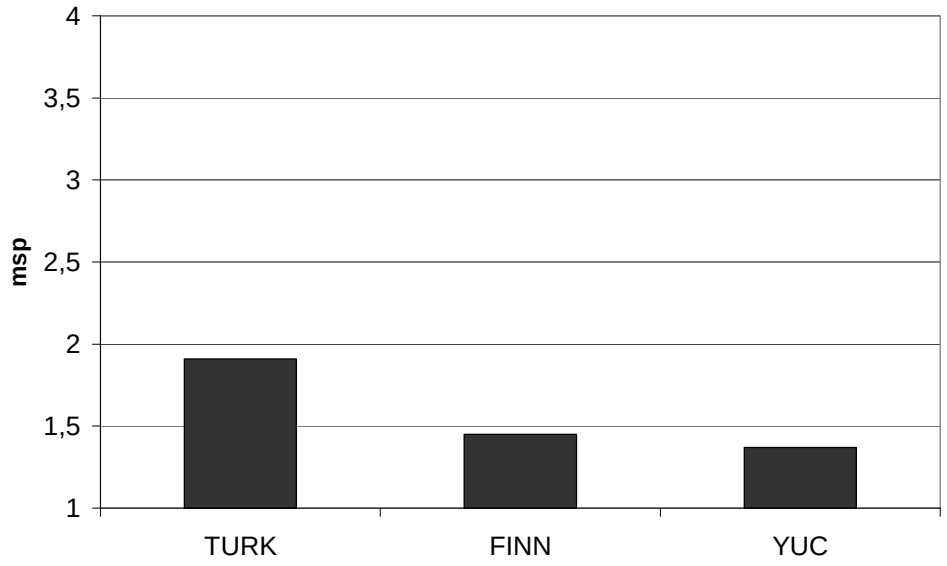

Figure 1a. Total input MSP (mean size of paradigm) in nouns for Turkish, Finnish and Yucatec

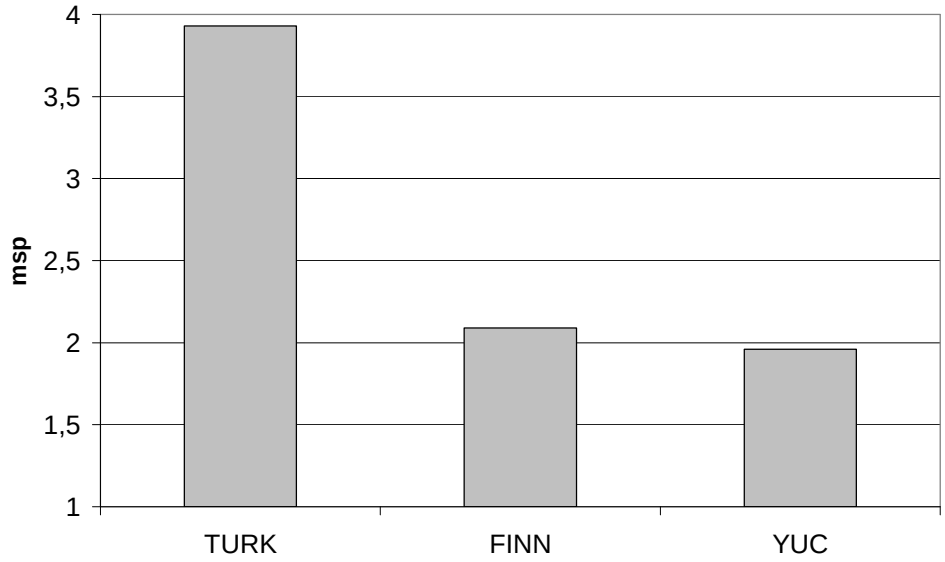

Figure 1b. Total input MSP (mean size of paradigm) in verbs for Turkish, Finnish and Yucatec

A2. Syntagmatic morphological richness

We now turn to syntagmatic morphological richness of the input data and first present the distribution of suffixes in nouns and verbs.

For nouns (Figure 2a), the highest number of suffixless basic forms is in Yucatec with about 85%, whereas in Turkish and in Finnish only about 40% of the noun forms are suffixless.

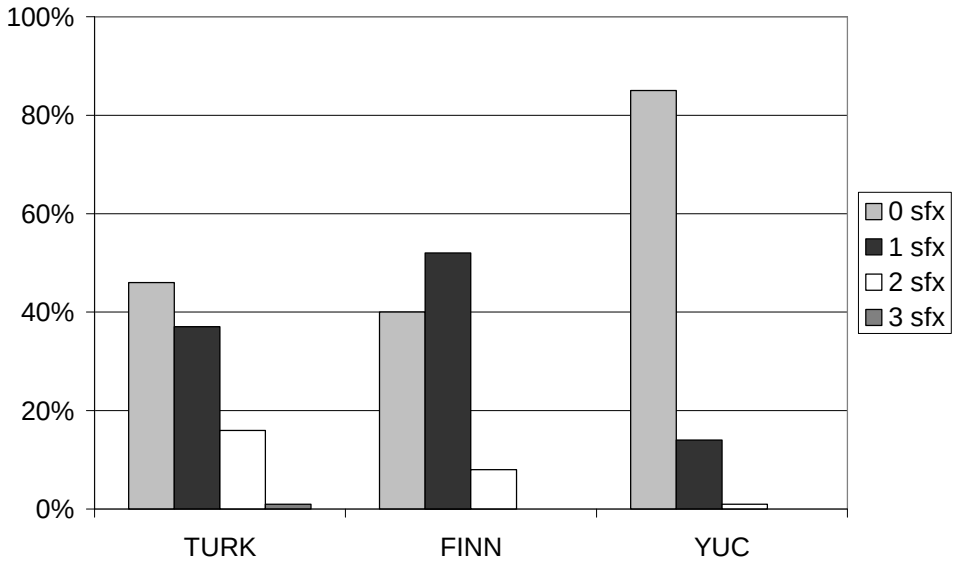

Figure 2a. Distribution of suffixes in nouns for Turkish, Finnish and Yucatec

In Turkish and in Yucatec Maya, the suffixless forms can be either the subject in the nominative, or a non-referential/non-specific object. In Finnish, suffixless forms are singular nominatives. The different form types in each language are illustrated by the following examples with the corresponding number of suffixes given in parentheses (Examples 1-3):

(1) Turkish

göz	eye	'eye' (0)
göz-üm	eye-POSS&1SG	'my eye' (1)
göz-ler-im	eye-PL-POSS&1SG	'my eyes' (2)
göz-ler-im-i	eye-PL-POSS&1SG-ACC	'my eyes' (Direct Object) (3)

(2) Finnish

jalka	leg	'leg' (0)
jala-n	leg-GEN	'of the leg' (1)
jalko-j-a	leg-PL-PART	'the legs' (2)

(3) Yucatec

xanab	shoe	'shoe' (0)
xanab-o'ob	shoe-PL	'shoes' (1)
(in=)nah-il-o'on	house-POSS-ABS1PL	'our houses' (2)

The situation is different in verbs (Figure 2b). In all three languages there are less than 20% verb forms without any suffix. The number of suffixes on verbs is lowest in Finnish and highest in Yucatec: about 90 % of the Finnish verb tokens are marked with only one suffix, whereas 60% of the Yucatec verb tokens have two suffixes. In Turkish, on the other hand, 30% of the verbs have one suffix, 40% two suffixes and 10% three suffixes.

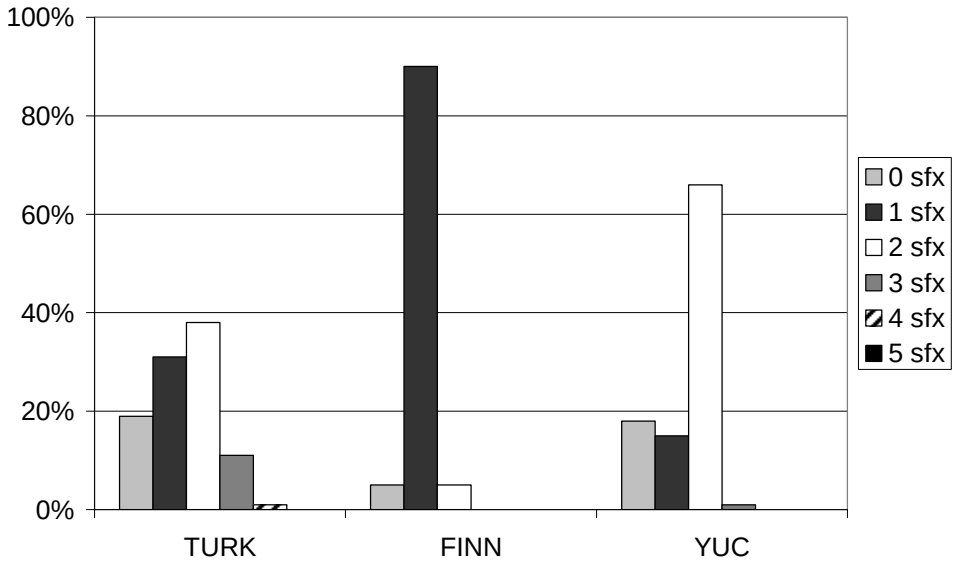

Figure 2b. Distribution of suffixes in verbs for Turkish, Finnish and Yucatec

Examples (4)-(6) below illustrate the different form types (number of suffixes given in parantheses):

(4) Turkish

yap	do	'do' (0)
yap-tı	do-PAST	'he did' (1)
yap-tı-m	do-PAST-1SG	'I did' (2)
yap-ıl-ır mı	do-PASS-AOR QUE	'can it be done' (3)
yap-a-ma-dı-m	do-ABIL-NEG-PAST-1SG	'I could not do' (4)
bul-a-mı-yacak-tı-m	find-ABIL-NEG-FUT-PAST-1SG	'I was not going to be able to find' (5)

(5) Finnish

tule	come	'come!' (0)
tule-t	come-2SG	'come' (1)
tul-isi-t	come-COND-2SG	'you would come' (2)

(6) Yucatec

(k-in=)t'aan	(INC=ERG1Sg) talk	'(I) talk' (0)
t'an-eh	talk-IMP	'talk!' (1)
(in=)k'ahóolt-ik-o'ob	(ERG1=)recognize(tv)-INC-ABS3Pl	'(I) recognize them' (2)
ganaart-ik-o'on-ech	win(tv)-INC-ABS1Pl-ABS2Sg	'you win over us' (3)

The following figures (see Table 5) express morphological richness of the input in terms of average number of suffixes (Variable A2). It is observed that Turkish presents the richest input both for nouns and verbs, Finnish and Yucatec are comparable in terms of the number of suffixes on verbs whereas Turkish and Finnish are more similar in case of nouns.

Table 5. Average number of suffixes in Turkish, Finnish and Yucatec

	Nouns	Verbs
TURK	0.70 *(Input Den)*	1.43 *(Input Den)*
FINN	0.60 *(Input Tuo 0.56, Input Tuu 0.63)*	1.01 *(Input Tuo 0.99, Input Tuu 1.02)*
YUC	0.16 *(Input Arm)*	1.12 *(Input Arm)*

B. Transparency

The three languages differ in terms of the morphophonological processes that affect word and form transparency (Variables B1 and B2). Examples (7), (8), and (9) illustrate transparent and opaque noun forms for each language:

(7) Turkish: morphophonological alternation due to suffixing
transparent: *kutu* 'box' : *kutu-n* 'box-POSS2SG'
opaque: *burun* 'nose' : *burn-un* 'nose-POSS2SG'

(8) Finnish:
transparent: *juusto* 'cheese' : *juusto-a* 'cheese-PART'
opaque: *hevonen* 'horse' : *hevos-ta* 'horse-PART'

(9) Yucatec: morphophonological contraction of the second syllable
transparent: *peek'* 'dog' : *peek'-o'ob* 'dog-PL'
opaque: *peek'* 'dog' : *peek'-oo* 'dog-PL-DEIC'

In Turkish and Yucatec, pure agglutination is common: in these languages, transparent combination of stem and suffix is more the rule rather than the exception, and noun forms are very little affected by morphophonological processes that reduce transparency. Non-transparent noun forms are at about 10% in the Turkish input and at about 5% in the Yucatec input. In Finnish inflection, especially in connection to plural formation, there are many morphophonological alternations which result in low word transparency: only 15% of the nouns maintain their phonological form in plural when inflectionally modified. But plural is

only seldom used in the input, and that is why form transparency in Finnish is about the same as in Turkish and Yucatec as shown in the following figure (Figure 3a).

