Exploring the facilitating effect of diminutives on the acquisition of Serbian noun morphology

Nada Ševa

Thesis submitted for the degree of Doctor of Philosophy

Department of Psychology

University of Stirling

August 2006

Table of Contents

List of figures

List of tables

Acknowledgments

I owe my special gratitude to my supervisor Dr Vera Kempe for her unconditional support and help both professionally and personally during these three years. Throughout our daily collaboration, she always inspired and motivated me to learn and seek for more. And of course, she is the most responsible for my tango addiction.

I would also like to thank my second supervisor Dr Peter Hancock for his help during my Phd studies.

I feel deeply indebted to Professor Patricia Brooks (City University of New York), whose guidance and advice played a very important role in writing of this thesis.

I would like to thank Darinka Anđelković and Professor Aleksandar Kostić from the Laboratory of Experimental Psychology, University of Belgrade and Professor Draga Zec from Cornell University, who inspired and supported me to enter the field of psycholinguistics.

I thank all my Belgrade friends: Ana, Milena, Maja, Dušica, Oliver, Nevena, Brana, Petar, Jelena Morača and of course Petar Đuričin for being always there for me.

Also, I would like to thank all my colleagues from the University of Stirling, especially my room colleagues Sonja, Rachael and Adel.

I thank my family in Serbia and my husband Srdjan, for their endless love and support.

This work is dedicated to my grandmother Mira, mother Ljiljana and father Srbislav.

Declaration

I hereby declare that the research reported in this thesis is my own work unless otherwise stated. No part of this thesis has been submitted for a degree at another university.

Parts of Chapter 3 were presented as Ševa, Hadjiconstantinou, & Kempe (2005) and are to be published in Ševa, Kempe, Brooks, Mironova, Pershukova, & Fedorova (in press). Parts of Chapter 4 are to be published in Ševa et al. (in press) and were presented as Ševa, Kempe, & Brooks (2005) and Ševa & Kempe (2005). Parts of Chapter 5 are published as Ševa, Kempe, & Brooks (2006-a) and were presented as Ševa, Kempe, & Brooks (2006-b).

Nada Ševa

Abstract

Studies of Russian, Polish, and Lithuanian language learners converge on the finding that morphological features of nouns are first generalized to word clusters of high morpho-phonological similarities such as diminutives, that grammatical categorisation is are more easily applied to novel words that fall into these clusters. The present thesis explores whether the facilitating effect of diminutives on the acquisition of complex noun morphology can be extended to Serbian, a south Slavic language, morphologically similar to Russian and Polish. Specifically, the thesis explores the role of parameters responsible for the obtained diminutive advantage: high frequency of a particular cluster of words in child-directed speech (CDS) and morpho-phonological homogeneity within this cluster.

A corpus analysis of the distribution of diminutives in Serbian CDS indicated a rather unexpected difference in frequency relative to Russian and Polish CDS, despite the high similarity of the diminutive derivation across these three Slavic languages. Out of the total number of nouns in Serbian CDS only 7% were diminutives, compared to 20-30% in Polish and 45% in Russian.

Two experimental studies explored whether the low frequency of diminutives in Serbian CDS attenuates the diminutive advantage in morphology learning compared to Russian and Polish. In the first two experiments, Serbian children exhibited a strong diminutive advantage for both gender agreement and case marking in the same range as Russian children, indicating that morpho-phonological homogeneity within the cluster of diminutives may play as important a role as their frequency for grammatical categorisation of novel nouns.

A third study investigated in more detail the effects of morpho-phonological homogeneity on the emergence of the diminutive advantage using a gender-agreement task with novel nouns in simplex and pseudo-diminutive form over four sessions with Serbian children. The results showed a pseudo-diminutive advantage for gender agreement by Session 2, suggesting that the categorisation of nouns into grammatical categories is based on morpho-phonological homogeneity of the word cluster, emerges relatively fast, and can occur despite the much lower frequency of diminutives in Serbian CDS.

Finally, a series of neural network simulations designed to capture the pattern of results from the third experimental study was used to examine to what extent a simple associative learning mechanism, relying on morpho-phonological similarity of the noun endings, can explain the findings. The performance of three models, a whole-word feed-forward network, a Simple Recurrent Network (SRN) and a last-syllable feed-forward network, was compared to the experimental data. The superior fit of the SRN suggests that gender learning is based on a very fast sequential build-up of representations of the entire word, allowing the system to exploit the predictive power of word stems to anticipate regularised endings.

Overall, the findings of this thesis contribute to our general understanding of mechanisms responsible for the acquisition of complex inflectional noun morphology in two ways. First, by extending experimental studies and neural network simulations to Serbian, the results underline the universality of the idea that noun morphology is learned and processed through a single-route associative mechanism based on the frequency and morpho-phonological structure of nouns. More specifically, the results from experimental studies and neural network simulations demonstrate that for diminutives, the low-level grammatical categorisation is based mainly on the morpho-phonological similarity of word endings, and can emerge after just a few exposures.

And second, the neural network simulations suggest that during the process of categorisation of nouns into gender categories, learners rely not only on predictable information from the noun endings, but also on phonological regularities in the stems of nouns. Taken together, these findings contribute also to a better understanding of the facilitating role of CDS in morphology acquisition.

Chapter 1

1. Introduction

1.1. Aims and outline of the thesis

A large number of studies in the last three decades demonstrated that child-directed speech (CDS) has a facilitating effect on language learning, ranging from the level of phonetics/phonology to the level of semantics and pragmatics. Interestingly, the positive effect of CDS on the acquisition of inflectional morphology was only demonstrated in recent cross-linguistic research, despite the fact that learning in this domain was in the focus of developmental psycholinguistics for decades. More specifically, in a series of experimental studies, it was shown that diminutives, one of the most prominent features of this register, facilitate the acquisition of complex inflectional systems in languages like Russian, Polish and Lithuanian.

This thesis explores the distribution of diminutives in CDS, and the possible facilitating effect of this derivation on the acquisition of noun morphology in Serbian, a south-Slavic language with rich inflectional morphology. Serbian CDS is interesting for two reasons. First, a detailed corpus analysis of Serbian CDS, based on what is probably one of the biggest CDS corpora in the Balto-Slavic languages, will provide us with additional information about the cross-linguistic variability in the distribution of diminutives in CDS. Second, there is a substantial morphological similarity between Serbian and other Slavic languages, like Russian and Polish. This similarity will allow me to replicate directly the diminutive advantage for noun morphology learning observed in experimental studies for Russian, Polish and Lithuanian thereby strengthening the cross-linguistic data base for this phenomenon. Moreover, I will use the experimental findings on Serbian children as a starting point for the further

exploration of the nature of the morpho-phonological and distributional factors responsible for the facilitating effect of diminutives on the acquisition of inflectional morphology. Specifically, I will examine whether the facilitating effect of diminutives will be altered by a difference in frequency of diminutives in CDS.

The following parts of this chapter provide a more detailed description of the general characteristics of CDS together with a description of previous corpus studies and experimental work on the facilitating effect of diminutives on the acquisition of languages with complex noun morphology.

Chapter 2 describes Serbian noun morphology, specifically the rich gender and case-marking systems.

Chapter 3 presents the first qualitative and quantitative corpus analysis of diminutive usage in Serbian CDS. In addition to coding of the nouns for the derivational status the nouns are coded for their gender and declension class. The distributional patterns observed for CDS will be compared with gender, declension and derivational distributions in adult-directed speech (ADS) and in written language.

Chapter 4 describes a series of experiments designed to test whether diminutives facilitate the learning of noun gender and case categories in Serbian. The results of the experimental studies will be compared with the results on the facilitating effect of diminutives on the acquisition of noun morphology observed in previous research.

Chapter 5 explores the relationship between frequency of diminutives and increased morpho-phonological similarity of word endings as possible factors which contribute to the facilitating effect of diminutives by using neural network simulations to model gender learning of pseudo-diminutive and simplex nouns in Serbian children.

Finally, Chapter 6 summarises the results obtained in this thesis and discusses them in the broader context of noun morphology acquisition and processing.

1.2. General characteristics of child-directed speech (CDS)

First language acquisition relies to a great extent on the adult capacity to adapt their language production to the level of children's communicative and language learning needs. Recent research has shown that mothers' adjusted responsiveness to early infant's non-verbal expressivity benefits language development. In a series of longitudinal observations of conversations between mothers and their nine and thirteen months old infants, Nicely, Tamis-LeMonda & Bornstein (1999) and Tamis-LeMonda, Bornstein & Baumwell (2001) demonstrated that mothers' affective and verbal sensitivity to the infants' early communicative attempts best predicted the timing of five milestones of the children's early language: first imitations, first words, 50 words in expressive language, combinatorial speech, and the use of language to address the past.

Opposite to that mothers with post-natal depression who did not establish an emotional bond with their children, produced responses which were less affective and less informative in content, in comparison to non-depressed mothers (Reissland, Shepherd, Herrera, 2003; Herrera, Reissland, Shepherd, 2004; Kaplan, Bachorowski & Zarlengo-Strouse, 1999; Kaplan, Bachorowski, Smoski & Zinser, 2001). As a consequence of non-attuned responsiveness towards their children, these mothers failed to promote associative learning in their 4-month old infants (Kaplan et al., 1999; Kaplan, Bachorowski, Smoski, Hudenko, 2002) which is considered to be one of the prerequisites for language learning (Altmann, 2002; Elman, 1993;

Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996; Tomasello, 2003). Consequently, three year old children of depressed mothers spoke less and had a poorer lexicon than children of healthy women (Breznitz & Sherman, 1987). Given how important language and communicative modifications of parental discourse addressed to children is for language learning, it is not surprising that these modifications appear almost universally across different languages and cultures in the form of a special register called child-directed speech (CDS) (Snow & Ferguson, 1977; Ferguson, 1978; Snow, 1985). [1]

CDS relies on the communicative abilities of competent speakers to vary their speech according to conversational situation, topic, medium of communication, role and age of interlocutor, etc. (Hymes, 1974, Hudson, 1996). From an evolutionary perspective, this register evolved primarily as a set of biologically relevant signals which serve to control infant attention and arousal (Fernald, 1992) and to help to create and maintain an emotional bond between parents and infants (Trainor, Austin & Dejardins, 2000). This suggests that the beneficial effects of the CDS register on the acquisition of language are mainly a by-product of those basic biological functions. Presumably, in the continuous process of interaction between parents and children, mother-child communication became partially conventionalised so that CDS represents a mixture of innate child-care behaviours and a set of culturally and socially shaped features which can be transmitted not only from generation to generation, but also to the other registers like pet-talk, talk addressed to lovers, to foreigners, etc. (Ferguson, 1978; DePaulo & Coleman, 1986).

[1] In addition to this term, researchers also use the labels "baby-talk," "parentese," "motherese," "caregiver speech," 'infant-directed speech (IDS)". The terms "baby-talk" and "infant-directed speech" are usually used to specially mark the speech which is addressed to newborn babies or infants.

Based on data from research on CDS characteristics in 27 languages[2], Ferguson (1978) identified a set of 23 language variables, mainly from the domain of phonetics/phonology and prosody, but also at the level of morphology, syntax, lexicon and pragmatics which are universally modified / simplified by parents and non-kin adults when addressing children across different languages and cultures. This set of characteristics has been expanded by subsequent corpus studies of CDS which have highlighted the universality of this register. Table 1.1. represents a short overview of the main universal features (from phonetics and phonology to pragmatics) of CDS register which are listed in the Ferguson's study (1978) and subsequent corpus analyses across different languages and cultures.

[2] Languages in Ferguson's study were: Arabic (Syrian), Aramaic (Neo-), Bengali, Berber, Cocopa, Comanche, Dutch, English, German, Gilyak, Greek, Hidatsa, Hungarian, Japanese, Kannada (Havyaka), Kipsigis, Latvian, Luo, Maltese, Marathi, Pomo, Portuguese (Brazilian), Romanian, Samoan, Serbian/Croatian, Spanish, Tzeltal.

Table 1.1. The list of main features of child-directed speech (CDS) register.