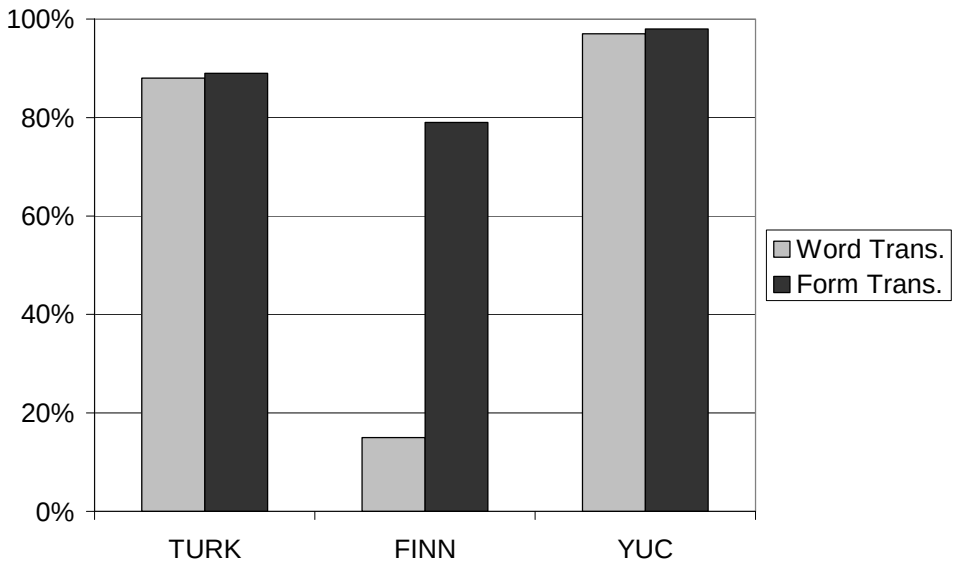

Figure 3a. Word and form transparency in nouns for Turkish, Finnish and Yucatec

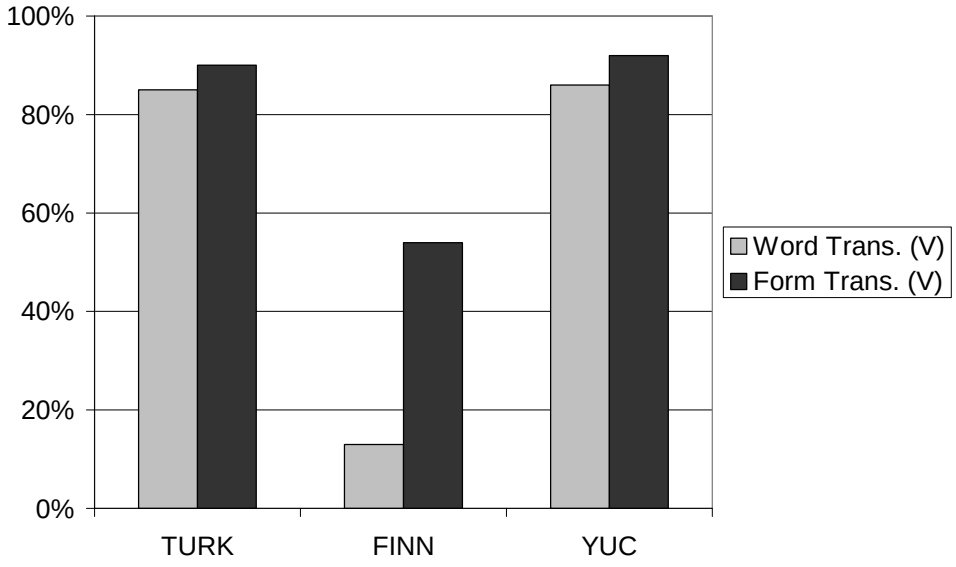

Figure 3b. Word and form transparency in verbs for Turkish, Finnish and Yucatec

Examples of transparent and opaque verb forms from each language are presented in examples (10)-(12) below:

(10) Turkish: morphophonological alternation due to suffixing
transparent: *kır* 'break' : *kır-ıl* 'break-PASS'
opaque: *devir* 'knock over' : *devr-il* 'knock-over-PASS'

(11) Finnish: certain suffixes cause morphophonological alternations with certain stems
transparent: *sano-a* 'say-INF1' : *sano-taan* 'say-PASS'
opaque: *alka-a* 'begin-INF1' : *ale-taan* 'begin-PASS'

(12) Yucatec: morphophonological contraction
transparent: *t'an* 'talk' : *t'an-eh* 'talk-IMP'
opaque: *haant* 'eat' : *haant-koo (=haant-ik-o'ob)* 'eat-INC-3PL'

The difference between the three languages is even more pronounced in verbs (see Figure 3b): in Turkish and Yucatec Maya the vast majority of verb lexemes and also verb forms are transparent, whereas in Finnish there are stem alternations in most verbs, thus word transparency is about as low as in nouns. But these stem alternations do not occur very frequently in the input, and form transparency is again higher than word transparency, yet not as high as in nouns.

C. Uniformity

Turkish has no non-uniform noun categories. The non-uniform categories in Finnish are the plural, partitive, and the illative which all occur relatively frequently; 24% of the noun forms are non-uniform. The non-uniform categories of Yucatec are the possessive and inalienable.

In verbs, Turkish displays some level of non-uniformity, particularly due to the behaviour of the aorist inflection. Non-uniformity is more pervasive both in Finnish and Yucatec. The verbal categories that show non-uniformity in Finnish are 3SG, passive, past participles, past, 1st and 2nd infinitives. The Yucatec non-uniform categories are four: Incompletive, Completive TV: SUBJ, IMP, NEG. In sum, in Yucatec there is a high proportion of non-uniform categories for verbs, in Finnish the verbal form categories are predominantly non-uniform.

The percentage of uniform categories (Variable C) in the three languages is given in Table 6.

Table 6. Uniform categories in Turkish, Finnish and Yucatec

	Nouns	Verbs
TURK	100.0% *(Input Den)*	93.0% *(Input Den)*
FINN	76.0% *(InputTuo 79%, Input Tuu 73%)*	20.0% *(Input Tuo 15%, Input Tuu 25%)*
YUC	96.0% *(Input Arm)*	45.0% *(Input Arm)*

D. Salience

As far as phonological-segmental salience (Variable D1) is concerned, there is no vowel reduction in suffixes in Turkish, but in Finnish and in Yucatec some vowel reduction occurs both in nouns and in verbs (see Table 7).

Table 7. Phonological-segmental salience of inflectional suffixes in Turkish, Finnish and Yucatec

	Nouns	Verbs
TURK	100.0% *(Input Den)*	100.0% *(Input Den)*
FINN	90.0% *(InputTuo 90%, Input Tuu 90%)*	90.0% *(Input Tuo 90%, Input Tuu 90%)*
YUC	90.0% *(Input Arm)*	70.0% *(Input Arm)*

As regards prosodic salience of inflectional suffixes (Variable D2), Finnish is very different from the other two languages. In Turkish stress is mostly word final; deviations are mostly due to the clitic nature of some of the suffixes which cause the primary stress in the word to shift to the syllable preceding it. Turkish displays such non-canonical patterns with clitics in nouns (8%) and with clitics and few other suffixes in verbs (34%). In Yucatec stress is word-final, Yucatec shows 100% canonical stress in both categories. So the suffixes in both languages are prosodically salient. However in Finnish stress is word-initial and the suffixes are prosodically non-salient (see Table 8).

Table 8. Prosodic salience of inflectional suffixes in Turkish, Finnish and Yucatec

	Nouns	Verbs
TURK	92.0% *(InputDen)*	66.0% *(Input Den)*
FINN	0.0% *(InputTuo 0%, Input Tuu 0%)*	0.0% *(Input Tuo 0%, Input Tuu 0%)*
YUC	100% *(Input Arm)*	100% *(Input Arm)*

5.4. Development of mean size of paradigm in child speech

The characteristics that have been summarized above for the three languages constitute our independent variables. The dependent variable is the speed of development. Below we present the curves representing the development of MSP for nouns (Figure 4a), and for verbs (Figure 4b).

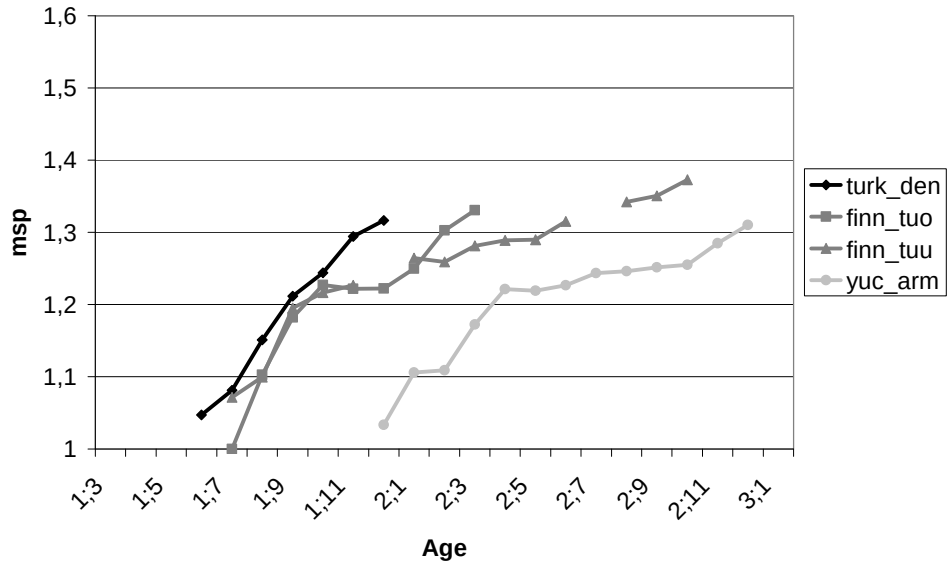

Figure 4a. Development of mean size of paradigm in nouns for Turkish, Finnish and Yucatec

It is observed that the curves for the development of MSP in nouns for the three languages have almost the same shape. Each curve rises steeply at first indicating a fast development at the beginning, until each child reaches MSP 1.2. After this point, development is slower for Finnish and particularly for Yucatec as compared to Turkish. In case of noun inflection, all three independent variables related to morphological richness predict that development would be faster in Turkish, next in Finnish and slowest in Yucatec – and this is borne out.

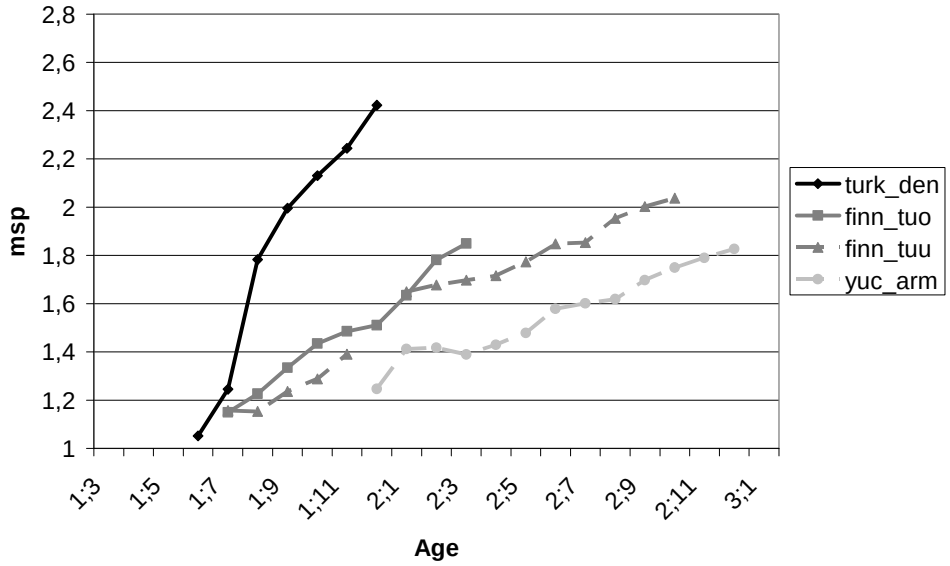

Figure 4b. Development of mean size of paradigm in verbs for Turkish, Finnish and Yucatec

For verbs, all independent variables related to morphological richness make Yucatec and Finnish more similar to one another as compared to Turkish. A similar grouping of the languages is observed with respect to the dependent variable: the developmental curve for mean size of paradigm for verbs is far steeper for Turkish than for Finnish and Yucatec, showing faster development.

To conclude: the development of nouns is more uniform across the three languages whereas the child acquiring Turkish goes ahead for verbs. The noun paradigms have more limited size than the verb paradigms, and this may result in a more similar pattern of acquisition in nouns than in verbs.

6. General results

Aris Xanthos

6.1. Introduction

In this chapter, we will discuss the results of this research at a more general level; in particular, we will examine our main hypothesis, i.e. that the speed of development of morphological richness in child speech is positively correlated with the degree of morphological richness of the input (see Dressler, this volume). We applied two standard statistical tools to the data, namely analysis of variance (ANOVA) and correlation analysis. The former was used to ensure that observed differences in speed of development between languages and language groups were significant; the latter specifically addresses the hypothesized correlation.

Each method and its results will be discussed in a separate section. Since the data that were used are described in detail in the chapter "Method" (Xanthos & Laaha, this volume), we will mainly highlight the aspects that distinguish the set-up of the ANOVA from that of the correlation analysis. Also, a more thorough discussion of the presented results will be deferred to the concluding chapter (Dressler et al., this volume).

6.2. Analysis of variance

6.2.1. Method

In order to assess the significance of differences in speed of development of MSP (see Xanthos & Laaha, this volume for details on the calculation of this variable) between languages and language groups, we applied an ANOVA to our data. Although this does not directly document our main hypothesis, it sheds some light on the cross-linguistic and typological relevance of the dependent variable, speed of development.