Language domain	CDS features
Phonetics/Phonology	-Higher fundamental frequency and larger frequency range (Ferguson, 1978; Ogle & Maidment, 1993); -Lengthening of vowels (Swanson, Leonard & Gandour, 1992; Swanson & Leonard, 1994); - Lengthening of word-final unstressed syllables (Albin & Echols, 1996); -Lengthening of the voice onset time for the alveolar and velar stops (Englund, 2005); -Greater stretching of vowel space (Kuhl, Andruski, Chistovich, Chistovich, Kozhevnikova, Ryskina, Stolyarova, Sundberg, Lacerda, 1997); -Less segmental variability (Berstain-Ratner, 1996); -General phonological simplification like cluster reduction, liquid substitution, reduplication, etc. (Ferguson, 1977, 1978).
Prosody	-Higher and longer pitch (Fischer & Tokura, 1996; Fernald, Taeschner, Dunne, Papousek, de Boysson-Bardies & Fukui, 1989); -Exaggerated pitch breaks of focused words in utterance-final position (Fernald & Mazzie, 1991); -More informative prosodic patterns (Fernald, 1989). -Longer pauses (Ferguson, 1978; Fisher & Tokura, 1996).

Lexicon/Semantics	-Lexicon limited to certain semantic areas like: kin terms and body parts and functions, games, food, animals, etc. (Ferguson, 1977, 1978; Hayes & Ahrens, 1988); -high percentage of diminutives and hypocoristics (Ferguson, 1978; Gillis, 1998; Dressler, 1997); - proper name used for attention-orienting and instruction to act (Durkin, Rutter & Tucker, 1982).
Syntax	-Shorter sentences (Snow, 1977; Ferguson, 1978); -Parataxis (e.g. listings without conjunctions) (Snow, 1977; Ferguson, 1978); -Telegraphic style (Snow, 1977; Ferguson, 1978); -Repetitions (Hardy-Brown & Plomin, 1985; Baron, 1990).
Pragmatics	-Higher percentage of questions, specially yes/no questions (Gleitman, Newport & Gleitman, 1984); -Frequent use of tags like *OK?*, *hm?* (Ferguson, 1978); -Prolonged pauses between utterances (Bloom, Russell & Wassenberg, 1987); -High degree of redundancy with regard to the referential component of speech – repetition of conversational episodes, words, etc. (Messer, 1980).

In addition to the studies described above which focused mainly on finding similarities between CDS across different languages, there is a growing number of corpus analyses which indicate that CDS register usage differs to some extent not only across languages and different cultures, but also between speakers of the same language, depending on their gender, age, conversational situation, social status, etc. Initial corpus studies on CDS production in languages spoken in communities which are far away from modern westernised conventions about the upbringing of children, showed some differences to the so far almost universally observed features of CDS. For example, no difference between the means and standard deviation of fundamental frequency and pitch range of CDS and ADS was observed in languages like Mi'kmaq (spoken in Nova Scotia) and Quiche Mayan (spoken in the western highland region of Guatemala) (Fee & Shaw, 1998; Bernstein-Ratner & Pye, 1984). Also, Rabain & Sabeau (1997) showed that Wolof mothers tended to introduce 3rd parties into conversations but used little reference to the environment in contrast to French mothers who were more likely to keep the conversation with their children in a dyadic organisation and centred on the immediate physical environment. In addition to these cross-linguistic comparisons, Shute & Wheldall (1989) showed that even in the different variants of English (British vs. North American), mothers tend to vary in the range of vocal pitch, with British mothers having smaller pitch increase when addressing their children in comparison to American mothers.

Other factors which seem to influence CDS are the socioeconomic status of the parents and the gender of the child. With respect to socioeconomic status, Hoff-Ginsberg (1991) and Hoff-Ginsberg & Tardif (1995) showed that mothers with high-socioeconomic status (upper-middle class) tended to use a more complex and advanced lexicon, provided topic-continuing replies to a greater proportion of their

children's utterances and with less directive behaviour in comparison to mothers of lower-socio economic status (working class).

With respect to the gender of the child, it has been shown that mothers of female children talked more, asked more questions, repeated their children's utterances more often, and used longer utterances compared to mothers of male children, who used more directives, clarification requests, feedback and confirmation in their parent-child dyads (Cherry & Lewis, 1976; Stoneman & Brody, 1981; Lanvers, 2004; Da Fonseca & Salomao, 2005; Clearfield & Nelson, 2006). These results indicate that mothers transmit different messages to their male and female infants, both through language and non-verbal interaction which may contribute to infants' gender role development. Additionally, Foulkes, Docherty & Watt (2005) showed that the maternal usage of phonetic variants of the phoneme [*t*] in word-medial and word-final prevocalic contexts depends on the gender of their two-year old children: speech to girls contained more standard variants than speech to boys which, by contrast, contained higher rates of vernacular variants of the same phoneme. This indicates that CDS may play a role in learning the social-indexical values of phonetic features.

Finally, one of the most commonly observed differences in the production of CDS were the variations between mothers and fathers. Tamis-LeMonda, Shannon, Cabrera & Lamb (2004) have demonstrated in a longitudinal study with two and three year old children that both fathers' and mothers' supportive parenting have facilitating effects on language and cognitive development. Still, in a variety of languages like French, Italian, German, Japanese, British English (Fernald, Taeschner, Dunne, Papousek, de Boysson-Bardies & Fukui, 1989) and American English (Warren & Bohannon, 1984; Fernald, et al., 1989), it was observed that mothers and fathers

exhibited differences in pitch and pitch range, with mothers producing higher pitch and a wider pitch range in comparison to fathers. Also, some research suggests that fathers use more advanced vocabulary, introduce more wh-questions, produce more directives and imperatives, cause more communicative breakdowns and generally elicit more speech from the child (Bernstein-Ratner, 1988; Tenenbaum & Leaper, 1998; Tomasello, Conti-Ramsden, & Ewert, 1990; Leaper, Anderson & and Sanders, 1998; Jin & Naka, 2002, and for a more detailed overview see Abkarian, Dworkin, & Abkarian, 2003).

The observed differences between parents can be explained by two similar hypothesises: The Bridge Hypothesis (Gleason, 1975; Barton & Tomasello, 1994) and The Differential Experience Hypothesis (McLaughlin, White, McDevitt & Raskin, 1983) which state that fathers speak more simply to children than to adults, but not as simply as mothers do, mainly due to the fact that fathers are less sensitive to the child's linguistic abilities. As a consequence, fathers are putting more demands on the child which improves children's performance and creates a bridge for children to communicate with strangers.

On the other hand, recent studies on the difference between mothers' and fathers' CDS which included socio-economic status, time spent with the child, and education of parents as possible confounding variables which can effect CDS production, directly questioned the Bridge and Differential Experience hypotheses. Specifically, the comparisons between upper-middle and working class families showed that working class parents tended to produce the same or moderately different CDS in contrast to upper-middle class parents which exhibited more pronounced differences between parents (Matthews, Ichile, Newman & Berstein-Ratner, 2004; Rowe, Coker & Pan, 2004). In addition to these studies, Davidson and Snow (1996)

showed that in families where both parents were highly educated (most of the participants had Master's or Doctoral degrees), mothers used more difficult vocabulary items, more questions and more complex speech in general.

The effects described above indicate that the roles of the primary and secondary caregiver in facilitating first language learning are highly dependent on socio-economic and cultural factors. As a result, it would be interesting to see in further research how CDS varies in diverse family configurations, what parts of the register remains constant, and most importantly what effects this has on the first language acquisition.

Taken together, the short overview of the main characteristics of CDS showed that the input presented to children in a form of this register contains a great deal of the potential cues to the grammatical and semantic features of the system that is acquired. The following part of the Introduction will present the summary of previous experimental and neural network studies exploring whether first and second language learners really benefit from the facilitating features of CDS register.

1.3. The facilitating effect of child-directed speech on language acquisition

In addition to the corpus based studies on the general characteristics of CDS, there is a growing number of research studies addressing the question as to what extent children are sensitive to the specific features of the input, and whether some of the features of CDS tend to facilitate language learning.

A number of experiments have shown that infants and children discriminate and prefer CDS from the earliest stages of their life. For example, Cooper & Aslin (1990) demonstrated that newborns and one-month old infants preferred infant-directed over adult-directed speech. A similar effect was observed for slightly older (four to ten months) children (Fernald, 1985; Werker & McLeod, 1989; Kemler-Nelson, Hirsh-Pasek, Jusczyk, Cassidy, 1989; Pegg, Werker & McLeod, 1992; Werker, Pegg & Mcleod, 1994; Kaplan, Goldstein, Huckeby & Cooper, 1995; Cooper, Abraham, Berman & Staska, 1997; Hayashi, Tamekawa & Kiritani, 2001). Moreover, recent studies showed that the children's preference towards CDS may facilitate language learning. The following parts of this section will provide an overview of the studies using corpus analyses, experimentation and connectionist modelling to examine the effect of CDS in various domains of language learning.

The first task which prelinguistic children must accomplish is to find a way to isolate meaningful chunks from the continuous streams of speech that they hear. Gerken (1996) pointed out that prosodic cues from CDS can provide reliable cues for successful word segmentation. Moreover, an extensive longitudinal corpus study of 26 mother-child dyads showed that mothers' tendency to segment words clearly at the children's 10-word stage resulted in fewer unanalysed phrases at the 50-word stage (Pine, Lieven & Rowland, 1997). In a set of neural network simulations, Brent &

Cartwright (1996) and Christiansen, Allen & Seidenberg (1998) showed that models performed better on the task of word segmentation if they were presented with information on distributional regularities, phonotactic constraints, relative lexical stress and boundaries between utterances obtained from a corpus of CDS as compared to a corpus of ADS. Kempe, Brooks & Gillis (2005) and Kempe, Brooks, Gillis & Samson (in press) argued that diminutives, one of the most frequent derivations in CDS, increase invariance of word endings in the input, and, thus, may serve as a word segmentation cue. In a set of experimental studies on second language learning of Dutch and Russian nouns, they demonstrated that the adult English speaking participant performed better in discriminating Russian and Dutch nouns in uninterrupted speech when those units contained diminutive suffixes at the end.

Also, in an experimental study on children's sensitivity to prosodic cues for word segmentation in CDS, Thiessen, Hill & Saffran (2005) showed that 7-8 months old infants were able to segment words in a nonsense speech stream spoken with CDS, but not with ADS intonation.

At the level of morpho-syntax, a number of studies for English demonstrated that CDS provide a number of phonological/prosodic cues (e.g. length of phonemes, length of syllables, presence of stress, phonological complexity of words, etc.) and distributional cues which tend to facilitate the process of grammatical categorisation (Fisher & Tokura, 1996; Hoff-Ginsberg, 1985; Kelly, 1996; Mintz, Newport & Bever, 2002; Mintz, 2003; Morgan, 1996; Morgan & Demuth, 1996; Monaghan, Chater, & Christiansen, 2005; Redington, Chater & Finch, 1998). Most recently, cross-linguistic studies indicated that phonological and distributional cues might also facilitate the categorisation of words onto nouns, verbs, adjectives, adverbs, etc. in

languages like German, Dutch, French and Japanese (Keibel & Elman, 2004; Monaghan, Christiansen, Chater, submitted).

The relationship between phonological and distributional cues in CDS and children's grammatical categorisations was confirmed in neural network simulations which showed better performance when presented with data from CDS (Cartwright & Brent, 1997; Freudenthal, Pine & Gobet, 2001).

Similar effects were observed in experimental studies which showed that infants are using acoustic cues like pitch and durational changes and distributional information from CDS for the discrimination of noun and verb phrasal units or, more specifically, discrimination of just nouns from the verbs (Jusczyk, Hirsh-Pasek, Kemler-Nelson, Kennedy, Woodward & Piwoz, 1992; Fitneva, Tobiah, Christiansen & Monaghan, 2005).

Finally, vocabulary acquisition also relies on the semantic and syntactic structure of CDS. Corpus based studies showed that there is a high correlation between socio-economic status (SES) of parents and the size of children's lexicon. Children from mid-SES families were exposed to a poorer vocabulary in comparison to children from high-SES families. This difference was reflected in children's overall lexical production (Hoff-Ginsberg, 1998). Also, for children with a relatively poor lexicon, it was shown that maternal language at 1;1 years predicts children's MLU at 1;8 years which suggest that the observed individual differences in early language acquisition might be attributed partially to the lexical richness of CDS (Hampson & Nelson, 1993). In addition, experimental studies with second language learners demonstrated that English-speaking adults acquired Chinese words better if the targeted/novel words were presented at the end of the sentences (a CDS-like situation)

in contrast to conditions when the targeted words were presented in the middle part of the sentence (an ADS like situation) (Golinkoff & Alioto, 1995).