The individuals entered into this analysis were monthly samples, characterized by their score for speed of development of MSP (nouns and verbs).[1] Figures 1a and 1b show the mean speed (and 95% confidence intervals) obtained for each individual language, for nouns and verbs respectively. Figures 2a and 2b show the corresponding values obtained by clustering languages into weakly inflecting, strongly inflecting, and agglutinating language groups. We examined the dependence between speed of development and each factor (language and language group) using Welch's variant of ANOVA (Welch, 1951), which does not rely on the assumption of the homogeneity of variances. In particular, we performed 4 separate one-way ANOVAs: language (9) and language group (3) factors were examined separately, and this was done for nouns and verbs separately.[2]

[1] Notice that the analysis included the six Dutch corpora for which no input data were available (see Laaha et al., this volume, Table 1').

[2] The language factor being nested within the language group factor, it is in principle possible to perform a *nested* ANOVA on these data; we are currently exploring this issue.

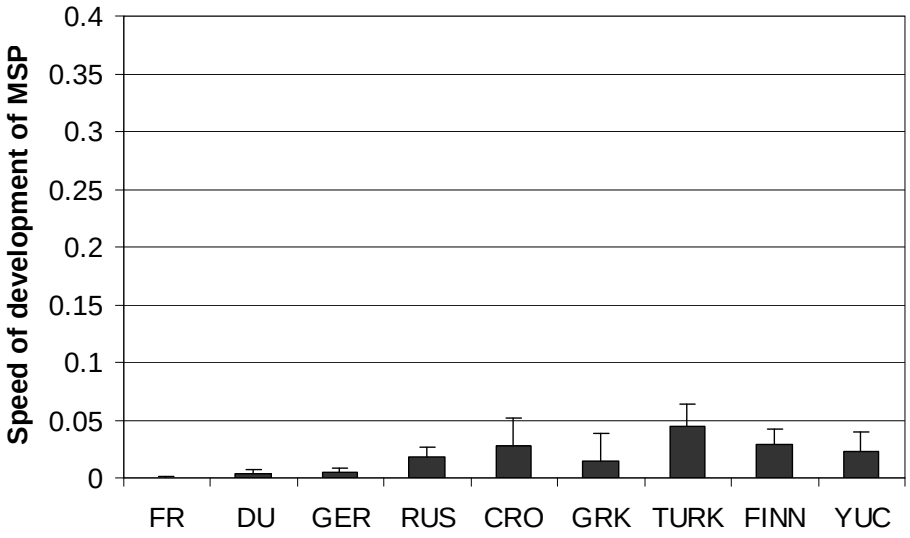

Figure 1a. Mean speed of development of MSP for each language: nouns

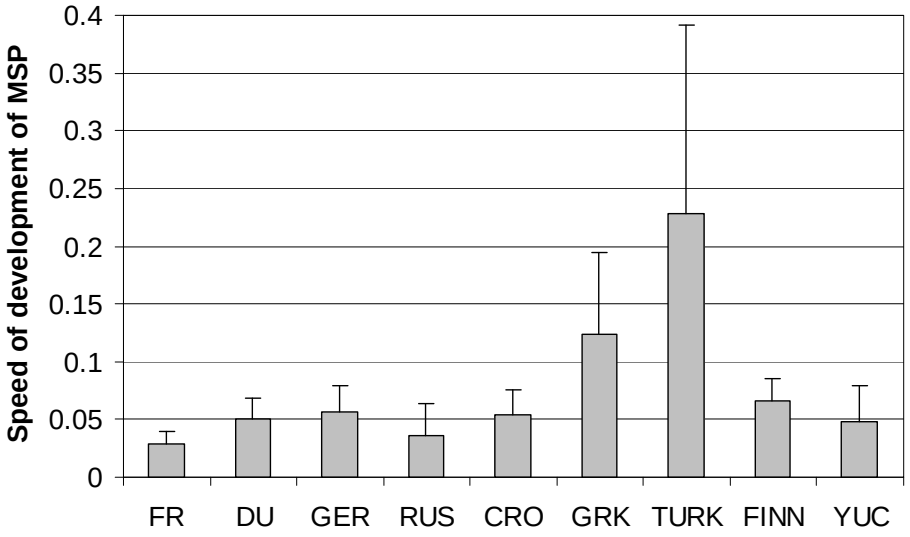

Figure 1b. Mean speed of development of MSP for each language: verbs

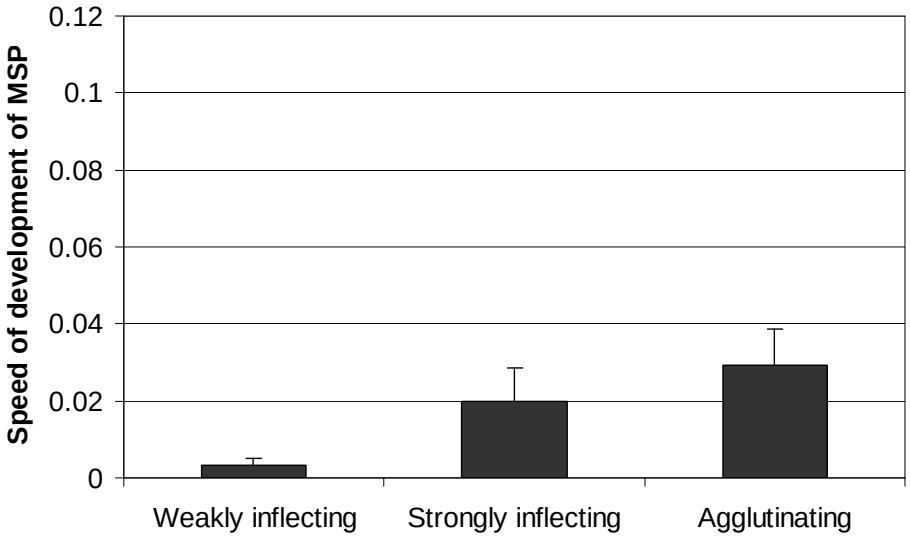

Figure 2a. Mean speed of development of MSP for each language group: nouns

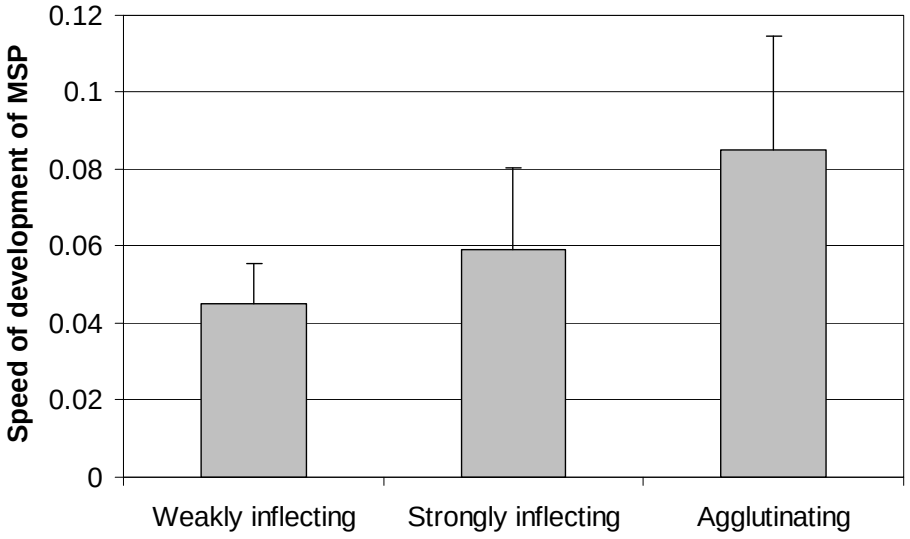

Figure 2b. Mean speed of development of MSP for each language group: verbs

6.2.2. Results

The Welch ANOVA results are presented in Table 1. As can be seen, the pattern of results differs between nouns and verbs. For nouns, the dependence of speed of development on both factors (language and language group) is highly significant ($p < .001$). For verbs, the dependence is significant, though generally weaker: it is still high for the language factor ($p = .002$), but less so for the language group factor ($p = .032$).

Table 1. Welch ANOVA results

Factor	Subsystem	*F*	Degrees of freedom		Sig. (*P*)
			1	2	
Language	Nouns	12.14	8	41.96	< .001
	Verbs	3.64	8	45.72	.002
Language group	Nouns	22.49	2	60.82	< .001
	Verbs	3.59	2	75.08	.032

On the whole, we may say that despite large confidence intervals, speed of development is significantly dependent on both factors. The degree of this dependence varies in a way that corresponds to the intuitive examination of Figures 1a to 2b. Nouns display a more contrasted profile than verbs, and a better correspondence with typological expectations (in relation to our main hypothesis). Verbs, on the other hand, are characterized by the prominence of Turkish and Greek, two rather small corpora (see Xanthos & Laaha, this volume, section 2.2), which account for most of the differences at the level of language groups.

At this point, it is difficult to say whether this diverging behavior of nouns and verbs should be explained by other factors than the particular sample of languages and children that was used. Nevertheless, more differences between the two subsystems will be reported in the next section.

6.3. Correlation analysis

6.3.1. Method

The second analysis was aimed at determining how well each of the seven independent variables (A1: paradigmatic morphological richness, A2: syntagmatic morphological richness, B1: word transparency, B2: form transparency, C: uniformity, D1: phonological-segmental salience, D2: prosodic salience) accounted for the observed differences in speed of development of MSP. This time, the individuals were whole corpora (as opposed to monthly samples). 13 corpora were entered into the analysis, most of them having a score for each of the seven independent variables (see Table 2 below and Tables 2a and 2b in Xanthos & Laaha, this volume for exceptions to this) besides their score for the dependent variable. Similarly to the ANOVA setup, nouns and verbs were examined separately.

Pearson correlation coefficient (r) is by far the most widely used measure of correlation. However, it is highly sensitive to the presence of outliers (individuals located far from the rest of the data), to the point that the addition of a single outlier to an otherwise random sample can yield a very high correlation coefficient. In our case, the speed of development of MSP measured for the corpus of Turkish verbs was much higher than the others (see Table 2b in Xanthos & Laaha, this volume), so we could not rely on Pearson correlation coefficient.

Instead of that, we used Spearman correlation coefficient (r_S), which is defined as the (Pearson) correlation coefficient between the *ranks* corresponding to the original scores. The

related statistical test is less powerful (i.e. more liable not to detect significant correlations), but it is recommended in the presence of one or more outliers; we applied the two-tailed version of the test, since there was no *a priori* reason to exclude negative correlations.

6.3.2. Results

The correlation coefficients between each independent variable and speed of development are plotted on Figures 3a (nouns) and 3b (verbs) below, where hatched bars represent values that are *not* significant at the 0.01 level (see also Table 2 below).

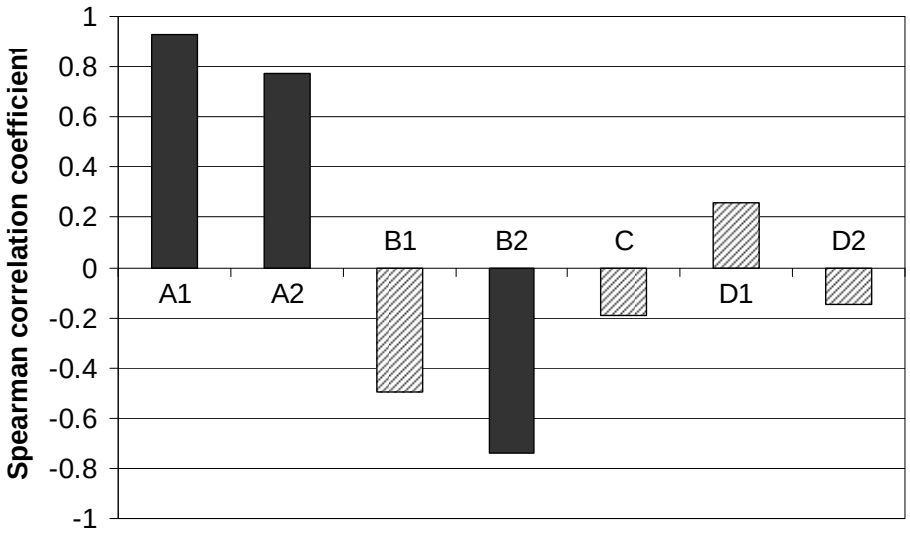

Figure 3a. Spearman correlation between each independent variable and speed of development of MSP: nouns

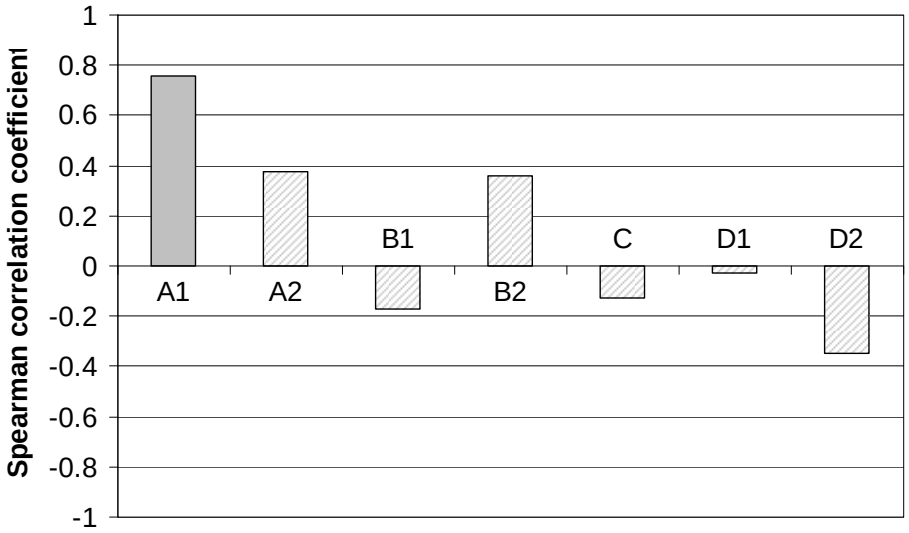

Figure 3b. Spearman correlation between each independent variable and speed of development of MSP: verbs

Table 2. Correlation analysis results

Independent variable		Nouns			Verbs		
		Corr. (r_S)	Sig. (P)	Nb. scores (N)	Corr. (r_S)	Sig. (P)	Nb. scores (N)
A1	Paradig. morph. richness	.930	< .001	13	.757	.003	13
A2	Syntag. morph. richness	.774	.002	13	.375	.206	13
B1	Word transparency	-.497	.084	13	-.169	.580	13
B2	Form transparency	-.739	.004	13	.361	.226	13
C	Uniformity	-.192	.595	10	-.125	.699	12
D1	Phonol.-segm. salience	.256	.399	13	-.025	.936	13
D2	Prosodic salience	-.149	.643	12	-.349	.243	13

The most striking result is that variable A1 (paradigmatic morphological richness) has the most significant correlation with speed of development of MSP, both for nouns (r_S = .93, p < .001) and for verbs (r_S = .757, p = .003). This is a strong confirmation of the hypothesized relationship between speed of development and the degree of morphological richness in the input.