Overall, the empirical and computational evidence underscores a strong relationship between specific features of CDS and children's language production in the first years of their lives. Nevertheless, the short overview presented above showed that most of the cross-linguistic research was mainly focused on the relation between CDS and the acquisition of phonetic/phonological features, word-segmentation or basic morpho-syntactic categorisations, i.e. part-of-speech categorisation. With the respect to the acquisition of other language domains, like morphology, complex syntactic functions or word meanings, studies were mainly carried out in English or with artificial grammars. This indicates that there is still a great need for additional cross-linguistic research on the universal nature of the facilitating effect of CDS on first language acquisition. For example, in contrast to English which is a relatively morphologically impoverished language, there is a large number of languages like the Balto-Slavic group (Russian, Polish, Serbian, Lithuanian, etc.), German, Turkish, Finnish etc. which exhibit very complex morphological systems. First attempts at cross-linguistic comparisons of morphology acquisition were conducted during seventies and eighties (Ferguson & Slobin, 1973; Slobin, 1985; Johnston & Slobin, 1979), and mainly presented the observational research of children's morphology comprehension and production. These studies were based on a limited source of data on possible facilitating effect of CDS on the acquisition of morphology. However, recent experimental studies on the acquisition of complex noun morphology in Russian, Lithuanian and Polish (Kempe & Brooks, 2001; Kempe, Brooks, Mironova & Fedorova, 2003; Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press; Ševa, Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press; Dabrowska,

2006; Savickienė, Kempe & Brooks, in preparation) have shown that the parental usage of diminutives (the terms for expression of affection and endearment) potentially benefit the acquisition of noun morphology.

Next, I will provide a description of the qualitative and quantitative features of diminutive usage in CDS registers across different languages, together with an overview of experimental studies on the facilitating effect of diminutives on the acquisition of complex noun morphology in Russian, Polish and Lithuanian.

1.4. Diminutives in child-directed speech and their facilitative effect on learning of complex noun inflectional morphology

1.4.1. General description of the distribution of diminutives in child-directed speech

Diminutives are a derivation which is used almost universally across languages for the expression of smallness, affection and endearment (Jurafsky, 1996). In most languages, diminutives are derived by simple attachment of diminutive suffixes to the simplex form of a noun (see Table 1.2.).

Table 1.2. Examples of derivation of diminutives across different languages

Simplex form of noun	**Diminutive form of noun**	**Language**
dog	dogg-*y*	English
raam	raam-*pje*	Dutch-window
žiraf	žiraf-*ik*	Russian-giraffe
estrella	estrell-*ita*	Spanish-star
die Mütze	das Mütz-*chen*	German-cap
il poeta	il poet-*ino*	Italian-poet

However, despite the universality of the semantic and pragmatic features of diminutives, languages differ in the level of productivity of this derivation, from systems like English, where the attachment of the diminutive *–y* suffix is limited to a small number of nouns like *doggy, bootie, Patty,* but not **liony*, **tably*, etc., to languages like Russian, Lithuanian, Serbian, Italian, Spanish, where diminutives can be derived not only from most concrete and some abstract nouns, but also from

adjectives, adverbs, verbs, etc. Bauer (1997, cited in Melzi & King, 2003) suggested the following universal hierarchical distribution of diminutives over different word-types: (a) Noun, (b) Adjective/Verb, (c) Adverb/Numeral/Pronoun/Interjection, (d) Determiner, indicating that "for a word-class to be used as the base in evaluative morphology in a particular language, word-classes from each step above that word-class must also be so used in that language" (Bauer, 1997: 540, cited in Melzi & King, 2003).

Despite cross-linguistic differences in the potential productivity of this derivation, diminutives are defined as one of the universal features of CDS. Parents start introducing diminuitivised words from the first months of their children's lives and continue using them frequently until children are four or five years old (Ferguson & Snow, 1977; Ferguson, 1978). Moreover, the first corpus analysis of child noun production in every day interactions with adults showed that diminutives are one of the first derivations acquired by two to three year old children, and that their use highly correlates with parental diminutive production. (Gleason, Perlmann, Ely, & Evans, 1994; Dressler, 1997; Gillis, 1998; Kempe, Brooks & Pirrot, 2001; Melzi & King, 2003). Still, the experimental studies which tested children's comprehension and production of the diminutive derivation of novel nouns have shown that the full acquisition of the structural, semantic and pragmatic features of diminutives continues until ten to twelve years of age (Gleason, 1958; Snow, Smith & Hoefnagel-Hohle, 1980; Herrera & Carvallo, 1987, cited in Melzi & King, 2003), suggesting that the first usage of diminutives is mainly a product of children's imitation of nouns from the input. Melzi & King (2003) also observed a significant number of parental imitations of children's diminutive usage and concluded that:

…"the strong relationship between diminutive imitation and overall use highlights the sensitivity of both interlocutors to each other's speech. As the mother picks up her child's diminutives and imitates them back to him or her, she appears to strengthen the child's use of diminutivised forms and thereby to facilitate greater productive use of diminutives".

In addition, the parental imitation of children's diminutives may be used as a means to increase the overall attachment between interlocutors, and to establish continuity and stability of communication flow which, in turn, can facilitate children's overall language production (King & Melzi, 2004).

Quantitative analyses of the frequency of diminutives in CDS were obtained for the following languages: English (Gleason et al., 1994), German (Kempe et al. 2001; Korecky-Kroell & Dressler, 2004), Russian (Voeykova, 1997; Kempe et al, 2001), Polish (Hamman, 2003; Dabrowska, 2006), Lithuanian (Savickienė, 1998), Italian (De Marco, 1998), Dutch (Gillis, 1997), three dialects of Spanish (Kempe et al., 2001; Melzi & King, 2003, 2004; Herrera & Carvallo, 1987, cited in Melzi & King, 2003) and Greek (Stephany, 1997). In these studies, the frequency of diminutives was described by various measurements like the percentage of diminutives out of total number of noun tokens[3] (Russian, Greek, Polish, Lithuanian, German, Dutch and Mexican Spanish), the percentage of diminutives out of total number of noun types (Russian, Mexican Spanish, German, Dutch, Italian), the percentage of diminutives out of total number of words (Venezuelan and Peruvian Spanish), the percentage of diminutives out of total number of all diminutivisible

[3] The term noun tokens is used for labelling occurrences of all word-forms, and the term noun types is used for labelling different word-forms.

words (Peruvian Spanish)[4], and an absolute number of diminutives over 100 sentences (English)[5]. In addition, the corpora also differ with respect to: a) the children's age at which conversations were recoded, covering a large age span from 13 months to 11 years; b) the number of mothers included in the studies, with the biggest samples for English (88 mothers) and Peruvian Spanish (32 mothers) to Lithuanian, Dutch or Italian with only one parental-child dyad and c) the different number of age samples, with only two age samples for English and Peruvian Spanish to ten or more very dense samples for Lithuanian and Spanish. Given these big methodological discrepancies in the recoding and coding of data, the cross-linguistic comparisons in this chapter are limited only to the languages which provided the same type of measurement of diminutive production, and which covered similar age groups. Also, I will only provide descriptive comparisons between languages because statistical tests of diminutive production are precluded by the fact that estimates of diminutive frequency were based on samples differing in number of words and number of participants.

As a result of these constraints, descriptive comparisons were possible only for data from the studies on CDS in Russian, Greek, Polish, Lithuanian, German, Dutch,

[4] In the case of Spanish, diminutivisible words are nouns, adjectives and adverbs. For Venezuelan Spanish, analyses showed that parents used 5% of diminutives out of all words when children were seven to ten year old, and for Pervuian Spanish 7% of diminutive out of all diminutivisible words and 2% of diminutives out of all words, when children were three and five year old. Preliminary calculations based on data on distribution of different part-of speech categories in Spanish (Farwell, Helmreich & Casper, 1995), indicate that for Peruvian Spanish, mothers produced not more than 8-10% of diminutives out of total number of nouns in contrast to Mexican mother who produced 45% of diminutives. This indicates that diminutive distribution may depend not only on socio-cultural factors, but also on small dialectological differences. Also, samples for Spanish differ in size (number of mothers) and age of children. For example, for Mexican Spanish the estimates were based only on one mother, producing CDS when child was two to three year old. On the other hand the sample for Peruvian Spanish covered larger population with 32 mother-child dyads, and larger age span, from three to five year old. This indicates that the observed percentage of diminutives is not only sensitive on the age of children, but also on the size of the sample used for the frequency estimates.

[5] Analyses for English showed that parents produced on average 1.9 diminutives per 100 sentences.

Italian and Mexican Spanish for which the diminutive distribution was provided as a percentage out of all noun tokens and types. The data for the frequency of diminutives in those languages are presented in Table 1.3.[6]

Table 1.3. Cross-linguistic comparison of the frequency of diminutives in CDS.

Language	**Percentage of diminutives out of all noun tokens**	**Percentage of diminutives out of all noun types**
German	3-7	6
Greek	30-40	30-40
Mexican Spanish	42	40
Russian	45	40
Polish	20-30	/
Dutch	20-30	/
Lithuanian	30-40	/
Italian	/	10-20

This quantitative comparison of diminutive usage in CDS confirms that diminutives are one of the main features of this register, with an average of one third of nouns produced with this derivation. The only exception was observed for German with only 3-7% of diminutives out of all noun tokens and 6% of diminutives out of all noun types. Given this, German CDS is closer to English CDS, with 1.9 diminutives over 100 sentences, despite the fact that German belongs to the group of languages with potentially high diminutive productivity. Kempe et al. (2001) hypothesised that the observed difference is not only due to socio-cultural factors, but also due to cross-linguistic differences in the derivation of diminutives between German and other languages: in many languages, diminutives not only increase the salience of the word endings but also regularise stress patterns, gender marking and case marking. In

[6] Sign / was used for marking fields were data was not present in studies.

contrast, German diminutives neutralise all nouns with the suffix *–chen*, *–lein*, *–l* and *–le*, rendering noun gender and case marking opaque. Thus, the high frequency of diminutives in CDS which is primarily a product of affective communication between parents and children, might also be influenced by the potential facilitating effect of this derivation on word segmentation (Kempe et al., 2001, 2005; Kempe, Brooks, Gillis & Samson, in press) and acquisition of noun morphology. In order to test whether a high frequency of diminutives in CDS really facilitates noun morphology learning, Kempe & Brooks, 2001; Kempe et al, 2003; Kempe, Brooks, Mironova, Pershukova & Fedorova, in press; Ševa et al., in press; Dabrowska, 2006; Savickienė et al., in preparation, conducted a series of experimental studies with first and second language learners of Russian, as well as with Polish and Lithuanian children. The last part of this section will present these studies and discuss the observed results in a broader context of the research on the acquisition of morphology.

1.4.2. The facilitating effect of diminutives on the acquisition of inflectional noun morphology

The acquisition of morphology in general is traditionally described as a U-shaped process which passes through three stages (Ervin, 1964; Bever, 1982; Bowerman, 1982, cited in Redington & Chater, 1998). Stage one is characterised by initial rote-learning, where children produce both regular and irregular forms correctly. Stage two is characterised by the identification of morphological structure and "over-regularisation" of previously correctly produced irregular forms (children start producing *goed* instead of *went*). Finally, during stage three, children start to produce regular and irregular forms accurately, and correctly apply morphological markers to novel items. As a result, during the process of morphology acquisition, children are confronted with two problems (Redington & Chater, 1998; Tomasello, 2003):

The first problem is related to the identification of relevant morphological units which are usually expressed in phonologically reduced, unstressed, monosyllabic bits at the beginnings or ends of the words. The second problem is associated with the mapping of those units onto appropriate grammatical functions and meanings. For example, the child has to learn that adding *-ed* means that the verb takes the past tense. This process is usually aggravated by the fact that in some cases these grammatical morphemes are polyfunctional, like the English suffix *–s which* is used for marking plural, third person simple present tense or possessive gentive.

Languages tend to differ in the complexity of the inflectional noun morphology. In languages with rich inflectional morphology, learners are faced with

the task of extracting the suffixes for marking categories like gender[7], case[8] and number[9]. For example, there are highly complex languages like Finnish with 16 different cases or Lithuanian with 7 cases distributed over 12 declension paradigms. In addition, in many languages, pronominal words (adjectives, pronouns and some numbers) can also be inflected and they have to agree in gender, case and number with the noun to which they are related (a detailed description of noun morphology will be presented in Chapter 2).

Early cross-linguistic studies showed that the complexity and transparency of morpho-phonological features at the ends of words affected the learning trajectories in languages with complex morphological systems (Johnston & Slobin, 1979; Slobin & Bever, 1981). Despite the observed cross-linguistic differences in learning rates, it has generally been assumed that complex morphological systems are fully mastered by children between two and four years of age. So far we know that parents tend to facilitate the acquisition of these grammatical categories in two ways. One way is through immediate recasts of the child's words that were missing in the form of reformulations, expansions, topic continuations, or replies. Farrar (1990, 1992) has shown that two year olds were two to three times more likely to imitate the correct grammatical morpheme after corrective recasts than after any other form of positive

[7] Most languages distinguish between two or more genders. The assignment of nouns to one of the gender categories is usually based on semantic/pragmatic and formal factors (word-structure, inflection or sound-structure). The classification of animate nouns very often corresponds to the real world distinction of sex. Thus, nouns referring to females are generally of feminine gender. On the other hand, in languages without clear formal cues (e.g. German), the assignment of gender category to inanimate and abstract nouns very often looks like a matter of arbitrary stipulation (Corbett, 1991). Probably one of the most famous examples of categorisation based on culturally different semantic and pragmatic factors is found in Dyirbal, an Australian language, spoken in north-east Queensland. Dyirbal differentiates four genders: one for men, most of the animals, moon, etc., second for women, fire and dangerous things, third for non-flesh food and fourth for the residue (Lakoff, 1987, Corbett, 1991).

[8] Cases represent morphologically different forms of the same noun which are used to mark different syntactic functions and meanings, like subject, object etc. For example, in Latin, the word *porta* 'doorNOM' in genitive case gets the suffix *–e*, *portae* 'of the doorGEN'. A paradigm of different inflected forms of the same noun is called declension.

[9] Most languages distinguish between singular and plural, and in some languages, dual – a separate number comprising two to four items.

evidence which indicates that mothers' recasts are helping children to identify elements which are low in salience. The other way to facilitate the acquisition of noun morphology is to introduce words like diminutives in CDS which increase the salience of the word endings and in some languages, additionally simplify gender and case marking. In a set of experimental studies, it was shown that diminutives tend to facilitate the acquisition of complex systems of noun morphology, in languages like Russian, Polish or Lithuanian.

The first study (Kempe & Brooks, 2001) was conducted with second language learners of Russian. Adult English speakers were presented with a set of Russian nouns. Half of the participants were presented with nouns in diminutive form, together with a colour adjective (e.g. *krasnij domik* 'red house (DIM MASC)', *krasnaja kozochka* 'red goat (DIM FEM)') and the other half were presented with simplex forms of the same nouns, presented in the same context with the colour adjectives (e.g., *krasnij dom* 'red house (SIM MASC)', *krasnaja koza* 'red goat (SIM FEM)'). After four sessions of exposure, both groups were tested, on a gender generalisation task that required them to produce colour adjective-noun phrases, similar to the ones they were exposed to during training, for a variety of familiar as well as novel Russian nouns. Responses in which adjectives and nouns agreed in gender were treated as correct gender categorisations of the presented nouns. The main result of this study was that the adults in the diminutive-exposure group produced significantly fewer adjective-noun gender-agreement errors on both simplex and diminutive novel nouns in comparison to adults who were exposed only to simplex nouns.

The second study (Kempe et al., 2003) tested gender categorisation with 2- to 4- year old Russian children. In a similar task to the one used in Kempe & Brooks,

2001, children were presented with a set of pictures of familiar and unfamiliar animals, with unfamiliar animals labelled with novel nouns. Half of the nouns were presented in the diminutive and the other half in the simplex form. The first occurrences of gender agreement (with adjectives or pronouns) were coded and indicated that children produced significantly less gender-agreement errors with diminutive than with simplex forms of the nouns. The results from this study were replicated in Ševa et al. (in press) with another group of two to four year old Russian children and also supported in a cross-linguistic comparison with Lithuanian, a richly inflected Baltic language (Savickienė et al., in preparation). In this study, two groups of Lithuanian children (three and six year old) were tested for gender agreement between nouns and pronominal words (adjectives and pronouns) using an elicited production task, similar in structure to the one used in the studies for Russian children. The main result from all these studies indicated that children produced fewer gender-agreement errors for novel nouns introduced in diminutive form.

Furthermore, Kempe, Brooks, Mironova, Pershukova & Fedorova, in press have shown a facilitating effect of diminutives on the acquisition of case marking in Russian. In this study, two to four year old Russian children were presented with a set of unfamiliar objects labelled with novel nouns. The novel nouns were introduced twice to a child before prompting the appropriate answers for the dative/genitive questions. The nouns were presented in three conditions. In simplex-simplex condition, each novel noun was introduced twice in simplex form. In the diminutive-diminutive condition, each noun was introduced twice in diminutive form. Finally, in the word-play condition, each noun was introduced once as simplex and once as diminutive, with order of presentation alternating. Dative and genitive constructions were elicited with a little toy elephant which was moving towards and from the

objects. The results revealed that Russian children made fewer errors when the novel words were presented in the wordplay and diminutive-diminutive condition compared to the simplex-simplex condition. The facilitating effect of diminutives on case marking was replicated in one more study with Russian children (Kempe et al., in preparation), but has also been demonstrated for Polish, another Slavic language with rich noun morphology. In a case-marking task, where children produced genitive, dative and accusative constructions, Dabrowska (2006) showed that two to four year old Polish children committed fewer case-marking errors with novel diminutive masculine and feminine nouns compared to novel simplex nouns.

Taken together, these studies demonstrate a strong diminutive advantage in the acquisition of inflectional morphology by children learning these three Balto-Slavic languages, as well as adults exposed to Russian as a foreign language. How do these findings fit with current accounts of the acquisition and processing of inflectional morphology? Traditionally, structuralist and nativist accounts on the production and comprehension of morphology by children and adults were mainly based on studies of English verb morphology (and in later stages on German noun plural forms). It was assumed that our cognitive system operates using a dual-route mechanism, containing two separate architectural components (Pinker, 1991; Prasada & Pinker, 1993; Marcus, Brinkmann, Clahsen, Wiese & Pinker, 1995). The first part of the system deals with regular forms (e.g. walk/walked) which are produced by a set of abstract symbolic rules, applied to the stem of the words. The second part is based on the lexical associative memory which is used for irregular forms (e.g. sing/sang), and it is sensitive to the words' type and token frequency, phonological structure and semantics. However, the last 20 years of research on acquisition and processing of inflectional morphology saw an increase in studies which assumed that this process

relies on a single-route mechanism. A series of item-based/connectionist models proposed that children move from learning morphological patterns for single words, to learning morphological patterns applying to narrow clusters of fairly similar words, and eventually to wider generalisations, encompassing groups of words commonly labelled as grammatical categories. It was hypothesised that this gradual process of categorisation of the morphological system is based on an associative learning mechanism which connects both regular and irregular nouns into a set of clusters/low-level schemata, taking into account a words' type and token frequency, semantics, and phonological structure. These accounts also presupposed that narrow (phonologically based) and wide generalisations, once acquired, co-exist in the adult system. Thus, while the system seems to favour low-level generalisations characterised by morpho-phonological homogeneity at a certain stage of learning, later generalisations (e.g. in Russian and Polish, the generalisation of feminine agreement and case marking to all nouns ending in *–a*) need not completely override earlier generalisations (Tomasello, 1992, 2003; Bybee, 1995; Lieven, Pine & Baldwin, 1997; Rumelhart & McClelland, 1986; Plunket & Marchman, 1991, 1993; MacWhinney & Leinbach, 1991; Elman et al., 1996; Plunkett & Nakisa, 1997; Kempe & MacWhinney, 1998; Joanisse & Seidenberg, 1999; Hahn & Nakisa, 2000; Ramscar, 2002).

As a result of these architectural differences, the two models (single vs. dual-route) make different predictions for the generalisations of novel nouns. The dual-route model states that grammatical categorisation of novel items will be based on default rules and will proceed separately for regulars and irregulars. On the other hand, single-route accounts assume that the generalisations of novel items are based on phonological and distributional regularities within the system. Consequently, the

outcomes for at least some novel words should be different, with better performance for the items which are closer to the phonological space of the familiar nouns.

Thus, the diminutive advantage observed in the experiments with Russian, Polish or Lithuanian three to six year old children performing gender agreement and case marking tasks complements the single-route accounts which stated that processing and acquisition of complex inflectional systems will rely on phonological similarities and high frequency of the clusters of nouns. The diminutive advantage in Russian, Polish and Lithuanian may be the result of several factors. First, the use of diminutives in CDS might regularise noun morphology within the observed systems. For example, in languages like Russian, in addition to the dominant transparent and regular class of nouns, there is a small subset of non-transparently gender-marked nouns. The diminutivisation of such nouns results in a form which is transparently gender marked which means that the frequent diminutivisation minimises the instances of non-transparently gender-marked nouns, thereby increasing the overall degree of gender-marking regularity in the input. Gender learning is easier if the input contains less non-transparently gender-marked nouns (Kempe & Brooks, 2001). Similarly, Lithuanian diminutives reduce the complexity of the system of noun declensions by decreasing the number of declension types from twelve to three.

Second, adding a diminutive morpheme to a noun inserts a phonologically invariant segment right before the inflectional suffix at the end of the word. It is possible that this 'island of invariance' may serve to mark and highlight the upcoming inflectional changes thereby drawing the learners' attention to morpho-phonological information such as the association between noun ending and noun gender or case.

Third, diminutive morphemes increase the phonological similarity within genders. For example, while Russian and Polish masculine simplex nouns can end in

any consonant, Russian and Polish masculine diminutives all end in *-k*. This renders masculine diminutive nouns much more similar to each other than masculine simplex nouns. The same is true for feminine and neuter nouns. Increased phonological similarity should make it easier to discover grammatical gender categories.

Fourth, diminutive morphemes result in a substantial degree of phonological similarity amongst the class of diminutives in general. Since these morphemes can sometimes encompass up to three syllables of a noun (e.g. Russian: *ruchonochka* 'handDIM-DIM-DIM'), diminutives constitute a noun cluster with high morpho-phonological similarity, thereby facilitating the extraction of gender categories for that particular cluster of words.

Finally, high type frequency of diminutives in CDS may also play a role in the facilitating effect of this derivation. Standard usage based accounts state that novel words are categorised more correctly if they fall into highly dense clusters of phonologically similar words. The role of high token frequency of diminutives observed in the corpus data was not discussed in any of the previously described studies, although there are some indications in the literature that high token frequency of one item (or even a category) might facilitate language learning by rendering the entire system more redundant (Casenhiser & Goldberg, 2005). Also, both high type and token frequency increase the chances for diminutive nouns to appear next to adjectives or pronouns. Given that in the Balto-Slavic languages pronominal words always take the same gender/case suffixes as the nouns they are referring to, this may provide additional distributional cues for noun categorisation.

In sum, this overview of possible factors responsible for the diminutive advantage observed in Russian, Polish and Lithuanian shows that children might rely both on phonological and distributional cues for gender and case categorisation. This

notion is in line with the Phonological-Distributional Coherence Hypothesis which states that language learning benefits from the integration of phonological and distributional information about a category provided in the input (Monaghan, Christiansen & Charter, submitted).

In the next chapter, I will provide a brief description of Serbian noun morphology to illustrate the complexities that need to be acquired by the language learner. This overview will be followed by the cross-linguistic corpus and experimental exploration of facilitating effect of diminutives on the acquisition of noun morphology to Serbian, a south Slavic language with rich inflectional noun morphology.

Chapter 2

2. Description of Serbian noun morphology

Serbian is a south Slavic language, with richly inflected nouns, verbs, adjectives and pronouns, and relatively free word order. In many of the world's languages, case/gender marking inflections are used to distinguish the roles that nouns play in sentence interpretation. In contrast to English which relies heavily on word order to convey who did what to whom (with case marking limited to a few pronominal contrasts such as *I* versus *me*), languages in the Balto-Slavic family, such as Serbian, Russian, Polish and Lithuanian, use very rich systems of inflectional suffixes to differentiate the functional roles of nominals. These fusional languages also differ from agglutinative languages, such as Turkish, Finnish or Hungarian, with respect to the complexity of their morpho-syntactic paradigms. In fusional languages, inflections typically mark combinations of grammatical features (e.g., noun gender, case, number and animacy) and may display considerable syncretism, where a suffix is used to mark several different cases across several different gender paradigms. For example, in Serbian, the suffix *-a* is used to mark several different gender-number-(animacy)-case combinations as in (1-3):

(1) *vod-a*: 'water-feminine' +singular+nominative or +plural+genitive

(2) *konj-a*: 'horse-masculine' +singular/plural+genitive or singular+animate+accusative

(3) *sel-a*: 'village-neuter' +singular/plural+genitive or +plural+nominative/accusative

Agglutinative languages, in contrast, may contain a considerable number of distinct affixes, each associated with a single function or meaning.