In the case of verbs, no other independent variable is significantly correlated with speed. For nouns, however, two other variables have a significant, correlation, though lesser than A1: syntagmatic morphological richness A2 (r_S = .774, p = .002), and form transparency B2 (r_S = -.739, p = .004). In our sample, A2 and B2 are even more strongly correlated with A1 (r_S = .802, p = .001 and r_S = -.83, p < .001 respectively). Therefore, it is possible that the correlation of one or the other with speed is a by-product of their relationship with A1.[3]

In order to investigate such effects, we calculated *partial* correlation coefficients, which account for the relationship between two variables after factoring out the possible relationship between each of them and a third one (see e.g. Bavaud, 1998). We first considered the case of variable A2. The partial correlation between speed and A1 (controlling for A2) is still very significant ($r_{S\ \text{speed,A1|A2}}$ = .818, p = .001); thus it is not likely that their direct correlation is influenced by A2. On the contrary, the partial correlation between speed and A2 (controlling for A1) falls drastically with respect to their direct correlation ($r_{S\ \text{speed,A2|A1}}$ = .128, p = .691), so that the latter may well be a spurious effect of the influence of A1.

We conducted the same verification with variable B2. The results are very similar, i.e. the correlation between speed and A1 decreases much less when controlling for B2 ($r_{S\ \text{speed,A1|B2}}$ = .842, p < .001) than the correlation between speed and B2 when controlling for A1 ($r_{S\ \text{speed,B2|A1}}$ = .160, p = .618). Thus, it is likely that, here as well, A1 is the underlying factor.

As was noted in the previous section, it is not clear whether the differences found between the nominal and verbal subsystems would be confirmed using another or a larger sample of children and languages. From a purely statistical point of view, given the less contrasted values measured for speed of development for verbs, it is not a surprise that the corresponding correlations with independent variables are generally weaker. Nonetheless, with regard to the data at hand, we may state that only variable A1 (paradigmatic morphological richness) is consistently correlated with speed of development.

[3] In the case of syntagmatic morphological richness, this is what the parameter of morphotactic transparency (see e.g. Kilani-Schoch & Dressler, 2005) predicts: all other things being equal, affixation is a more transparent pattern of inflection than stem modification (or suppletion); to that extent, it is *natural* that an increase in paradigmatic morphological richness should be matched by an increase in syntagmatic morphological richness.

6.4. Summary

In this chapter, we have surveyed the results of the general analyses that were applied to our data. Using Welch ANOVA, we found that the speed of development of MSP varied significantly as a function of the language and language group factors, though this effect is stronger for nouns than for verbs, and for the first factor than for the second one. A second analysis focused on the Spearman correlation between speed of development and each independent variable (at the level of corpora). It revealed that speed was highly correlated with variable A1 (paradigmatic morphological richness), both for nouns and for verbs, as predicted by our main hypothesis. For nouns, other significant correlations were found for variables A2 (syntagmatic morphological richness) and B2 (form transparency); however, the examination of partial correlations suggests that these may be spurious effects of the relationship of A2 and B2 with A1.

7. Discussion and Conclusion

Wolfgang U. Dressler, Ursula Stephany, Ayhan Aksu-Koç, and Steven Gillis

Our main hypothesis, that "the richer noun or verb morphology is in the input, the more stimulated the child will be to develop noun or verb morphology rapidly" (Dressler, this volume: 8), has stood the test. This hypothesis has been embedded in a typological framework which considers noun and verb inflection as entities of a two-fold comparison: (1) a comparison of nine languages according to three different language types, (2) a comparison of the morphological subsystems of noun inflection and verb inflection. This is the first time that models of ordering typology and quantitative typology (in the sense of Hempel & Oppenheim, 1936; see Dressler, this volume) have been applied to a cross-linguistic study of first language acquisition.

In the preceding chapter (Xanthos, this volume) *classificatory typology* has been used to relate the three language types of weakly inflecting, strongly inflecting and agglutinating languages to speed of development, i.e. by clustering the scores obtained for speed of development of each individual language into the three language groups. The expected significant differences between language groups have been found with respect to speed of development, i.e. greatest speed (in terms of speed of development of MSP) for the group of agglutinating languages, lower speed for the strongly inflecting languages, and lowest speed for the weakly inflecting ones (see also Figures 2a, 2b and Table 1 in Xanthos, this volume). It must be kept in mind, however, that these results only represent a first approximation, due to the oversimplified concept of typological classes of languages (see Dressler, this volume), which informs us only about cross-class differences, but not about cross-language differences.

In order to compare all nine languages of our sample individually, they were ordered following the model of *ordering typology*, namely, according to the degree to which they display characteristics of the ideal agglutinating, inflecting and isolating types. The characteristics used for ordering the languages are morphological richness and related typological criteria, as these are reflected in the input.

In the domain of noun inflection, the (mainly) agglutinating languages were placed between the agglutinating and inflecting prototypes (1), and the (mainly) inflecting languages were ordered between the inflecting and the isolating prototype (2), as follows (see also Table 1 in Dressler, this volume):

Ordering typology: Noun inflection
(1) Agglutinating – inflecting: Turkish – Finnish – Yucatec
(2) Inflecting – isolating: Croatian – Russian – Greek – German – Dutch – French

This order is identical for paradigmatic and syntagmatic richness.

As far as verb inflection is concerned, the order is somewhat different (see also Table 1 in Dressler, this volume). The ordering for paradigmatic and syntagmatic richness is nearly identical.

Ordering typology: Verb inflection
(3) Agglutinating – inflecting: Turkish – Yucatec – Finnish
(4) Inflecting – isolating: Croatian – Greek – Russian – …

Within weakly inflecting languages the order differs for paradigmatic and syntagmatic richness:
(4a) Paradigmatic axis: French – German – Dutch
(4b) Syntagmatic axis: German – Dutch – French

How do the differences in morphological richness found in children's input in the nine languages relate to speed of development?

As far as noun inflection is concerned, the order of speed of development on the whole equals the typological order of languages, i.e. Turkish – Finnish – Yucatec for the agglutinating languages, and Croatian – Russian – Greek – German – Dutch, and again, with a sizable interval, French for the inflecting ones (see also Figure 1a in Xanthos, this volume).

For verb inflection, however, the order of speed of development differs from the typological ordering of languages (see also Figure 1b in Xanthos, this volume):
1) agglutinating languages: Turkish – Finnish – Yucatec, as for nouns, but with an inversion of Yucatec and Finnish, contrary to the order predicted by the typological ordering of verb inflections.
2) inflecting languages: Greek – German – Croatian – Dutch – Russian – French.

Despite these ordering differences, however, the results show the same pattern as above: speed of development in the group of (mainly) agglutinating languages is faster than in that of (mainly) inflecting languages, and over time, there is a better match to the number of target categories in nouns, than to the verbs.

How are these results to be interpreted? One possibility is that there are fewer morphological categories to be acquired in the noun than in the verb. Therefore the limited number of categories acquired over time shows a better match to the number of target categories in nouns.

Another possibility is that verb morphology is acquired differently from noun morphology (Behrens, 1999; see below on noun bias). Unlike nouns, which are typically referring, verbs are typically predicating (Gentner, 1978, 1982) and are therefore much more involved in syntax than nouns (despite the contextual character of case). For this reason syntactic factors play a bigger role in the acquisition of verbs than of nouns. Hence the degree of morphological richness will tend to be more relevant for the acquisition of nouns than for that of verbs. This might explain the closer correspondence between morphological richness and speed of development in nouns than in verbs.

Comparing typological ordering of languages with rank order of speed of development poses two major problems: First, the amount of agreement between typological ordering and rank order of speed of development cannot be properly measured, i.e. it can only be evaluated whether the typological order of languages and the order of speeds of development agree, but not to what extent they do. The second problem is that the ordering relation between agglutinating and strongly inflecting languages is a priori unclear, because they are ordered between different poles of ideal language types (agglutinating – inflecting vs. inflecting – isolating). Relative speed of development cannot be predicted across these language types. The question is whether cumulative morphology, typical of inflecting languages, where, for example, a unique suffix expresses both number and case of nouns, will stimulate children to develop inflection more or less strongly than the concatenative technique of agglutinating languages, where these grammatical categories are expressed by two subsequent suffixes. The evidence from omissions of grammatical morphemes quoted by Peters (1997: 182) seems to favour the early acquisition of agglutinating inflection as opposed to cumulative morphology (we will come back to this question below).

These two problems of ordering typology disappear when the model of *quantitative typology* is applied to our data. As can be seen in Figures 3a, 3b and Table 2 of the preceding chapter (Xanthos, this volume), the quantitative measuring of Variable A1 paradigmatic

morphological richness (total input MSP) and Variable A2 syntagmatic morphological richness (average number of suffixes) is unequivocal: the richer inflection is in the input, the more stimulated children are to develop inflection in this domain, and the more rapid is development. On the one hand, they appear to be sensitive to the number of distinct, non-homophonous inflected forms per lemma, i.e. to the paradigmatic axis of morphological richness as measured by mean size of paradigm (MSP). On the other hand, they appear to be sensitive to the number of sequential suffixes, i.e. to the syntagmatic axis of morphology, or on-line complexity, so to say. These findings have important theoretical consequences, since both the hypothesis that morphological richness has no effect on acquisition, and the counter-hypothesis that morphological richness has a negative effect on acquisition are not supported.

Similar to what has been found in a descriptive way in ordering typology, in quantitative typology as well, the correlations between morphological richness and speed of development are less significant for verb than for noun inflection: paradigmatic morphological richness (Variable A1) is predictive for both nouns and verbs, but syntagmatic morphological richness (Variable A2) is predictive only for nouns. One reason may again be the lower complexity of nominal as compared to verbal inflection, as stated above. An additional and related reason may be that the correspondence between the independent and dependent variables of paradigmatic richness, i.e. between mean size of paradigm in the input and speed of development of mean size of paradigm in the output, is more direct than that between syntagmatic independent variables (mean number of suffixes) and paradigmatic dependent variables (mean size of paradigm).

From a typological perspective, inflecting languages differ from agglutinating ones on the syntagmatic axis of nouns more so than on that of verbs. Prototypically, the inflecting-fusional type expresses the grammatical categories of declension by a single cumulative morpheme (a suffix in Indo-European languages) while, in conjugation, often several and non-cumulative inflectional morphemes are concatenated, although to a lesser degree than in the agglutinating type. There is thus in our language sample less typological variety to be found on the syntagmatic axis in verbs than in nouns. The substitution of synthetic by analytic morphology, such as synthetic preterits or futures by compound pasts or periphrastic futures, is typical for weakly inflecting languages as opposed to strongly inflecting ones.

Not surprisingly, the two corresponding units of measurement of mean size of paradigm in the input and children's speed of development of contrasting inflectional forms of nouns and verbs were found to correlate significantly. Thus, the bigger the mean size of paradigm is in the input, the greater the speed of development of oppositions within paradigms in children's output. MSP, which represents type frequency of morphological distinctions, should therefore be a better predictor of speed of development than the overall inventory of form types occurring in the input.

As far as MSP is concerned, the question is whether zero morphemes contribute to paradigmatic morphological richness. Clearly, suffixless base forms such as English singular nouns and base forms of verbs, e.g. *go*, must be counted as belonging to the paradigm. After all, children must acquire the opposition between suffixless base forms and different suffixed forms of the same paradigm. Since suffixless forms must be integrated into the paradigm of suffixed forms they have been counted as contributing to MSP (see Xanthos & Laaha, this volume). The difference between suffixless and suffixed forms has been covered by the determination of morphological richness on the syntagmatic axis.

The different roles played by suffixless forms on the syntagmatic and the paradigmatic axis make it possible to explain why, e.g., French verb inflection, which is morphologically richer on the paradigmatic axis than that of Dutch or German, stimulates children less to speedy development of MSP than in Dutch or German. The reason is that suffixless base forms of verbs have a much higher type and token frequency in French verb inflection than in the other two languages.