2.1. The general structure of the Serbian noun system

Traditionally, grammarians (Stevanović, 1964; Stanojčić & Popović, 2003) distinguish the following morphologically marked categories in the Serbian noun system:

a) gender (masculine, feminine, neuter);

b) case (nominative, genitive, dative, accusative, vocative, instrumental, locative);

c) number (singular and plural)

d) animacy (only with masculine nouns).

The three gender categories are highly transparent in the nominative case, such that most masculine nouns end in *–∅* (i.e., a consonant), feminine nouns in *–a*, and neuter nouns in *–o* or *–e*. In addition to these 'transparent' nouns, there are several clusters of 'non-transparent' masculine and feminine nouns that constitute approximately 10% of all Serbian nouns. These comprise masculine nouns ending in *–a* which are mostly kinship terms, some occupations, male proper names, and nicknames of animals (e.g., *meda* 'teddy-bear' versus *medved* 'bear') and feminine nouns ending in consonants, with many of these nouns referring to abstract concepts (e.g., *ljubav* 'love').

Most of the nouns (common and proper) can be declined through singular and plural, with the exception of a small group of *sigularia tantum* (nouns which only appear in singular form) – e.g., *hrabrost* 'courage' and *pluralia tantum* (nouns which only appear in plural form) – e.g., *vrata* 'door' or *makaze* 'scissors'.

Nouns are declined through four declension groups organized around the three transparently-marked gender categories plus a fourth declension paradigm for non-

transparent feminine nouns (i.e., feminine nouns ending in consonants). Non-transparent masculine nouns (i.e., ending in *–a*) follow the regular feminine declension, but require masculine adjective-noun and pronominal agreement. The Serbian case marking system exhibits a considerable degree of inflectional syncretism. For the 14 possible conditions defined by the 7 cases and 2 numbers (singular/plural), there are only 8 different suffixes for the masculine declension, 7 for the transparent feminine declension, 5 for the neuter declension, and 4 for the non-transparent feminine declension. For the 56 conditions defined by the 7 cases x 2 numbers x 4 declensions, there are only 9 distinct suffixes in total. In addition to the four main paradigms, masculine nouns differentiate into two subclasses, organized around the semantic criterion of animacy. The difference between animate and inanimate nouns is marked in the accusative singular case, where animate nouns take the suffix *–a*, in contrast to masculine inanimate nouns for which the accusative form is the same as in the nominative case (animate: *konj* 'horseNOM' vs. *konj-a* 'horseACC' and inanimate: *prozor* 'windowNOM' vs. *prozor* 'windowACC').

Within each paradigm, some of the suffixes have allomorphic variants. For example, the masculine instrumental distinguishes two forms: *–om* and *–em*, depending on the phonological structure of masculine noun stems (masculine nouns ending in non-palatal consonant take the suffix *–om*: *učenik* 'studentNOM' vs. *učenik-om* 'studentINS' in contrast to masculine nouns with stems that end in palatal consonants: *učitelj* 'teacherNOM' vs. *učitelj-em* 'teacherINS').

An overview of the four main paradigms in the Serbian noun system, together with the distribution of suffixes over cases is presented in Table 2.1 and Table 2.2.

Table 2.1. Declension classes in Serbian language

Gender	Case	Singular	Plural
Masculine Animate	Nominative	konj[10]	konji
	Genitive	konja	konja
	Dative	konju	konjima
	Accusative	konja	konje
	Vocative	konju	konji
	Instrumental	konjom	konjima
	Locative	konju	konjima
Masculine Inanimate	Nominative	prozor[10]	prozori
	Genitive	prozora	prozora
	Dative	prozoru	prozorima
	Accusative	prozor	prozore
	Vocative	prozoru	prozori
	Instrumental	prozorom	prozorima
	Locative	prozoru	prozorima
Feminine	Nominative	voda[10]	vode
	Genitive	vode	voda
	Dative	vodi	vodama
	Accusative	vodu	vode
	Vocative	vodo	vode
	Instrumental	vodom	vodama
	Locative	vodi	vodama
Neuter	Nominative	selo[10]	sela
	Genitive	sela	sela
	Dative	selu	selima
	Accusative	selo	sela
	Vocative	selo	sela
	Instrumental	selom	selima
	Locative	selu	selima
Feminine Opaque	Nominative	ljubav[10]	ljubavi
	Genitive	ljubavi	ljubavi
	Dative	ljubavi	ljubavima
	Accusative	ljubav	ljubavi
	Vocative	ljubavi	ljubavi
	Instrumental	ljubavlju	ljubavima
	Locative	ljubavi	ljubavima

[10] *konj* [horse], *prozor* [window], *voda* [water], *selo* [village], *ljubav* [love].

Table 2.2. Distribution of suffixes and their allomorphs across cases within masculine animate/inanimate, feminine, neuter and feminine opaque paradigms.

Gender	Suffix	Case
Masculine Animate	-Ø (-e, -o)	nom. sg. + voc. sg.
	-a	gen. sg. + acc. sg. + gen. pl.
	-u	dat. sg. + loc. sg.+ voc. sg.
	-om (-em)	ins. sg.
	-i	nom. pl. + voc. pl.
	-e	acc. pl.
	-ima	dat. pl. + loc. pl.+ ins. pl.
Masculine Inanimate	-Ø (-e, -o)	nom. sg. + acc. sg. + voc. sg.
	-a	gen. sg. + gen. pl.
	-u	dat. sg. + loc. sg.
	-om (-em)	ins. sg.
	-i	nom. pl. + voc. sg.
	-e	acc.pl.
	-ima	dat. pl. + loc. pl.+ ins. pl.
Feminine	-a	nom. sg. + gen. pl. (+ voc.sg.)
	-e	gen. sg. + nom. pl. + acc. pl. + voc. pl.
	-i	dat. sg. + loc. sg. (+ gen. pl.)
	-u	acc. sg.
	-o	voc. sg.
	-om	ins. sg.
	-ama	dat. pl. + loc. pl.+ ins. pl.
Neuter	-o (-e)	nom. sg. + acc. sg. + voc. sg.
	-a	gen. sg. + nom. pl. + gen. pl. + acc. pl. + voc. pl.
	-u	dat. sg. + loc. sg.
	-om	ins. sg.
	-ima	dat. pl. + loc. pl.+ ins. pl.
Feminine Opaque	-Ø	nom. sg. + acc. sg.
	-i	gen. sg. + dat.sg. + loc.sg.+ nom.pl. + gen.pl. + acc.pl. + voc. sg. + voc. pl.
	-ju (-i)	ins. sg.
	-ima	dat. pl. + loc. pl.+ ins. pl.

2.2. Cases and genders within sentences

The general function of cases is to express various syntactic functions and meanings carried by nouns within broader units like phrases or sentences. Until today grammarians have not arrived at a full consensus about what distinguishes functions from meanings, nor have they completed taxonomy of those classes (Blake, 2001). This is maybe best illustrated in a list of functions and meanings for Serbian nouns presented in Appendix 1. The list was compiled from six standard Serbian grammar books in the early sixties (Kostić, 1965) and it contains some unusual solutions for the depiction of syntactic functions and meanings in Serbian. However, despite the differences in description and classification of functions and meanings, most of grammarians agree on the typical functions and meanings which are used for the general differentiation of cases. Table 2.3 represents typical functions and meanings for the Serbian case system (Stevanović, 1964, Stanojčić & Popović, 1999; Kostić, 1965).

Table 2.3. List of Serbian cases and basic syntactic functions and meanings.

Case	**Main meaning**	**Example**
Nominative	Subject (always without prepositions)	***Marko*** *čita knjigu.* (**MarkoNOM** is reading a bookACC).
Genitive	Partitive genitive	*parče* ***kolača*** (peaceNOM of **cakeGEN**)
	Possessive genitive	*kuća moga* ***dede*** (houseNOM of myGEN grandfatherGEN)
	Ablative genitive	*Ja se sećam* ***Ane.*** (I remember **AnaGEN**.)
Dative	Indirect object	*Marko čita knjigu* ***bebi****.* (MarkoNOM is reading a bookACC to a **babyDAT**).
	Direction	*Marko ide prema* ***Ani****.* (MarkoNOM walks towards AnaDAT.)
Accusative	Direct object	*Marko čita* ***knjigu****.* (MarkoNOM is reading a **bookACC**).
Vocative	Communicative, in exclamations	*Ivane!* (IvanVOC)
Instrumental	Instrument or tools	*On je posekao prst* ***nožem****.* (He cut his finger with a **knifeINS**.)
	Accompaniment	*Marko je došao sa* ***drugovima****.* (Marko came with friendsINS.)
Locative	Place	*Ana sedi na* ***stolici****.* (AnaNOM sits on chairLOC.)

In addition, the list of functions and meanings presented in Appendix 1. demonstrates that the same functions can be carried by several cases. This indicates that the case a noun takes in given syntactic position is determined primarily by the verb's argument structure. For example, depending on the verb, the same subject noun may appear in four different cases (nominative, genitive, dative and accusative), as in (4-7).

(4) *Ana čita knjigu.*

AnaFEM-NOM read bookFEM-ACC.

'Ana reads a book'

(5) *Ani se spava.*

AnaFEM-DAT REFL sleep

'Ana wants to sleep'

(6) *Ane nema.*

AnaFEM-GEN not-exist

'Ana is not there.'

(7) *Anu boli glava.*

AnaFEM-ACC hurt headFEM-NOM

'Ana has a headache'

Serbian cases vary with respect to whether the noun can be used in combination with a preposition. Nominative and vocative cases never allow a preposition, whereas locative case is always used with one of six different prepositions. The other four cases can be used with or without prepositions, with the number of possible prepositions differing from case to case. Genitive case is the most

widespread, collocating with 57 different prepositions. Most Serbian prepositions take several different cases, e.g., the preposition *na* 'on' is used with accusative and locative, as in (8,9):

(8) *na stolicu*

onto a chairFEM-ACC

(9) *na stolici*

on a chairFEM-LOC

The particular case that will be used with a given preposition is determined by the verb's argument structure, as in (10, 11).

(10) *Ana se penje na stolicu.*

AnaFEM-NOM REFL climb on chairFEM-ACC

'Ana is climbing on the chair.'

(11) *Ana sedi na stolici.*

AnaFEM-NOM sit on chairFEM-LOC

'Ana is sitting on the chair.'

Examples (4)-(11) are showing a unidirectional relation between nouns and their heads, where verbs and prepositions are determining the case of a noun. In contrast to this type of relation, the case and gender in which nouns are used can affect the grammatical status of other words, like adjectives, pronouns, some numbers and the past participle of verbs. All those words are usually labelled as pronominal

because they have to agree or to be congruent with the noun they are attached to. Examples (12)-(15) give illustrations for all types of agreement:

(12) adjective agreement

lep Jovan

'beautifulNOM.SG.MASC Jovan NOM.SG.FEM'

(13) pronoun agreement

tvoja Ana

'your.NOM.SG.FEM Ana NOM.SG.FEM'

(14) number agreement

jedno dete 'one.NOM.SG.NEUT child NOM.SG.NEUT'

(15) past participial agreement

Jovan je došao.

'Jovan NOM.SG.MASC to be AUX come PAST.PARTICIPLE.MASC.SG.'

The overview of Serbian inflectional noun morphology provided in this chapter showed that Serbian has a complex, but also relatively regular noun morphology. Previous corpus studies and experimental research on the acquisition of Balto-Slavic languages with similar noun morphology systems (Kempe et al., 2003; Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press; Ševa et al., in press) showed that diminutives, as one of the most prominent features of CDS may facilitate the acquisition of the gender and case categories. The next chapter will present the first corpus analysis of the distribution of diminutives in Serbian CDS, in order to extend the cross-linguistic database on distribution of diminutives in different CDS. Specifically, it would be important to see whether the frequency of diminutives in

Serbian CDS matches the frequency observed in morphologically similar Russian, Polish and Lithuanian CDS.