In regard to discussions of whether there exists a general noun bias in the early phases of acquisition both on the lexical and the syntactic level (Tomasello et al., 1997; Lieven, 1998; Bassano, 2000; Gentner & Boroditsky, 2001), the question has been raised whether morphology emerges earlier in nouns than in verbs (Tomasello et al., 1997; Brown, 1998; Bassano, 2000; Klampfer(-Laaha) & Korecky-Kröll, 2002; Pfeiler, 2002). As far as the speed of early development of inflectional paradigms is concerned, our results are unequivocal: in all languages of our sample speed of development is higher in verb inflection than in noun inflection, even when there is a time lag between the earlier construction of noun and that of verb paradigms. This may be explained by the greater communicative importance of verb inflection as compared to noun inflection within each of our nine languages, which may also be the main reason for the existence of more morphological categories in the verb than in the noun.

In contrast to our results on morphological richness, we did not find clear-cut positive effects of the degree of morphological transparency, uniformity or salience in our data. The independent variables of word and form transparency (B1, B2), uniformity (C), phonological-segmental and prosodic salience (D1, D2) correlate with the dependent variable of speed of morphological development in unpredictable ways, namely weakly, not at all or even negatively (see Figures 3a, 3b, Table 2 in Xanthos, this volume).

However, we can observe one possible facilitating effect of transparency for children's development of inflectional morphology when we compare word and form transparency: in the inputs of all nine languages, form transparency of nouns and lexical verbs is higher than (or at least as high as) word transparency (with the only exception of French verbs). This means that actual usage of morphology in children's input is more transparent than the system itself.

Our results do not mean that the relative degree of transparency, uniformity and salience of morphological patterns in the input do not have any role to play in acquisition. Rather, these parameters have been shown to be important when compared with competing morphological patterns of the same language which are less transparent, uniform or salient (Aksu-Koç & Slobin, 1985: 847; Slobin, 1985: 1216; Peters, 1997: 181; Lieven, 1997: 223; De León, 1998: 154ff; Savickiene, 2003). Thus, diminutives may be acquired earlier than their bases if the former are more transparent or more salient than the latter (see Savickiene et al., this volume). In Dutch, root infinitives emerge earlier than competing finite forms because of their higher degree of salience (Wijnen et al., 2001; Gillis, 2003: 197-199). Note also that we measured only two of the many aspects of salience (more in Köpcke, 1993). Also, the results might change if the effect of these three factors in question were combined. But in the present study each factor (starting with paradigmatic morphological richness) was correlated separately with speed of development.

The facilitating effect of transparency, uniformity and salience studied so far does not seem to extend to the domain of inflectional morphology as a whole. It is not sufficient to claim that children will acquire the inflectional morphology of nouns or verbs in a given language more rapidly because it is relatively more transparent, uniform or salient than the corresponding morphology in another language. It is rather the importance of the role played by inflectional noun vs. verb morphology in language structure and, more directly, in language use, that will induce the child to focus more on the acquisition of noun morphology vs. verb morphology vs. syntax.

Moreover, it may be promising to compare the impact of transparency, uniformity and salience in typologically very similar languages. Future comparisons within a typologically similar group such as more or less agglutinating languages, where the condition of "all-other-things-being-equal" may be assumed to be met, might be more informative than comparisons across a much wider spectrum as in our case, where the languages vary from highly agglutinating to very weakly inflecting.

It may well be that the lack of evidence for a global facilitating role of transparency, uniformity and salience in the present study is due to methodological problems: in order to facilitate quantitative typological comparison we have pulled each child's input together into one point of comparison. This overlooks the wide-spread phenomenon of parents' fine-tuning of child-directed speech to the child's course of linguistic development (see Snow, 1989, 1995; Ketrez, 2003).

The methodological preference – necessitated by concerns for equalizing the data – for the use of a single average to represent the input values for different time points may have affected our results at a more general level. In fact, we think that higher correlations would be obtained if it had been possible to map the output at different points in development onto the input at corresponding points in development. This remains an issue for future analyses.

Another serious limitation that needs to be mentioned is the restriction of the present study to morphological form, excluding consideration of the development of morphological meanings and functions. This choice was necessitated by the scope of the task at hand: comparing nine languages only on the basis of formal categories was complicated enough to justify the exclusion of meaning and function as objects of investigation.

Finally, can we generalize what we have found for the development of the inflection of nouns and verbs to other children acquiring these languages? Although our conclusions are based on the study of a relatively small number of children, individual differences in speed of development within one and the same language are very limited: in Dutch, where seven children were studied, individual differences of speed of development are quite restricted. The same holds for the two Swiss-French children, the two Finnish children and, despite different acquisition strategies, for the two Russian children as well. One of the two Austrian children is an early talker while the other one is a late talker, but their difference in speed is relatively small. All this makes us confident of the robustness of our results, at least until further results will be available from future studies.

To conclude, what is developmentally most significant is the close relationship between mean size of paradigm in the input and speed of development of form oppositions within both nominal and verbal inflection. We have found that the bigger the mean paradigm size (MSP) in the input, the greater is the speed of development of paradigmatic oppositions in children's output. Furthermore, **c**hildren do not seem to acquire the inflectional morphology of nouns or verbs of one language more rapidly because it is relatively more transparent, uniform or salient than that of another language; nor do they seem to acquire the noun vs. verb morphology of their language with different speeds for this reason. It is rather the major or minor role that inflectional morphology of nouns and verbs plays within language structure (as reflected in the input) and, more directly, in language use, that induces the child to put more emphasis on the acquisition of noun morphology vs. verb morphology vs. syntax. Such sensitivity to typological characteristics has been established in different studies (Aksu-Koç & Slobin, 1985; Bowerman, 1985; Slobin, 1985; Stephany, 1985, 1997; Bowerman et al., 1995; Caselli et al., 1999; Bowerman & Choi, 2001; Berman, 2002; and Devescovi et al., 2005). The present study has advanced our knowledge of the impact of language typology on acquisition by quantifying with a new methodology different typological parameters of the input and speed of development in the child's speech for many more languages of different types.

8. Bibliography

edited by Sabine Laaha

Aksu-Koç, A. & Ketrez, F. N. (2003). Early verbal morphology in Turkish: emergence of inflections. In D. Bittner, W. U. Dressler & M. Kilani-Schoch (eds.), *Development of verb inflection in first language acquisition*. Berlin: Mouton de Gruyter.

Aksu-Koç, A. & Slobin, D. I. (1985). The acquisition of Turkish. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition.* Vol. 1. Hillsdale: 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.

Bates, E., Dale, P. S. & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney. (eds.), *The handbook of child language.* Oxford: Blackwell.

Bavaud, F. (1998). *Modèles et données*. Paris: L'Harmattan.

Bavin, E. (1998). Factors of typology on language acquisition: some examples from Warlpiri. In A. Siwierska & J. J. Song (eds.), *Case, typology and grammar.* Amsterdam: Benjamins.

Behrens, H. (1999). Was macht Verben zu einer besonderen Kategorie im Spracherwerb? In J. Meibauer & M. Rothweiler (eds.), *Das Lexikon im Spracherwerb*. Tübingen: Francke.

Bennis, H., den Dikken, M., Jordens, P., Powers, S. & Weissenborn, J. (1995). Picking up particles. In D. MacLaughlin & S. McEwen (eds.), *Proceedings of the 19th Annual Boston University Conference on Language Development*. Somerville: Cascadilla.

Berman, R. A. (2002). Crosslinguistic study of language acquisition. In N. J. Smelser & P. B. Baltes (eds.), *International encyclopedia of the social and behavioral sciences.* Oxford: Pergamon Press.

Berman, R. A. (2004). Between emergence and mastery: the long developmental route of language acquisition. In R. A. Berman (ed.), *Language development across childhood and adolescence.* Amsterdam: Benjamins.

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

Bowerman, M. F. (1985). What shapes children's grammars? In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition.* Vol. 2. Hillsdale: Erlbaum.

Bowerman, M. & Choi, S. (2001). Shaping meanings for language: universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (eds.), *Language acquisition and conceptual development.* Cambridge: Cambridge University Press.

Bowerman, M., de León, L. & Choi, S. (1995). Verbs and particles, and spatial semantics: learning to talk about spatial actions in typologically different languages. In E. V. Clark (ed.), *The proceedings of the 27th Annual Child Language Research Forum.* Stanford: Center for the Study of Language and Information.

Brown, P. (1998). Children's first verbs in Tzeltal: evidence for an early verb category. *Linguistics* **36**, 713-753.

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.

Christofidou, A. (2003). Genos ke klisi stin Elliniki: mia fisiki prosegisi. In A. Anastasiadi-Symeonidi, A. Ralli & D. Chila-Markopoulou (eds.), *To genos*. Athens: Patakis.

Christofidou, A. & Stephany, U. (2003). Early phases in the development of Greek verb inflection. In D. Bittner, W. U. Dressler & M. Kilani-Schoch (eds.), *Development of verb inflection in first language acquisition*. Berlin: Mouton de Gruyter.

De León, L. (1998). Raices verbales tempranas en maya tzotzil: factores del input vs. restricciones cognoscitivas. *Funcion* **18**, 147-174.

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. U. (1997) (ed.). *Studies in pre- and protomorphology*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

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

Dressler W. U., Kilani-Schoch, M., Gagarina, N., Pestal, L. & Pöchtrager, M. (2006). On the typology of inflection class systems. *Folia Linguistica* **XL/1-2**, 51-74.

Dressler, W. U., Kilani-Schoch, M. & Klampfer(-Laaha), S. (2003). How does a child detect bound morphology? In R. H. Baayen & R. Schreuder (eds.), *Morphological structure in language processing.* Berlin: Mouton de Gruyter.

Dressler, W. U., Mayerthaler, W., Panagl, O. & Wurzel, W. U. (1987). *Leitmotifs in Natural Morphology.* Amsterdam: Benjamins.

Geckeler, H. (1984). Le français est-il une langue isolante? – V. Skalička et la typologie du français. In F. J. Oroz Arizcuren (ed.), *Navicula Tubingensis. Fs. A. Tovar*. Tübingen: Narr.

Gentner, D. (1978). On relational meaning: the acquisition of verb meaning. *Child Development* **49**, 988-998.

Gentner, D. (1982).Why nouns are learned before verbs: linguistic relativity versus natural partitioning. In S. A. Kuczaj (ed.), *Language development.* Vol. 2. Hillsdale: Erlbaum.

Gentner, D. & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. C. Levinson (eds.), *Language acquisition and conceptual development.* Cambridge: Cambridge University Press.

Gillis, S. (1998) (ed.). *Studies in the acquisition of number and diminutive marking* (= Antwerp Papers in Linguistics 95). Antwerp: University of Antwerp.

Gillis, S. (1999). Procedure for automatic minMOR coding. Unpublished manuscript.

Gillis, S. (2003). A case study of the early acquisition of verbs in Dutch. In D. Bittner, W. U. Dressler & M. Kilani-Schoch (eds.), *Development of verb inflection in first language acquisition: a cross-linguistic perspective.* Berlin: Mouton de Gruyter.

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

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

Jordens, P. (1990). The acquisition of verb placement in Dutch and German. *Linguistics* **28**, 1407-1448.

Jucquois, G. (1993). *Le comparatisme.* Louvain: Institut de Linguistique de Louvain.

Kempe, V. & MacWhinney, B. (1998). The acquisition of case marking by adult learners of Russian and German. *Studies in Second Language Acquisition* **20**, 543-587.

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

Ketrez, N. F. & Aksu-Koç, A. (in press). Early nominal morphology in Turkish: emergence of case and number. In U. Stephany & M. Voeikova (eds.), *Development of noun inflection in first language acquisition: a cross-linguistic perspective.* Berlin: Mouton de Gruyter.

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

Klampfer(-Laaha), S. & Korecky-Kröll, K. (2002). Nouns and verbs at the transition from pre- to protomorphology: a longitudinal case study on Austrian German. In M. Voeikova & W. U. Dressler (eds.), *Pre- and protomorphology: early phases of morphological acquisition in nouns and verbs.* München: Lincom.

Köpcke, K.-M. (1993). *Schemata bei der Pluralbildung im Deutschen.* Tübingen: Narr.

Korecky-Kröll, K. & Dressler, W. U. (in press). The acquisition of number and case in Austrian German nouns. In U. Stephany & M. Voeikova (eds.), *Development of noun inflection in first language acquisition: a cross-linguistic perspective.* Berlin: Mouton de Gruyter.

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

Lees, R. (1961). *The phonology of modern standard Turkish.* Bloomington: Indiana University.

Lehmann, C. (1983). The present state of linguistic typology. *Proceedings of the International Congress of Linguists* **13***,* 950-956.

Lieven, E. V. M. (1997). Variation in a crosslinguistic context. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition.* Vol. 5. Mahwah: Erlbaum.

Lieven, E. V. M. (1998) (ed.). *Developing a verb category: cross-linguistic perspectives* (= Linguistics 36/4). Berlin: Mouton de Gruyter.

Lucy, J. A. (1994). The role of semantic value in lexical comparison: motion and position roots in Yucatec Maya. *Linguistics* **32**, 623-656.

MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk.* Mahwah: Erlbaum.

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

Peters, A. M. (1985). Language segmentation: operating principles for the perception and analysis of language. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition.* Vol. 2. Hillsdale: Erlbaum.