Chapter 3

3. Corpus analysis of diminutive usage in Serbian child-directed speech

This chapter will provide the first detailed analysis of the distribution of diminutives in Serbian CDS. As described in the Introduction, the cross-linguistic corpus analyses have shown that the frequency of diminutives in CSD can vary from 3% in German to 45% in Russian, despite their almost universal semantic and pragmatic features. Thus, the detailed corpus analysis of Serbian CDS will extend descriptive cross-linguistic comparisons[11] of the variability in diminutive usage in this register. Moreover, given that most of the previously described corpus studies on the distribution of diminutives in languages other than English were mainly based on examples of utterances for only one parent-child conversational dyad recorded over several sessions, the extensive corpus analysis of Serbian CDS will give us a unique opportunity to examine the distribution of diminutives in a larger sample of one language. In the case of Serbian, I will have access to longitudinal data for eight couples of mothers and fathers, obtained from the *Serbian Corpus of Early Child Language* (Anđelković, Ševa & Moskovljević, 2001), where parents were addressing four girls and four boys over a time course of two years. After a description of the general morphological, semantic and pragmatic features of Serbian diminutives, I will present the results of the coding of Serbian nouns for their derivational (diminutives vs. simplex) and grammatical status (gender and declension categories). Moreover, in order to provide an adult baseline of diminutive usage for Serbian, I will compare the results for the CDS register with the distribution of diminutives in Serbian ADS and written language,

[11] A more detailed discussion of the limitations of cross-linguistic comparisons of the production of diminutive in CDS was provided in the Introduction.

based on the samples from the *Conversational corpus of Serbo-Croatian language* (Savić & Polovina, 1989) for the ADS register and the *Frequency Dictionary of Serbian Contemporary Language* (Kostić, 1999) for written language.

In addition, I will look in more detail at the extent to which age and gender of the children affected parental diminutive usage, as well as whether there is any difference between mothers and fathers.

3.1. Terms for the expression of affection and endearment in Serbian

This part of the chapter will present a detailed morphological analysis of Serbian diminutives, a derivation which is used by parents for the expressions of affection and endearment. An elaborate description of this cluster of words will give us the opportunity to spot small linguistic differences between Serbian and other Slavic languages which may influence the frequency of diminutives in Serbian CDS. In addition to the diminutives, I will provide a description of hypocoristics, another derivation which is also used for the expression of endearment and affection towards children. The existence of this derivation which expresses similar meanings may also influence the overall distribution of diminutives in Serbian CDS.

3.1.1. Morphological characteristics of Serbian diminutives

Diminutivisation is a productive process in Serbian, i.e. diminutives can be derived from:

a) almost all concrete nouns – *sto* 'table' – *stoč**ić*** 'tableDIM';

b) some abstract ones – *želja* 'wish' – *želj**ica*** 'wishDIM';

c) some adjectives/adverbs – *hladan,-a, -o* 'cold' – *hladnj**ikav**,-a,-o* 'coldDIM'

Several suffixes are used in the process of diminutive derivation. In Serbian, the most frequent noun diminutive suffixes are:

a) *-ić* for masculine – *lav* 'lion' – *lav**ić*** 'lionDIM'

b) *-ica* for feminine – *krava* 'cow' – *krav**ica*** 'cowDIM'

c) *-ce* for neuter nouns – *selo* 'village'– *seo**ce*** 'villageDIM'

In addition, there is a set of complex derivatives of masculine and neuter suffixes: *–čić*, *–ance, –ence, –ašce, –ešce*, as well as more archaic and regional forms like *–če*, *–ak* and *–ac* and their derivatives: *–inče, –uljak, –onjak, –arak, –urak, –ečak*.

Note that in Serbian, as in Russian and Polish, diminutive suffixes retain the grammatical gender of the simplex form of the noun. Moreover, for the non-transparently gender-marked nouns described earlier, diminutives provide an ending which is transparently gender marked, e.g. *stvar* 'thing (FEM)' – *stvarč**ica*** 'thingDIM (FEM)'. In addition to this, diminutives can highlight the stem used in the derivation of the word within the transparently marked gender classes which contain certain morpho-phological alternations like, for example, a 'shift from *l* to *o* at the end of the word': *posao* 'job (NOM. MASC)' vs. *posla* 'job (GEN.MASC)' – ***posl**ić* 'jobDIM (NOM. MASC)' vs. ***posl**ića* 'jobDIM (GEN.MASC)'.

Also, Serbian has lexicalised or 'frozen' diminutives, i.e. nouns ending in a diminutive suffix which have taken on a meaning quite different from the corresponding simplex noun. For example, the Serbian *četka* means 'brush' but the diminutive *četkica*, in addition to 'small brush' is usually used to denote a tooth brush. Apart from denoting smallness, endearment and affection, Serbian diminutives suffixes can be used in some other derivational processes like nominalisation of adjectives and adverbs or derivation of compound nouns. For example, the Serbian suffix *–ica* which is considered to be the most productive suffix in Serbian (Stevanović, 1964; Klajn, 2003), can be used not only for the derivation of diminutives, but also as a suffix which changes the gender of simplex nouns from masculine to feminine (e.g. *lav* 'lion (MASC)' - *lavica* 'lioness (FEM)' or as a suffix for simple noun derivations, where the new noun is semantically related to the stem (e.g. *sto* 'table (MASC) – *stolica* 'chair (FEM)'). Similar polyfuctionalitaty of the

diminutive suffixes exists also in Polish, where *–ka* suffix can be used to derive the feminine form of the base noun (e.g. aktor 'actor (MASC)' – aktorka 'actress (FEM)', but it seems that this process is more productive in Serbian than in other Slavic languages.

3.1.2. Morphological characteristics of Serbian hypocoristics

In addition to diminutives, Serbian also distinguishes a class of hypocoristics, a similar derivation for the subjective expression of endearment and affection. The most common suffix used for the derivation of hypocoristics is *–a* (and derivates *–ca*, *–ča*) which is used for all genders. Unlike diminutives, hypocoristics can only be derived from a limited number of words, usually from: a) *proper nouns*: *Aleksandar* (MASC) – *Aca* (FEM.HYP), *Nikola* (MASC) – *Nidža* (FEM.HYP), *Marija* (FEM) – *Maca* (FEM.HYP), etc.; b) *some animal names*: *medved* MASC 'bear (MASC)' – *meda* 'bearHYP (FEM)', *pas* 'dog (MASC)' – *kuca* 'dogHYP (FEM)', *mačka* 'cat (FEM)' – *maca* 'catHYP (FEM)' and c) *kinship terms*: *ujak* 'uncleHYP (MASC) – *ujka* 'grandfatherHYP (FEM)', *baba* 'grandmother (FEM)' – *baka*FEM 'grandmotherHYP (FEM)', etc. In contrast to diminutives which maintain the gender and declension of simplex nouns, hypocoristics change the gender and declension class of masculine nouns by feminising them. In addition to this, hypocoristics can be diminutivised by adding the suffix *–ica*, e.g. *baka* 'grandmotherHYP (FEM)' – *bak**ica*** 'grandmotherHYP+DIM (FEM)' or *tata* 'fatherHYP (FEM)' – *tat**ica*** 'fatherHYP+DIM (FEM)'.

Although hypocoristics exist in other Slavic languages, like Russian, it seems that this derivation is a little bit more productive in Serbian than in Russian, where hypocoristics are usually derived only from proper nouns and a few animal names.

3.2. Diminutives and hypocoristics in Serbian CDS

3.2.1. Description of the corpus and the sample

The estimation of the distribution of diminutives and hypocoristics in Serbian CDS was based on the *Serbian Corpus of Early Child Language* (Anđelković et al., 2001). The corpus was compiled out of longitudinally video recorded sessions of natural and free face-to-face interactions between a child and its family members. The sample of children included 4 boys and 4 girls. The recording procedure started when the children were 18 months old, and ended at 48 months of age. The sessions took place every two months (16 samples per child), and lasted 90 minutes per session. In addition to the recording of language development, the general cognitive development of all children was assessed. Different psychological scales and inventories were applied every six months. These assessments were recorded too which added an additional 30 minutes in some sessions (120 minutes in total). The whole material was transcribed using the *Child Language Data Exchange System* (CHILDES, MacWhinney, 2000). With this system, we tried to cover all aspects of natural communication, and therefore our transcripts included not only words and utterances with codes for repetitions, interruptions, etc., but also additional information on activities, nonverbal and gestural turns, paralinguistic information, comments, information about the addressee, etc.

Overall, the corpus contained over 1,000,000 word tokens with two thirds of utterances produced by adults addressing children and one third of children's utterances.

In order to estimate the distribution of diminutives out of the total number of nouns in CDS, I sampled the utterances of eight couples of parents, produced when the children were 1;8, 2;2, 2;8, 3;2 and 3;8 years old. This sample contained around 30,000 utterances and approximately 112,000 word tokens.

Since the structure of interactions between adults and children was not controlled during the recordings in terms of who was going to participate in the conversational situations and how much, a balanced presence of mothers and fathers in all sessions could not be obtained. In most cases the mothers were carrying the interactions with approximately 80% of utterances and word tokens.

3.2.2. Coding of the corpus

3.2.2.1. Lemmatisation

In order to determine the percentage of diminutives out of all nouns, the utterances produced by the parents were first lemmatised. This means each token in the utterance was coded for its part of speech category: noun, verb, adjective, etc. and labelled with the token's lemma, i.e. the citation form (basic form of a word which is used as dictionary entry). For example, for the token *kuću* 'houseACC' which is a noun in accusative singular the citation form is *kuća* 'houseNOM', i.e. the same noun in nominative singular.

In order to achieve this level of coding (semi)automatically, I had to adapt the MOR program from the CHILDES package (MacWhinney, 2000) which provides a method for the (semi)automatic tagging of corpora. Appendix 2 describes in more detail the procedure for lemmatisation, conducted with the version of the MOR program adapted for Serbian. In addition, I manually coded the unidentified tokens (approximately 8%) and checked the coding of the nouns.

Out of approximately 112,000 tokens, the parents produced 15,786 noun tokens and 2033 different lemmata, with 12,250/1660 common noun tokens/lemmata. The mothers produced 10,147/1479 common noun tokens/lemmata and the fathers produced 2107/600 common noun tokens/lemmata. In this thesis, estimates of frequency of diminutives in Serbian CDS were based only on the set of common nouns.[12]

[12] In the corpus analysis of diminutive frequency in Russian CDS, Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press, excluded proper Russian nouns from their calculations of diminutive frequency, mainly because diminutives are used very productively for the derivation of nick names (*Ivan* SIM MASC-*Vanjechka* DIM MASC, *Tatjana* SIM FEM – *Tanjechka* DIM FEM). A similar effect was observed in Serbian where both diminutive and hypocoristics are used for the derivation of nick names (*Nikola* SIM. MASC – *Nidža* HYP MASC, *Nada* SIM. FEM – *Nadica* DIM.FEM). Since these items can be treated in both languages as frozen elements of parent-child discourse, including them into the general count of diminutives would unjustifiably inflate the number of diminutives and hypocoristics. Thus, for both languages, estimates based only on common nouns would provide more realistic approximations of the true diminutive productivity in CDS. In order to maintain the same sampling criteria as in Russian, I coded only common nouns which will allow for a more direct comparison of the two languages.

3.2.2.2. Coding of nouns

Common nouns were manually coded for the following morphological categories: a) derivation; b) grammatical gender and declension classes and c) word-play conditions.

a) The **derivational coding** contained the following categories:

1. **simplex nouns** – nouns which are in their simplex form, but which can be productively diminutivised (e.g., *kuća* 'hous' – simplex vs. *kućica* 'house' – diminutive form]
2. **diminutive nouns** – nouns which are in the diminutive form, ending in one of the diminutive suffixes, like *–ica*, *–ić* or *–ce*.
3. **lexicalised or 'frozen' diminutives** – nouns which originated as diminutives, but over time became highly lexicalised or lost their diminutive meaning, e.g. *devojčica* 'little girl' originated as the diminutive form of *devojka* 'girl', but today the term *devojčica* is the only way to express reference to a female child.
4. **hypocoristic nouns** – nouns used for the subjective expression of smallness, endearment and affection like: *tata* 'fatherHYP', *deka* 'grandfatherHYP', *ujka* 'uncleHYP'.
5. **non-diminutivisible nouns** – nouns which can be used only in the simplex form, mainly abstract nouns like *ljubav* 'loveSIM'.

b) Coding of **grammatical gender and declension classes** – The following chapters in this thesis will describe experimental studies of the effect of the diminutive distribution in CDS and noun gender and case acquisition. They address

the question as to what extent information from the input is related to the children's performance on gender categorisation and case marking of novel nouns, both in diminutive and simplex form. For that reason, I have coded the nouns additionally for their:

1. **grammatical gender**: masculine, feminine and neuter, based on a noun's agreement with pronominal words; and

2. membership in one of the four **declension classes** in Serbian: the feminine class ending in *–a*, the masculine class ending in consonants, *–o* or *–e*, the neuter class ending in *–o* or *–e* and the feminine opaque class ending in consonants.