Peters, A. M. (1997). Language typology, prosody, and the acquisition of grammatical morphemes. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition.* Vol. 5. Mahwah: Erlbaum.

Pfeiler, B. (2002). Noun and verb acquisition in Yucatec Maya. In M. Voeikova & W. U. Dressler (eds.), *Pre- and protomorphology: early phases of morphological acquisition in nouns and verbs.* München: Lincom.

Ramat, P. (1995). Typological comparison: towards a historical perspective. In M. Shibatani & T. Bynon (eds.), *Approaches to linguistic typology*. Oxford: Clarendon Press.

Savickienė, I. (2003). *The acquisition of Lithuanian noun morphology*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Sezer, E. (1983). On non-final stress in Turkish. *Journal of Turkish Studies* **5**, 61-69.

Sgall, P. (1993). Skaličkas Sprachtypologie und ihre Fortsetzungen. *Sprachtypologie und Universalienforschung* **46**, 318-329.

Sgall, P. (1995). Prague school typology. In M. Shibatani & T. Bynon (eds.), *Approaches to linguistic typology.* Oxford: Clarendon Press.

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

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (eds.), *Studies in child language development.* New York: Holt, Rinehart and Winston.

Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition.* Vol. 2. Hillsdale: Erlbaum.

Slobin, D. I. (1997). The universal, the typological and the particular in acquisition. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition*. Vol. 5. Mahwah: Erlbaum.

Smoczyńska, M. (1985). The acquisition of Polish. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition*. Vol. 1. Hillsdale: Erlbaum.

Snow, C. (1989). Understanding social interaction and language acquisition: sentences are not enough. In M. H. Bornstein & J. S. Bruner (eds.), *Interaction in human development.* Hillsdale: 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.), *Handbook of child language.* Oxford: Blackwell.

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

Stephany, U. (1997). The acquisition of Greek. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition.* Vol. 4. Mahwah: Erlbaum.

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.

Voeikova, M. D. (2002). The acquisition of case in typologically different languages. In M. Voeikova & W. U. Dressler (eds.), *Pre- and protomorphology: early phases of morphological acquisition in nouns and verbs.* München: Lincom.

Voeikova, M. D. & Dressler, W. U. (2002) (eds.). *Pre- and protomorphology: early phases of morphological development in nouns and verbs.* München: Lincom.

Voeikova, M. D. & Savickienė, I. (2001). The acquisition of the first case oppositions by a Russian and a Lithuanian child. *Wiener Linguistische Gazette* **67-69**, 165-188.

Welch, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika* **38**, 330-336.

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

Xanthos, A. & Gillis, S. (in prep.). Quantifying the development of inflectional diversity.

PART II

Acquisition of diminutives

Diminutives as pioneers of derivational and inflectional development – a cross-linguistic perspective

Ineta Savickienė, Wolfgang U. Dressler, Virág Barcza, Péter Bodor, Nihan Ketrez, Katharina Korecky-Kröll, Marijan Palmović, Ursula Stephany, and Evangelia Thomadaki

Introduction

Although diminutives are commonly viewed as being typical of child speech and child-directed speech, their acquisition has so far neither been studied in a cross-linguistic perspective nor been related to recent theoretical developments in the study of diminutives and of the associated evaluative classes of augmentatives and pejoratives (Dressler & Merlini Barbaresi, 1994, 2001; Jurafsky, 1996). However, the cross-linguistic study of the development of diminutives is apt to shed light on much debated theoretical issues in the acquisition of morphology and on the impact of language typology on acquisition. We expect productivity, morphological transparency and salience in the input language to favour diminutive acquisition, which, in turn, may even facilitate the development of inflectional morphology. Our paper is meant to address these issues based on extensive longitudinal child data from typologically different languages. The data are transcribed and morphologically coded according to CHILDES (MacWhinney, 2000) within the "Cross-linguistic Project on Pre- and Protomorphology in Language Acquisition". The main languages to be considered in this paper are the Indo-European inflecting-fusional languages Lithuanian, Croatian, Greek and German, as well as the agglutinating languages Turkish and Hungarian.

Adult language systems

Languages differ in the richness of diminutives in several ways. First, in the generality of application to different word classes. Thus, diminutives can be formed from many more word classes in Lithuanian than in Hungarian (see Table 1, row 1). The most productive diminutive formations of the six languages, all of them suffixes, are enumerated in row 2 (e.g., Lithuanian masc. suffixes *-elis, -ukas, -utis, -ytis* and fem. *-elė, -ukė, -ytė, -utė*). Diminutives must be distinguished from hypocoristics derived from personal names, although the same suffixes may be used for both (row 3). Hypocoristics may be used by children and caretakers as names designating family members; thus, 'mummy' is often used as a name-like form of address (row 4). A property that is characteristic of languages rich in diminutives, is multiple suffixation, either simultaneous suffixation of different suffixes (row 5) or recursiveness (repetition) of the same suffix (row 6). Here again, Lithuanian appears to be richer than the other languages. For example, from *puodas* 'pot' a diminutive with six suffixes, namely *puod-el-ait-uk-ėl-yt-ėl-is* 'cup-DIM (6)' can be formed, which is known from folk tales. Diminutive suffixes may contrast with other evaluative suffixes, such as augmentative or pejorative ones (row 7). In some languages gender and word class of the base are maintained in the diminutive while in others one of them or both may change (row 8).[1] Therefore, in German, diminutives are uniform and easy to identify, because they are always nouns and always neuters. The most productive Greek diminutive suffix *-aki* also neutralizes gender distinctions since all diminutives ending in *-aki* are neuter.

[1] In German, the gender of the base is not maintained (except, of course, in neuters).

Table 1. Diminutives in the target adult languages: Lithuanian, Croatian, Greek, Austrian German, Turkish, Hungarian

	Lithuanian	Croatian	Greek	Austrian German	Turkish	Hungarian
1. Word classes	nouns (even toponyms), adjectives, adverbs, verbs, interjections, numerals, colloquial greetings	nouns, adjectives, verbs, adverbs	nouns, adjectives, adverbs, numerals, colloquial greetings	nouns, rare derivations from adjectives and numerals, inter-rogatives, colloquial greetings	nouns, cardinal numerals, adverbs, con-junctions, adjectives, locative pronouns, interjections, colloquial greetings	nouns, rarely from adjectives, pronouns, numerals, colloquial greetings
2. Most productive suffixes	*-elis, (i)ukas, -utis, -ytis* (m.), *-elė, -(i)ukė, -utė, -ytė* (f.)	*-ić, -čić* (m.), *-ica, -čica* (f.)	*-aki* (n.), *-ula, -itsa* (f.), *-ulis* (m.) innovative use of *-ina* (f.)	*-i, -erl, -chen, -lein, -li* (n.)	*-CIK, -CA, -iş / -oş* the last two are hypo-coristics	*-cska/-cske*, *-ka/-ke, -i*
3. Hypocoristics	*Rūt-a* → *Rūt-elė, Rūt-ytė, Saulius*→ *Saul-iukas, Saul-ytis*	*Antonija* → *Toni, Tončica or Tonija, Renata* → *Renatica, Helena* → *Helenica*	*Dhimitris* → *Dhimitrakis* → *Takis, Meri* → *Mer-ula* → *Rula, Konstandinos* → *Kostas/ Kostis/Dinos*	*Anna* → *Ann-i, Ann-erl, Hans* → *Hans-i, Hans-l, Häns-chen*	*Nihan* → *Nih-os Fadime* → *Fa-dik, Mustafa* → *Musti*	*István* → *Ist-i-(ke), Pista, István-ka*
4. Special names in CDS	*mama* → *mam-ytė* 'mummy', *tėvas* → *tėv-elis* 'daddy'	*mama* → *mamice* 'mummy'	*papus* → *pap-ulis* 'grandpa', *mama* → *mam-aka* 'mummy'	*Papa* → *Pap-i, Pap-i-lein, Pap-i-li* 'daddy'	very rare	*anya* → *any-u, any-u-ci, any-u-ci-ka* 'mummy'
5. Multiple suffixation	*dalis* → *dal-el-yt-ė* 'particle', *žmog-el-iuk-as* 'man'	-	*mam-ak-ul-itsa* 'mummy', *Takis* → *Tak-ulis, Rula* → *Rul-itsa*	only in hypo-coristics: *Ann-i-li, Ann-i-lein*	*büyük-ce-cik* 'big' [two different diminutive suffixes]	*mam-i-ka* 'mummy'
6. Recursiveness (repetition)	*žmog-el-ėl-is* 'man' and hypo-coristics	*jagod-ič-ic-a* 'small fingertip'	-	-	*mini-ci-cik* 'small' [the same suffix]	-
7. Augmentatives and pejoratives	only pejoratives have their own suffixes, e.g. *bern-iukas - bern-iūkštis* 'boy'	augmen-tatives *- ina, - čina, -etina, - urina*	*-ára*, *-úkla*, *-aros*	-	-	-

Table 1 (continued). Diminutives in the target adult languages: Lithuanian, Croatian, Greek, Austrian German, Turkish, Hungarian

	Lithuanian	Croatian	Greek	Austrian German	Turkish	Hungarian
8. Word class shift / Category shift	class shift to nouns from numerals, gender shift	-	possible gender shift: *bala* (f.) 'ball' → *bal-aki* (n.), along with *bal-itsa* (f.)	m./f. base → neuter dim., gender is constant in hypo-coristics in *-i,* Adj., Num. → N	*büyük-çük* 'big-dim' [the big one] adjective → noun [with stress shift]	no gender; word-class maintained
9. Inflectional class shift: unprod. → prod.	fem. in *-is*: *žuvis* → *žuv-ytė* 'fish' masc. in *-is*: *dantis* → *dant-ukas* 'tooth' masc. in *-uo*: *vanduo* → *vand-en-ukas* 'water'	suffix fem. *-(č)ica*: *stvar* (Gen. *stvari)* 'thing' → *stvarčica* (Gen. *stvarčice*)	neut. *-os* , e.g. Nom./Acc. Sg. *dhasos,* Gen.Sg. *dhasus* Nom./Acc. Pl. *dhasi,* Gen. Pl. *dhason* 'forest' → Nom./Acc. Sg. *dhas-aki* Nom./Acc. Pl. *dhasakia* (no genitive)	marginally	no inflectional class in the language; all declension is uniform and productive	e.g. *ló,* Pl. *lov-ak* 'horse' → *lovacska,* Pl. *lovacská-k*
10. Transparency	*dviratis, dviračio, dviratį, dviračiui* → *dviratukas*, *dviratuko*, *dviratukui* 'bike', etc. palatali-zation	*ruka*, *ruci*, *ruku* → *ručica, ručice, ručici*, *ručicu* 'hand', etc. palatali-zation	less → more transparent: *kreas* → *kreataki* 'meat' no change: *ora* → *oritsa* 'hour' more → less transparent: *alepu* → *alepudhitsa* 'fox'	marginally	nearly all inflection is transparent anyway	diminutives are not necessarily more transparent than their corres-ponding simplicia
11. Stress shift	*knygà* → *knygẽlė* 'book' (always final trochees, except Sg. Inst., Loc., Pl. Acc.)	basically no, except in one-syllable word; change is in tone, not in position	nearly always final trochees and stable stress	always final trochees, stress stable	mostly final trochees	general word-initial stress

Diminutives belong to productive inflectional classes, therefore bases of unproductive classes switch to productive inflection when diminutivised (row 9). Thus, the Lithuanian inflection class of the simple noun *dantis* 'tooth' is unproductive, whereas the derived diminutive belongs to a productive class. Since Turkish inflection is always productive, diminutivisation does not provide such an advantage in this language. Likewise, in Lithuanian, Croatian and partly in Greek the inflection of diminutives may be more

transparent than that of their simple bases (row 10).[2] For example, in Croatian, root-final *-k* in the Nom. Sg. form *ruka* 'hand' changes to *-c* in the Gen. Sg. *ruci*, whereas in the diminutive the stem-final *-c* of *ručica* remains unchanged throughout the paradigm. In Greek, the stem of *kreas* (Nom./Acc. Sg.) 'meat' is *kreat-* in the Gen. Sg. and in the Pl. whereas it is stable in the diminutive paradigm (*kreat-aki*). Thus, the derivation of diminutives often leads to inflectional simplification or inflection class shift (row 9). In Greek, however, the inflection of diminutives may also be less transparent than that of their bases. Thus, the inflectional paradigm of the simple noun *alepu* (Nom./Acc. Sg.) 'fox' is transparent (*alepus* Gen. Sg., *alepudhes* Nom./Acc. Pl., *alepudhon* Gen. Pl.) while the diminutive paradigm *alepudhitsa* (Nom./Acc. Sg.) is not, as demonstrated by Gen. Sg. *alepudhitsas* vs. Nom./Acc. Pl. *alepudhitses.* In several of the languages studied, the derivation of diminutives leads to prosodically more uniform words with final trochees, which are therefore more salient (row 11). In Lithuanian, for example, a base form may be stressed on the final syllable as in *knygà* 'book', whereas in diminutives nearly all case forms are accented on the penultimate. This prosodic characteristic of diminutives may represent another advantage for the early development of diminutive formation.

Data

Table 2 summarizes basic information on our data. Lithuanian, Croatian, Greek and Turkish use data of one child (girl), whereas German and Hungarian provide data on two children, a girl and a boy.