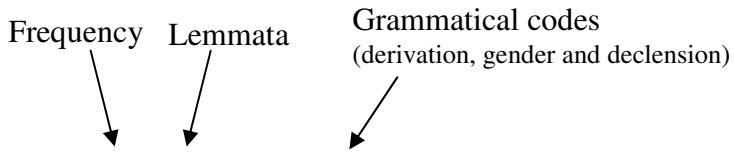
c) Coding of **word-play conditions** - In addition to the facilitating effect of high frequency and morpho-phonological saliency of diminutives, Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press argued that diminutives also facilitate acquisition of noun morphology through parental word-play. Parents often alternate between simplex and diminutive forms of the same noun stem, without introducing a semantic distinction, and usually in close proximity within the same conversational episode. As a result, approximately 9% of noun stems are produced both in simplex and diminutive form in Russian CDS. Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press showed that children were performing better in a case-marking task, when the experimenter alternated between the simplex and diminutive forms of novel nouns or presented novel nouns only in diminutive form (a more detailed description of this study is provided in the Introduction). In order to see whether Serbian parents are also using word play between simplex and diminutive forms of the same nouns, each noun lemma within one recording session was coded whether it was used exclusively in simplex form, exclusively in diminutive form, exclusively in hypocoristic form, or as alternating between the derivations.

3.2.3. Results

3.2.3.1. Frequency of diminutives in Serbian CDS, as well as the distribution of gender and declensions classes

A cumulative list for all nouns and separate lists for different age groups as well as boys vs. girls and mother vs. father, with the frequency of noun lemmata, and the codes for derivation, gender and declension class for all samples were extracted from the corpora with the FREQ program (CHILDES package, MacWhinney, 2000). Example (1) presents the output for one of the lists extracted with FREQ program:

(1)



3 n|Tanja&HYP:FEM:FEM

1 n|Zemun&NDM:MASC:MASC

1 n|album&SIM:MASC:MASC

1 n|autić&DIM:MASC:MASC

The frequency information indicates how many tokens were used in the different morphological categories. In addition, the number of different lemmata shows how rich the lexicon is within each of the grammatical classes. Figures 3.1.-3.3. present the distribution of different derivational, gender and declension classes for common noun tokens and lemmata.

The cumulative list of all lemmata revealed that parents used 7% diminutives out of all noun tokens, and 11.7% out of all noun lemmata. In contrast, the percentage of hypocoristics was 17.2% out of all noun tokens and 2.2% out of all noun lemmata.

These results indicate that the parents were deriving the diminutives more readily than hypocoristics, but that these diminutives were used only once or twice, in contrast to the few hypocoristics which were repeated continuously. Moreover, this analysis demonstrated a rather surprising difference in the distribution of diminutives in Serbian CDS, in comparison to morphologically similar Russian and Polish, with 25-40% of diminutives.

The analysis of diminutive usage across gender and declension classes showed that, overall, diminutives are mainly derived from feminine nouns (approximately 80% of diminutive nouns), indicating that the suffix *–ica* is the most frequent diminutive suffix in Serbian. Detailed results for the distribution of diminutives and other derivational classes across the three genders and four declension classes are presented in Tables A3.1.-A3.2. in Appendix 3.

Also, the analysis of the overall distribution of gender and declension classes showed that parents repeated nouns in feminine gender and declension more frequently than masculine, with 52% of noun tokens in feminine gender and over 60% of nouns declined with feminine declension. At the lemma level, masculine and feminine nouns were distributed almost equally.

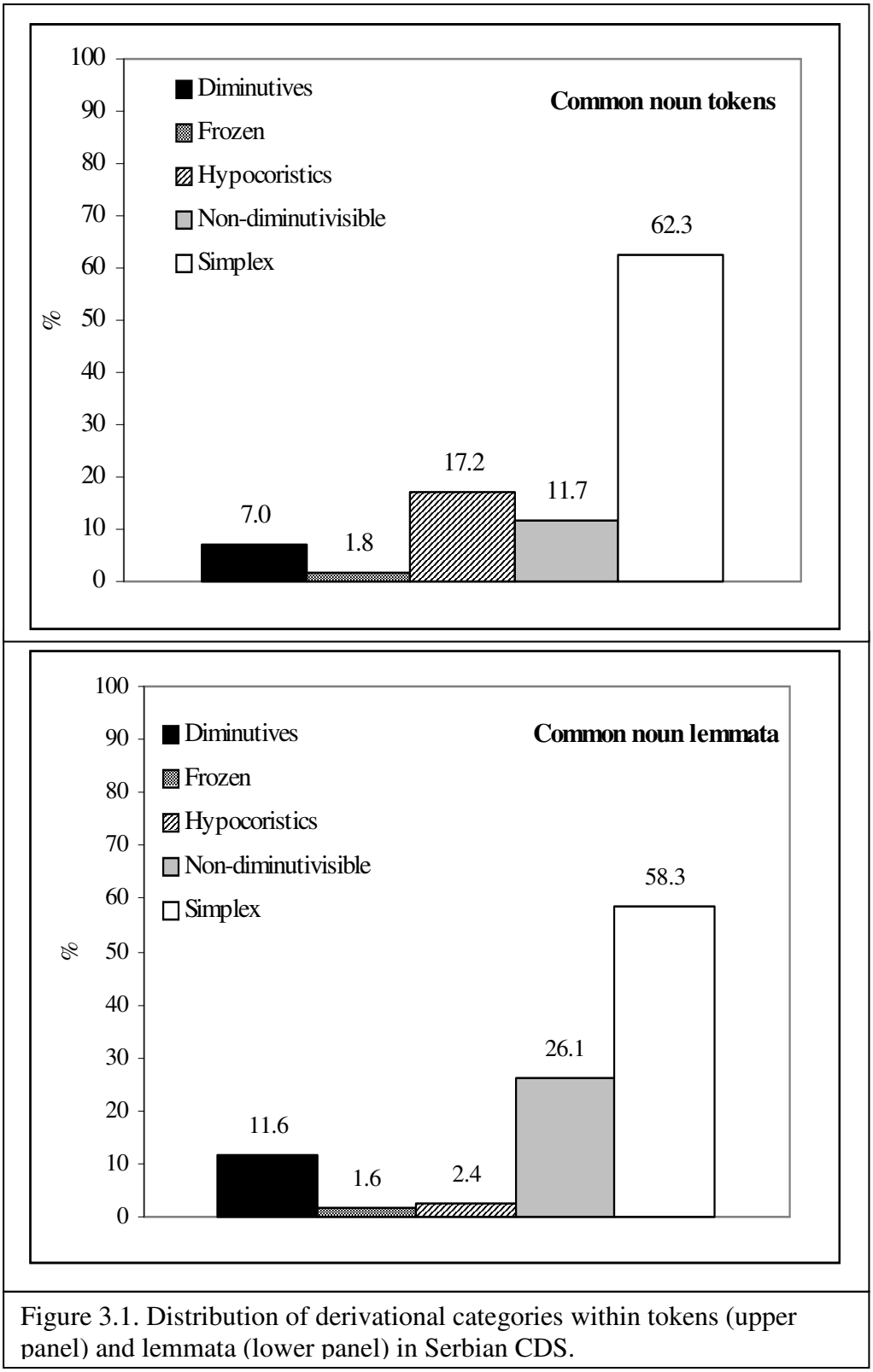


Figure 3.1. Distribution of derivational categories within tokens (upper panel) and lemmata (lower panel) in Serbian CDS.

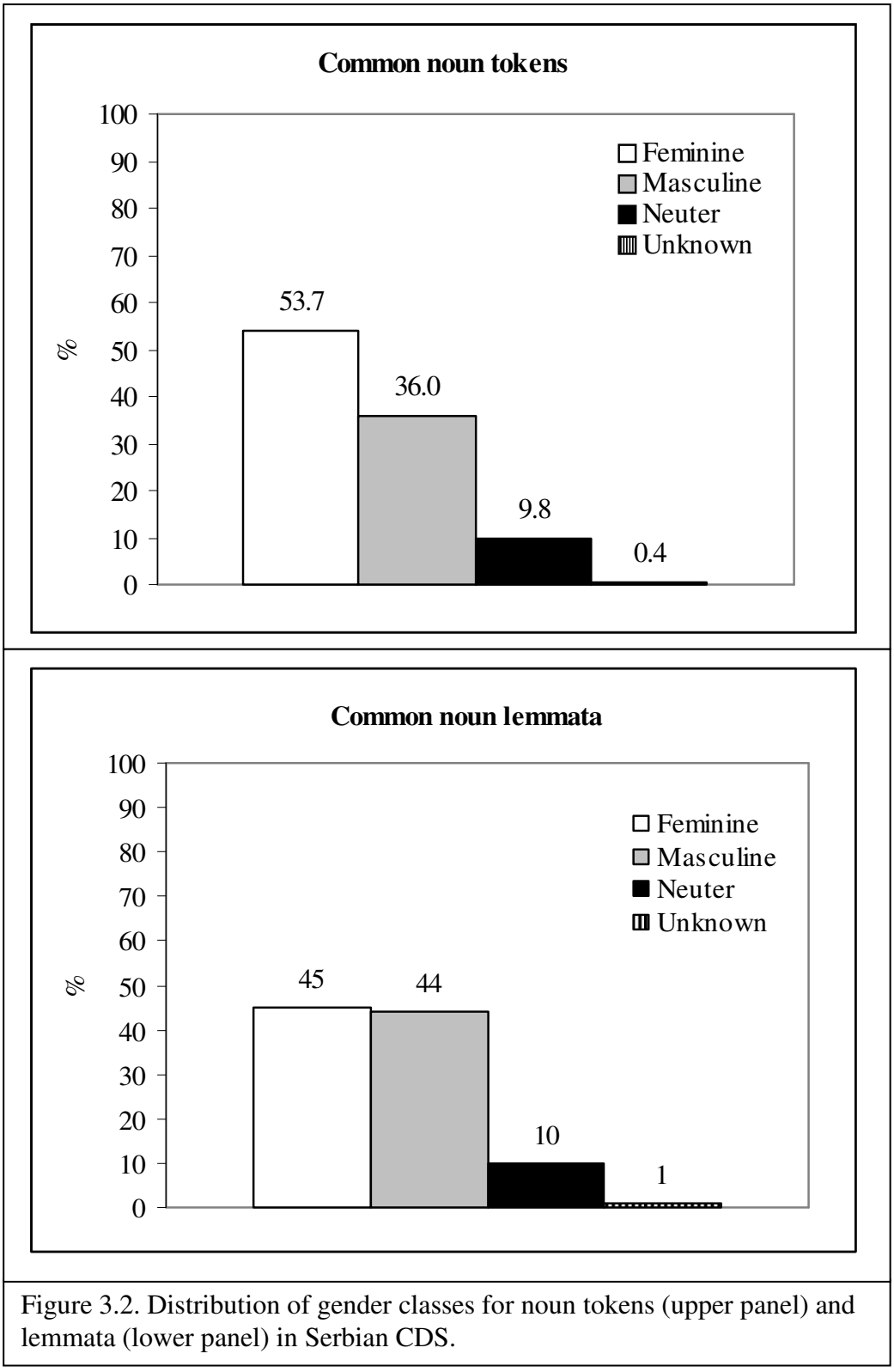


Figure 3.2. Distribution of gender classes for noun tokens (upper panel) and lemmata (lower panel) in Serbian CDS.