Table 2. Diminutives in eight corpora from six different languages

Language	Child	Age range	Duration (in hours)	Diminutives (types)	Diminutives (tokens)
LITH	Rūta (girl)	1;7 - 2;5	35	249	4754
CRO	Antonia (girl)	1;3 - 2;8	42	184	265
GRK	Anna (girl)	1;8 - 3;0	37.5	249	1664
GER	Jan (boy)	1;3 - 2;8	33	24	167
	Katharina (girl)	1;6 - 3;0	14	30	377
TURK	Deniz (girl)	1;3 - 2;4	9.5	4	7
HUN	Panna (girl)	1;11 - 2;11	41	40	414
	Miki (boy)	1;11 - 2;11	44	80	805

The data analysed in these corpora cover the period from the onset of morphology until the latter part or end of their third year of life. Most of the data were drawn from recordings of 33 to 44 hours, except for the Turkish girl Deniz (9.5 hours) and the Austrian girl Katharina (14 hours).

The relative frequency of diminutives (based on their likelihood of occurrence per hour) is higher in Greek and Lithuanian than in the other languages, both type- and tokenwise (GRK 6.6 types / 44.4 tokens, LITH 7.1 types / 135.8 tokens). The Austrian children use less than 1 diminutive type per hour on average and the Turkish girl does so both type- and tokenwise (see Table 3). The Hungarian children and the Austrian girl Katharina use more diminutive tokens than the Croatian girl.

[2] In Greek, there are also inflectional classes which switch in the opposite direction and others without any difference in transparency.

Table 3. Average usage of diminutives both type- and tokenwise

Child	Types per hour	Tokens per hour	Types/Tokens
Rūta (LITH)	7.1	135.8	0.05
Antonia (CRO)	4.4	6.3	0.70
Anna (GRK)	6.6	44.4	0.15
Jan (GER)	0.7	5.0	0.14
Katharina (GER)	0.5	26.9	0.02
Deniz (TURK)	0.4	0.7	0.57
Panna (HUN)	1.0	10.4	0.10
Miki (HUN)	1.8	18.3	0.10

The development of the semantics and pragmatics of diminutives and their possible role in the development of inflection

Semantics and pragmatics

Adult diminutives have two different basic meanings, the semantic meaning of smallness and a pragmatic meaning indicating not only endearment, sympathy, empathy and pleasure, but also irony. Diminutives furthermore mark child-centered speech situations and serve the mitigation of face-threatening speech acts and the like. The question whether the semantic meaning of smallness rather than pragmatic meanings is basic in diminutives has been much debated in the literature (e.g. Kiefer, 2004; see Bates & Rankin, 1979 vs. Dressler & Merlini Barbaresi, 1994, 2001).

As far as child speech is concerned, no semantic distinction of diminutives as compared to simple nouns is identifiable before contrasts between given diminutives and their base forms emerge. But even after the emergence of contrasts of a given base and the diminutive derived from this, the observer may be unable to ascertain any meaning difference in the child's output between a diminutive and simple noun occurring in the same speech situation. This is the case for some of the children studied during a period lasting for several months (Rūta (LITH) 1;8 - 2;2, Panna (HUN) 2;0 - 2;7, Miki (HUN) 2;3 - 2;11, Anna (GRK) 1;9 - 2;5 (with one exception at 1;9, see below), Deniz (TURK) 1;7 - 1;9).

In view of this situation the question is whether children produce diminutives alternating with their base forms merely for the sake of playfulness marking child-centered speech situations. This assumption poses a problem for Clark's principle of contrast (Clark, 1993: 69-83, 1995). According to this principle children should rapidly develop a semantic or pragmatic meaning distinction between diminutives and their base nouns used alongside each other. As will be shown below, this appears not to be the case, however.

In most of the six languages studied, diminutive formation is the first pattern of word formation to emerge (see Stephany, 1997b and Thomadaki & Stephany, submitted, for Greek; Savickienė, 1998, 2003 for Lithuanian). In those cultures in which diminutives are frequently used by adults in child-centered situations, children will encounter them very often and may use them as 'names' for things just as they would use the respective base nouns if the latter were preferred by adults in their child-directed speech.

First pragmatic effects of the use of diminutives versus base nouns can be detected in the speech of the Austrian girl Katharina (2;0) and the Lithuanian girl Rūta (2;3), who both show a strong tendency to use the hypocoristic 'mother-DIM' form in positive interactions with their mothers but the base noun 'mother' in negative ones.

Another kind of pragmatically oriented choice between hypocoristics and base nouns is found in Greek. At 1;11, the Greek girl Anna uses hypocoristics and diminutives non-standard in adult speech expressing endearment such as *aghap-in-es* 'love-DIM-PL' to refer to her twin

sisters but does not use them for inappropriate self-reference. In a conflicting interaction with his mother at 1;9.23, the Austrian boy Jan even produces a back-formation: Insisting on finally getting his *muesli* [my:sli] 'cereals', he asks for *mues* [my:s]. The earliest Hungarian example of the strategic use of a diminutive is found in an indirect request made by the girl Panna at 2;8 (see Example 1). Here, the girl is using the diminutive strategically for impressing the mother to give her what she would like to have.

(1) Panna (2;8)
Ott, szejetném [: szeretném] nusikám [: nyuszikám] lenne, de szejetném [: szeretném].
There – I would like – bunny-DIM-POSS – would be – so much – I would like
'there, I would like if I had a bunny, I would like it so much'.

There is no evidence for the pragmatic function of diminutives to be found in the Turkish data, since diminutives are very rarely used in the Turkish input. In Croatian, such evidence only emerges after the first months of observation (at 1;11 and 2;2).

Thus, contrary to what has been claimed in the literature (Avrutin, 1999; Ladegaard, 2003), at least first pragmatic strategies can be acquired by children well before the age of three (see also Stephany, 1997a: 298 on the use of the imperative as opposed to the subjunctive in Greek requests observed at 1;9).

At least as far as Hungarian, German and Lithuanian are concerned, the semantic meaning of smallness is at first most often expressed by adjectives and only later by adjectives accompanying diminutivised nouns or by diminutives as such. Lithuanian examples not only show this but also the early use of diminutivised forms of the adjective 'small': *mažytė, mažytė* 'small-DIM, small-DIM' (1;8), *te maža katytė buvo* 'there was a small cat-DIM' (1;11), *maža katytė, o cia šuniukas* 'small cat-DIM, and this is a dog-DIM' (2;3).

The examples in the Greek corpus do not completely conform to this acquisitional sequence. The construction of a diminutive with the adjective 'small' occurs already at 1;10 as in *arkudhaki mikro* 'bear-DIM small', *esi echis meghalo cheri, esi echis mikro* 'you have big hand, you have small' (2;11), *mikro pedhaki* 'small child-DIM' (3;0).

There is an isolated but interesting piece of evidence for the early emergence of the semantic meaning of smallness of hypocoristics in the German data at 1;9, when the Austrian boy Jan, instead of referring to his elder brother by the habitual hypocoristic form *Paul-i* calls him *Paul* thereby insisting that he is a big boy. Note that this exception concerns a hypocoristic name where the semantic dimension of smallness is usually irrelevant.[3]

In addition to such examples, there is ample evidence in the Lithuanian data that the child may use two different diminutives of the same basis without a noticeable difference in meaning during the same month period (number of tokens and ages are indicated in parentheses): mešk-iukas (5) = mešk-utė (6) = mešk-ytė (1) 'teddy bear' (1;8); *kač-iukas* (1) = *katin-ėlis* (1) = *kat-ytė* (10) 'cat' (1;9); *mašin-ėlė* (1) = *mašin-ytė* (3) 'car' (1;9); *arbat-ėlė* (3) = *arba-tytė* (2) 'tea' (1;10); *kišk-utis* (4) = *kišk-ytis* (1) 'hare' (2;0); *rank-utė* (1) = *rank-ytė* (1) 'hand' (2;0). We also have the evidence of free variation within the same micro-context where the Lithuanian girl uses two different diminutives (see Examples 2 and 3). Note also the girl's own name: *Rūt-ytė* = *Rūt-elė* (starting with 1;7) (Example 4).

(2) Rūta (1;8)
Mother: *O kas čia?* 'What is here?'
Rūta: *Mešk**utė**.* 'Teddy-bear-DIM-FEM'
Father: *Mešk**utis.*** 'Teddy-bear-DIM-MASC'
Rūta: *Mešk**iukas**.* 'Teddy-bear-DIM-MASC'

[3] Note also that in the above-mentioned Greek example *aghap-in-es* 'love-DIM-PL' an additional semantic meaning may not be totally excluded.

(3) Rūta (1;9)
Rūta: *Čia kitas katin**ėlis**.* 'Here is another cat-DIM-MASC'
Mother: *Kač**iukas**.* 'Cat-DIM-MASC'
Rūta: *Kat**utė**.* 'Cat-DIM-FEM'

(4) Rūta (2;0)
Mother: *O koks vardas mano mažytės?* 'What is the name of my little one?'
Rūta: *Rūt**ytė**, Rūta*. 'Rūta-DIM, Rūta'
Mother: *Rūt**ytė**, Rūta?* 'Rūta-DIM, Rūta?'
Rūta: *Rūta, Rūt**ytė**, Rūt**elė***.
Mother: *Tave galima vadinti Rūta, Rūt**ytė**, Rūt**elė**, taip?* 'Is possible to call you
Rūta, Rūta-DIM, Rūta-DIM, isn't it?'
Rūta: *Taip.* 'Yes'

Examples of an apparently indiscriminate use of different hypocoristics based on the same name exist in the child's speech and are especially frequent in child-directed speech in other languages as well, i.e. Croatian (*Ant-onija, Ton-čica*, *Ant-ica*), Greek (*An-ula, An-ul-itsa*), Hungarian (*Panna, Panka, Panni*), but there are very few or no examples where a child might use two different diminutives of the same basis (a single example for Greek of two diminutives derived from the same base of a common noun co-varying within one and the same speech situation: 2;4 *portula/portitsa* 'door-DIM').

All this shows that Clark's principle of contrast runs into problems during the early phases of the acquisition of diminutives, since for many months (sometimes even for more than one year), diminutives appear to be used alongside their base forms without any meaning difference (the only one German example of *Paul-i* vs. *Paul* may contradict our statement here). If pragmatic meaning is included in the scope of Clark's principle of contrast, the period of apparently indiscriminate use of diminutives (including hypocoristics) and base nouns will be reduced but will still extend over a long period and will thus pose a problem to this principle of lexical development. In view of the evidence found in the languages studied in the present paper the following conclusion may be proposed: Similar to what has been found with the variable use of strongly and weakly inflected forms of English and German verbs, in the early period of derivational development, children appear to be not yet able to properly choose between what appear to be variant forms for designating given referents and at first tend to use them indiscriminately (pace Clark, 1990: 424-5, 1993: 81). Of course, there is the epistemological problem of who has the *onus probandi*: the one who claims a meaning difference or the one who denies it.

How diminutives may simplify the acquisition of inflection

In the languages studied in the "Cross-linguistic Project on Pre- and Protomorphology in Language Acquisition", diminutive formation becomes productive at the same time as the first inflectional contrasts in nouns and verbs (and compounding in German). Not only do diminutives have an important role to play in the development of derivational morphology as the first derivational patterns to emerge but also they may even simplify the acquisition of declension in certain languages and may show a precocious development as compared to that of their base forms (Olmsted, 1994; Laalo, 1998; Savickienė, 2001, 2003; Kempe et al., 2003).

Thus, the transfer of a base noun from an unproductive and/or opaque declension class of the base to a productive and transparent class of the respective diminutive can explain why the latter type of declension is acquired earlier, see the following Lithuanian, Croatian and

Hungarian examples (Examples 5-7). (For different structural reasons no evidence can be found for this claim in German and Turkish).

(5) Lithuanian
masc. Nom. Sg. *vanduo* 'water', Gen. *vandens* vs.
masc. Nom. Sg. *vanden-ukas*, Gen. *vanden-uko* 'water-DIM'

(6) Croatian
fem. Nom. Sg. s*tvar* 'thing', Gen. *stvari* vs.
fem. Nom. Sg. *stvar-čica*, Gen. *stvar-čice* 'thing-DIM'

(7) Hungarian
Sg. *ló* 'horse', Pl. *lov-ak* vs.
Sg. *lov-acska*, Pl. *lov-acská-k* 'horse-DIM'

As far as Greek is concerned, there is no evidence for this claim since there is only one form each of the simple non-transparent noun *kreas* 'meat' (2 tokens) and the respective diminutive *kreataki* (1 token) and no example of the diminutive *dhasaki* 'wood-DIM' derived from the unproductive and non-transparent base *dhasos* 'wood' (7 tokens Sg. *dhasos*, 1 token Pl. *dhasi*). However, diminutives derived from most of the nouns belonging to these unproductive declensional classes are not commonly used by adults either.

A productive declension class is one which is freely used in integrating loan-words, inflecting abbreviations and other neologisms. As opposed to productive, transparent declension classes, non-transparent ones include an opacifying morphological or morphonological alternation such as lenition of the stem-final consonant by affrication, as in Lithuanian masc. Nom. Sg. *dviratis* 'bicycle', Gen. Sg. *dviračio* as opposed to the transparent diminutive lacking such an alternation (Nom. Sg. *dvirat-ukas*, Gen. Sg. *dvirat-uko*) (see Dressler et al., 1995-1996; Dressler, 2003). Although there is a tendency for transparent morphological patterns to be more productive than opaque ones (see Hay & Baayen, 2002), transparent patterns may be unproductive (e.g. the English suffix *-dom* in *king-dom*) while opaque ones may be productive (e.g. velar softening of the word-final stop in English *electri*[k] vs. *electri*[s]-*ity*).

Productivity and transparency

According to the theoretical tenets of Natural Morphology (Kilani-Schoch & Dressler, 2005) it may be predicted that, *ceteris paribus*, both child and adult learners as well as mature speakers will prefer productive to unproductive and transparent to non-transparent inflectional morphology. This prediction will be true of infrequent grammatical forms and may even be valid for very frequent irregular and opaque forms at certain stages of first language acquisition. It should follow from this that children will prefer diminutives if their inflectional paradigms are more transparent than those of their bases, but that they will rather choose base forms if these are more transparent than the corresponding diminutives. Clearly there are many disturbing variables which decrease the predictive power of productivity and transparency, even provided that there is a difference between diminutives and their bases in regard to these two criteria.

In Turkish, the inflection of diminutives as well as base nouns is fully productive and there is accordingly no difference in productivity between these two classes. In the other five languages, all or nearly all diminutives belong to productive inflectional classes, whereas the inflectional patterns of many of their bases are unproductive. It therefore seems interesting to investigate the possibility that diminutives derived from unproductive bases are found more

frequently in the speech of children acquiring these languages than diminutives derived from productive ones. This will be done by comparing the ratio of the token frequencies of diminutives divided by their bases in the languages at hand (keeping the variable of transparency constant). For those languages and children where sufficient data exist, the ratios for the variable of productivity are given in Table 4.[4]

Table 4. Diminutives and productivity

Language	Ratios (tokens)	
	D pt / S ut	D pt / S pt
CRO	1.60	0.68
GRK	0.49	0.52
HUN	2.67	0.56
LITH	3.81	2.08

[D = diminutive, S = simple noun, p = productive, u = unproductive, t = transparent]

The results from Croatian, Hungarian and Lithuanian show that children appear to prefer diminutives to base nouns when this involves a switch from an unproductive to a productive inflection class, but this is not the case for Greek, where nearly all simple nouns occurring in child speech as well as diminutives are productive.

Next, let us focus on the variable of transparency (keeping the variable of productivity constant). As shown in Table 5, Croatian, Greek and Hungarian children prefer diminutives instead of base nouns when this involves a switch from an opaque to a transparent inflection class. In Lithuanian, there are too few instances of opaquely inflected nouns and their diminutives in child speech for allowing quantitative analyses. Still, diminutives are almost always preferred in such cases.

Table 5. Diminutives and transparency

Language	Ratios (tokens)	
	D pt / S po	D pt / S pt
CRO	0.93	0.60
GRK	0.60	0.03
HUN	2.90	0.33

[D = diminutive, S = simple noun, p = productive, t = transparent, o = opaque]

Therefore, wherever more productive and transparent diminutives are systematically preferred to their unproductive and opaque bases, we may assume that the acquisition of diminutives simplifies the declensional system and thus facilitates its acquisition. In other words, diminutive formation may be assumed to constitute a derivational boot-strapping device (see Weissenborn & Höhle, 2001) for the acquisition of nominal inflection.

If, however, the simple noun and the respective diminutive both belong to productive and transparent classes, no such boot-strapping effect can be assumed. This seems to be the case with Turkish. Because of their greater length and higher morphological complexity in comparison with their bases, diminutives might prove to be a "disadvantage" to acquisition. Although a reason for the less frequent use of diminutives in Turkish as compared to Hungarian can be sought in the more strongly agglutinating character of the former language, it must not be forgotten that diminutives very rarely occur in the input since they are mainly used as hypocoristics.

In spite of the fact that Austrian German diminutives are at large more productive and transparent than base nouns they cannot in principle facilitate the acquisition of declension for

[4] German and Turkish are not included due to insufficient data for a quantitative analysis.

the following reasons. Not only do they rarely occur in the input of one of the two children, presumably due to pragmatic reasons, but also the only inflectional forms of Austrian German diminutives carrying an overt inflection marker are the Gen. Sg. of masculines and neuters, a very rarely occurring form, and the plurals of the diminutive suffixes *-erl* and *-i* (i.e. *-erl-n, -i-s*). Although the frequency of diminutives is very high in Greek, both in child speech and child-directed speech, their difference in productivity is not significant and also the most frequent diminutive type ending in *-aki* only inflects for number, but not for case.

Salience

Prosodic salience is a potentially relevant factor in the acquisition of diminutives since diminutives may regularize stress patterns and thus simplify the task of segmenting the speech stream into words (Jusczyk, 1997). This would render them easier to identify than simple nouns. As far as the languages studied in the present paper are concerned, no extra salience can be assumed for Hungarian which has obligatory word-initial stress and a non-homogeneous prosodic make-up of diminutives. Their prosodic structure constitutes either a rime only (*-i*, e.g. in *Ist-i*) or one syllable (*-(cs)ka*, e.g. in *Istvan-ka*) or more, when a vowel is inserted before these two suffixes (e.g. in *lov-acska*). Therefore short Hungarian diminutives may be either trochaic or dactylic, when disyllabic or trisyllabic, respectively, whereas longer ones may also have different structures according to secondary accents after the main initial stress. With the exception of Croatian, in the other four languages of our sample, diminutives constitute word-final trochees in their base forms and most of their oblique case forms. This might help to perceive them more easily. Salience could therefore be an additional factor responsible for the greater frequency of diminutives in child-directed speech than in adult-directed speech.

Conclusions

The cross-linguistic study of the development of diminutives in the inflecting-fusional languages Croatian, Greek, German and Lithuanian, as well as the agglutinating languages Turkish and Hungarian leads to the following conclusions:

In all languages investigated diminutives present the earliest emerging derivational category.

The frequency of diminutives in the input as well as the output of our child corpora clearly depends on their pragmatic role in the respective languages. In addition, the higher degree of morphological productivity, morphonological transparency and phonological salience may favour the use of diminutives as compared to base nouns. Only in this case (which does not hold for Turkish) they may facilitate the acquisition of inflection.

There is evidence to be found in our data that in the early phases of development children produce diminutives alternating with their base forms without any identifiable difference in meaning. Both the assumptions that the use of diminutives has at first no distinct meaning or that diminutives simply indicate child-centered speech situations, appear to pose a problem for Clark's principle of contrast. Also children's simultaneous use of different diminutive suffixes attached to the same base without any apparent meaning difference is problematic for this principle. Further evidence from both longitudinal and transversal studies is needed for settling this question.

Evidence for the emergence of the pragmatic meaning of diminutives prior to their semantic meaning comes from all the languages investigated. Contrary to what has been claimed in the literature, first pragmatic strategies can be acquired well before the age of three years.

References

Avrutin, S. (1999). *Development of the syntax-discourse interface*. Dordrecht: Kluwer.

Bates, E. & Rankin, J. (1979). Morphological development in Italian: connotation and denotation. *Journal of Child Language* **6**, 29-52.

Clark, E. (1990). On the pragmatics of contrast. *Journal of Child Language* **17**, 417-432.

Clark, E. (1993). *The lexicon in acquisition*. Cambridge: Cambrige University Press.

Clark, E. (1995). Later lexical development and word formation. In P. Fletcher & B. MacWhinney (eds.), *Handbook of child language*. Oxford: Blackwell.

Dressler, W. U. (2003). Degrees of grammatical productivity in inflectional morphology. *Italian Journal of Linguistics* **15**, 31-62.

Dressler, W. U., Drążyk, R., Drążyk D., Dziubalska-Kołaczyk, K. & Jagła, E. (1995-1996). On the earliest stages of acquisition of Polish declension. *Wiener Linguistische Gazette* **53-54**, 1-21.

Dressler, W. U. & Merlini Barbaresi, L. (1994). *Morphopragmatics: diminutives and intensifiers in Italian, German, and other languages*. Berlin: Mouton de Gruyter.

Dressler, W. U. & Merlini Barbaresi, L. (2001). Morphopragmatics of diminutives and augmentatives: on the priority of pragmatics over semantics. In I. Kenesei & R. Harnish (eds.), *Perspectives on semantics, pragmatics and discourse*. Amsterdam: Benjamins.

Hay, J. & Baayen, H. (2002). Parsing and productivity. *Yearbook of Morphology 2001*, 203-235.

Jurafsky, D. (1996). Universal tendencies in the semantics of the diminutive. *Language* **72 (3)**, 533-578.

Jusczyk, P. W. (1997). *The discovery of spoken language.* Cambridge, MA: MIT Press.

Kempe, V., Brooks, P., Mironova, N. & Fedorova, O. (2003). Diminutivization supports gender acquisition in Russian children. *Journal of Child Language* **30**, 1-15.

Kiefer, F. (2004). Morphopragmatic phenomena in Hungarian. *Acta Linguistica Hungarica* **51**, 325-349.

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

Laalo, K. (1998). Diminutives in Finnish child-directed and child speech. In S. Gillis (ed.), *Studies in the acquisition of number and diminutive marking* (= Antwerp Papers in Linguistics 95), 138–148.

Ladegaard, H. J. (2003). Politeness in young children's speech. Context, peer group influence and pragmatic competence. *Journal of Pragmatics* **36**, 2003-2022.

MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk.* Mahwah: Erlbaum.

Olmsted, H. (1994). Diminutive morphology of Russian children: a simplified subset of nominal declension in language acquisition. In C. Gribble & R. Rothstein (eds.), *Alexander Lipson: in memoriam*. Columbus: Slavica Publishers.

Savickienė, I. (1998). The acquisition of diminutives in Lithuanian. In S. Gillis (ed.), *Studies in the acquisition of number and diminutive marking* (= Antwerp Papers in Linguistics 95), 115-135.

Savickienė, I. (2001). The role of diminutives in Lithuanian child language acquisition. *Linguistica Baltica* **9**, 109-118.

Savickienė, I. (2003). *The acquisition of Lithuanian noun morphology*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Stephany, U. (1997a). The acquisition of Greek. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition*. Vol. 4. Mahwah: Erlbaum.

Stephany, U. (1997b). Diminutives in early Greek. In W. U. Dressler (ed.), *Studies in pre- and protomorphology*. Wien: Verlag der Österreichische Akademie der Wissenschaften.

Thomadaki, E. & Stephany, U. (submitted). Diminutives in Greek child language.

Weissenborn, J. & Höhle, B. (2001) (eds.). *Approaches to bootstrapping: phonological, syntactic and neurological aspects of early language acquisition*. Amsterdam: Benjamins.

List of contributors

AYHAN AKSU-KOÇ, Boğaziçi University, Istanbul, Turkey; e-mail: koc@boun.edu.tr

VIRÁG BARCZA, Hungarian Academy of Sciences, Budapest, Hungary; e-mail: bvirag@nytud.hu

PÉTER BODOR, University of Budapest, Budapest, Hungary; e-mail: bodorp@hotmail.com

ANASTASIA CHRISTOFIDOU, Greek Academy of Sciences, Athens, Greece; e-mail: christo@academyofathens.gr

WOLFGANG U. DRESSLER, Austrian Academy of Sciences & University of Vienna, Vienna, Austria; e-mail: wolfgang.dressler@univie.ac.at

NATALIA GAGARINA, Center for General Linguistics and Typology (ZAS), Berlin, Germany; e-mail: gagarina@zas.gwz-berlin.de

STEVEN GILLIS, University of Antwerp, Antwerp, Belgium; e-mail: steven.gillis@ua.ac.be

GORDANA HRZICA, University of Zagreb, Zagreb, Croatia; e-mail: gordana.hrzica@zg.htnet.hr

NIHAN F. KETREZ, Yale University, New Haven (Connecticut), USA; e-mail: nihan.ketrez@yale.edu

MARIANNE KILANI-SCHOCH, University of Lausanne, Lausanne, Switzerland; e-mail: marianne.kilanischoch@unil.ch

KATHARINA KORECKY-KRÖLL, Austrian Academy of Sciences, Vienna, Austria; e-mail: katharina.korecky@easyon.at

MELITA KOVAČEVIĆ, University of Zagreb, Zagreb, Croatia; e-mail: melita.kovacevic@public.srce.hr

SABINE LAAHA, Austrian Academy of Sciences, Vienna, Austria; e-mail: sabine.laaha@oeaw.ac.at

KLAUS LAALO, University of Tampere, Tampere, Finland; e-mail: klaus.laalo@uta.fi

MARIJAN PALMOVIĆ, University of Zagreb, Zagreb, Croatia; e-mail: mpalmovic@yahoo.com

BARBARA PFEILER, National Autonomous University of Mexico, Mexico City, Mexico; e-mail: bpfeiler@avantel.net

INETA SAVICKIENĖ, Vytautas Magnus University, Kaunas, Lithuania; e-mail: i.savickiene@pmdi.vdu.lt

URSULA STEPHANY, University of Cologne, Cologne, Germany; e-mail: am001@uni-koeln.de

EVANGELIA THOMADAKI, Democritus University of Thrace, Komotini, Greece; e-mail: ninthom@otenet.gr

MARIA D. VOEIKOVA, Russian Academy of Sciences, St. Petersburg, Russia; e-mail: m_voeikova@iling.nw.ru

ARIS XANTHOS, University of Lausanne, Lausanne, Switzerland; e-mail: aris.xanthos@unil.ch