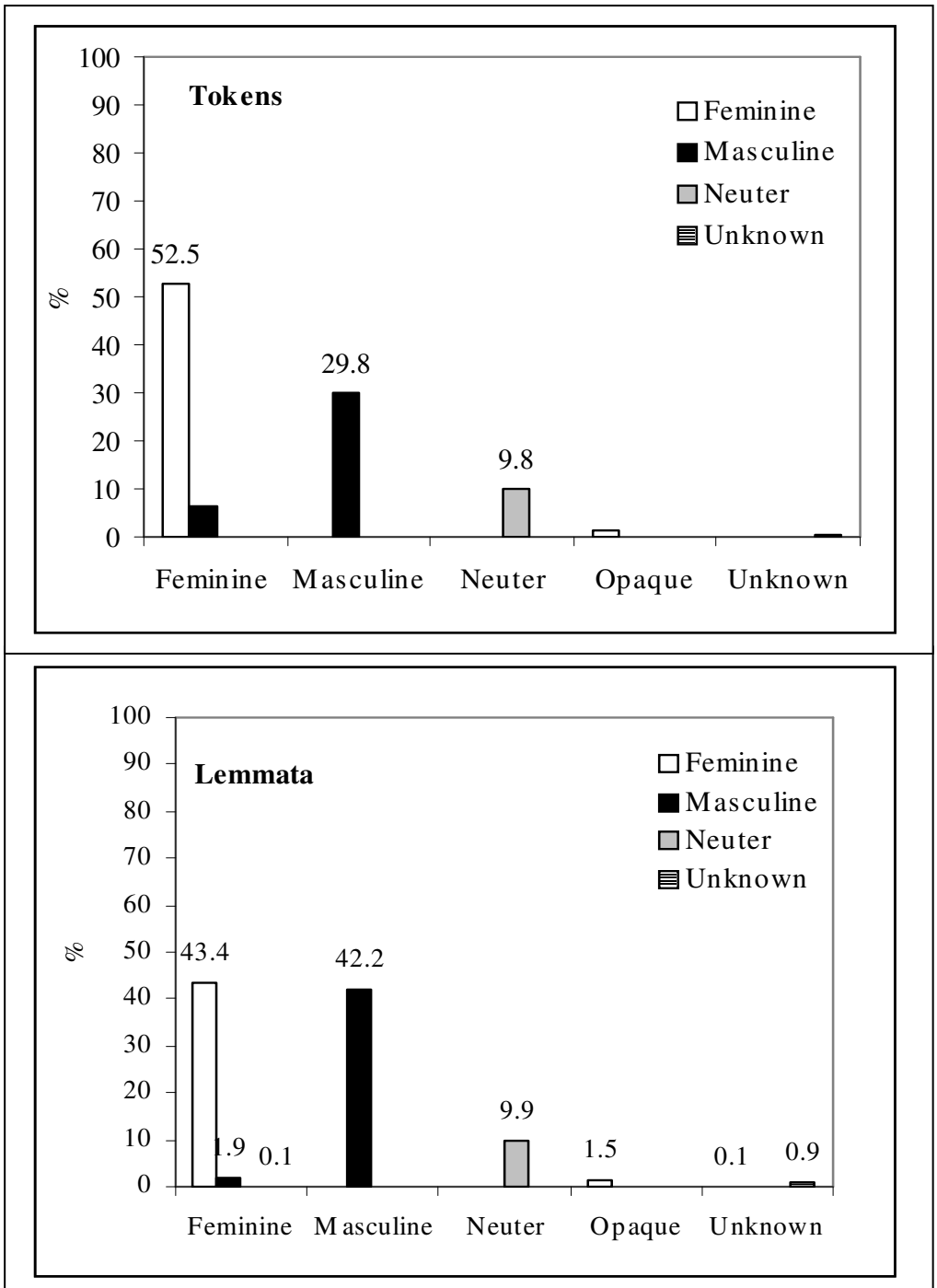


Figure 3.3. Distribution of declension classes, together with gender distribution within each declension class, for noun tokens (upper panel) and lemmata (lower panel) in Serbian CDS.

3.2.3.2. Qualitative and quantitative description of word-play in Serbian CDS

The word-play between simplex and diminutive forms of the same noun stem, present in Russian CDS, was observed in Serbian CDS as well.

Consider example (1) recorded when M.G. was 1;8 years old, where his mother is taking him to play in the backyard and ride a bike. Within the same conversational turn, she used the noun *bicikl* 'bike' both in simplex and diminutive form, with no clear semantic motivation, referring to the same four-wheel bike which children usually ride at that age.

(1) **1;8 Boy M.G.**

*MOT: oćemo nositi biciklić ?

'Shall we take a **bikeDIM.**'

*CH1: da .

'Yes.'

*MOT: da poneseš, a?

'Do you want to carry it?'

* MOT: ajde da ponesemo bicikl da malo voziš .

'Let's take the **bikeSIM**, so you can ride it a little bit.'

Example (2) contains word play within the same utterance, when the mother of A.Nj., 2;2 years old, is talking about their visit to the child's aunt and their trip by plane.

(2) 2;2 A.Nj.

*MOT: u avionu, da vidiš oblake, pa ćemo ić(i) da vidimo čiku što vozi avončić .

'In the **airplaneSIM**, so you can see the clouds, and we will see the manHYP, who drives the **airplaneDIM**.'

The next example (3) is a very nice illustration of word-play which is actually introduced by the child, and where the mother accepts the simplex-diminutive alternation in her utterances by imitating the child's variants of the words. Spontaneous word-play introduced by children was also observed with Russian children in the experimental study on word-play in novel nouns (Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press), where children independently introduced word-play for approximately 12% of words which were presented in the diminutive only and in the simplex only condition. In this example, A.Nj., 2;2 years old, is talking to her mother about playing in the sand-box and about the tools she needs: a rake, a can, a shovel, etc. The child starts first with the diminutive form of the word rake, *grabljice* and uses the simplex form, *grablje,* in the next utterance. The mother starts first with the diminutive form as well, and as soon as the child changes the word into simplex, she imitates the change and introduces the simplex form. This example is very interesting because it shows that parents not only are using word-play when addressing their children, but they also encourage children's usage of word-play, by imitating them. In this way, parents maintain emotional responsiveness with

their children (King & Melzi, 2003, 2004; Dabrowska, 2006), but also reinforce the facilitating effect of word-play on morphology learning.

(3) 1;8 Girl A.Nj.

*CH2: gabice .

'A **rakeDIM.**'

*MOT: i grabljice, i šta još nedostaje uz grabljice ?

'And the **rakeDIM**, and what else is missing apart from the **rakeDIM**.'

*CH2: deke peska .

'uninteliligible sandSIM.'

*MOT: peska, i šta još ?

'SandSIM and what else?'

*CH2: zato peska gabe .

'unintelligible sandSIM **rakeSIM**'

*MOT: grablje, pesak i ?

'**RakeSIM**, sandSIM, and?'

*CH2: i: .

'And.'

*MOT: kantica i ?

'**CanDIM** and?'

*CH2: kantica gabe .

'**CanDIM rakeSIM.**'

*MOT: grablje i lopata .

'**RakeSIM** and **shovelSIM**.'

In order to quantitatively compare the word-play production in Serbian CDS with Russian CDS, I performed an analysis of the cumulative list of noun lemmata for the utterances of each parental couple across the five age groups (40 samples). Table 3.1. presents the mean percentage for each condition across five age groups. The results indicate that the Serbian parents tended to use slightly less word-play, with 4% of noun stems on average in comparison to the 9% of noun stems in Russian CDS. Still, the qualitative comparisons showed that Serbian parents interchange simplex and diminutive forms in very similar manner to Russian mothers which is also reflected in spontaneous word-play both by Russian and Serbian children.

Table 3.1. Mean percentage of noun stems used in diminutive-simplex and diminutive-hypocoristic forms, as well as with simplex, diminutive and hypocoristic forms witin on recording session.

	1;8	**2;2**	**2;8**	**3;2**	**3;8**
diminutive	7.1	4.1	4.3	4.3	3.1
diminutive-hypocoristics	0.2	0.2	0.1	0.3	0.1
diminutive-simplex	4.7	4.7	4.3	2.5	2.3
hypocoristics	7.6	7.8	7	3.5	4.6
simplex	80.4	83.3	84.3	89.3	90

3.2.4. Discussion

The corpus analysis of distribution of the diminutives in Serbian CDS showed that Serbian parents used only 7% of diminutive out of all common noun tokens. The observed result is very interesting because it presents the first exception from the previously observed high frequency of diminutives in CDS of other Balto-Slavic languages. Specifically, the diminutive production of Serbian parents is closer to German CDS, with 3% of diminutives, than to morphologically similar Russian and Polish, with 20-45% of diminutives. On the other hand, despite this large discrepancy in the frequency of diminutives between Serbian CDS and the CDS of other Balto-Slavic languages, a qualitative and quantitative analysis of word-play showed that Serbian parents tend to interchange simplex and diminutive forms of the same word in a similar way to Russian parents. Furthermore, with the respect to general distribution of gender and declension classes, Serbian CDS exhibits a high frequency of feminine noun tokens both for diminutives and other derivational classes.

In order to determine whether the observed distributional patterns are a consequence of parental adaptations to the children's communicative and language learning requirements, we need information on the distribution of diminutives and gender/declension classes in an adult directed baseline, provided by corpora of adult-directed speech (ADS) and written language. Also, with respect to the low frequency of diminutives in Serbian CDS, one could argue that Serbian parents prefer hypocoristics as a marker of affection and endearment in conversations with children. Thus, in order to see whether there is an increase in diminutive usage compared to the adult baseline, I will present a corpus analysis of the distribution of diminutives in Serbian ADS and written language. These data will also serve as a point of reference

for comparisons of potential shifts in the distribution of gender and declension classes which may influence noun morphology acquisition.

3.3. Distribution of diminutives in Serbian adult-directed speech and written language

This part of the chapter will present a corpus analysis of Serbian ADS and written language which will provide the adult baseline for the comparisons of the diminutive production and the production of gender and declensions categories in Serbian CDS and other registers. So far the only quantitative comparison between CDS and ADS was provided for Russian, showing that Russian parents used significantly more diminutives when addressing their children (45% of diminutives) than when addressing adults (3% of diminutives) (Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press). Thus, the description of the adult baseline in Serbian will provide important information for further cross-linguistic comparisons on diminutive production.

3.3.1. Distribution of diminutives in Serbian adult-directed speech

3.3.1.1. Corpus description

The distribution of diminutives in Serbian adult-directed speech (ADS) was obtained from the *Conversational corpus of Serbo-Croatian language* (Savić & Polovina, 1989).

The corpus was collected during the 1980's and it contains 23 different spontaneous conversations between adults, in situations like family and friends' gathering, shopping, visits to doctors, phone messages, etc.

The overall corpus contains 31,073 words produced in 5,233 utterances. The utterances were lemmatised semi-automatically with the MOR program (MacWhinney, 2000), with additional manual tagging of unidentified words and checking of noun coding.

3.3.1.2. Noun coding

Nouns were manually coded for their derivational status, grammatical gender and declension class in the same way as for the CDS samples.

3.3.1.3. Results and discussion

The frequency of the different derivations was calculated based on the cumulative FREQ list which contained all conversational situations. Figure 3.4. depicts the distribution of all derivational classes, showing that in ADS both diminutives and hypocoristics are used considerably less than in CDS (7% for CDS vs. 0.7% for ADS). This indicates that despite the low frequency of diminutives in Serbian CDS, compared to Russian and Polish CDS, parents still tend to increase diminutive usage by about ten times when they are addressing their children in contrast to when they are talking to adults.

The analysis for the gender classes (depicted in Figure 3.5.) revealed a considerable difference between Serbian ADS and CDS, with the frequency of feminine tokens that was increased in CDS by 10% in comparison to ADS. On the other hand, noun lemmata were distributed similarly in CDS and ADS across genders, indicating that in CDS, parents were just repeating feminine nouns more often than

masculine or neuter nouns. This is probably due to the fact that lexemes which are repeated frequently and are closely semantically related to the CDS register, like body parts (e.g. *ruka* 'hand (FEM)', *glava* 'head (FEM)), kinship terms (e.g. *mama* 'mommy (FEM)', *baka* 'granny (FEM)') or toy names (e.g. *lopta* 'ball'), etc. are mainly feminine in Serbian. However, the noun lemmata in ADS were distributed similarly across different gender classes as in CDS. Thus, the shift in gender distribution is entirely due to token frequency, i.e. more repetition of feminine nouns.

The analysis of declension classes in Serbian ADS (see Figure 3.6.) showed that adults also change the distribution of those classes in addressing their children with an increase of the frequency of the feminine declension by almost 20% compared to 40% in ADS. This increase is not only due to the repetition of nouns which agree with the pronominal words in feminine gender, but also nouns which are in masculine gender, but which decline as feminine nouns, like *tata* 'daddy (MASC)', *deka* 'grandad (MASC)', *zeka* 'bunny (MASC)', etc. A more detailed distribution of declension classes across three genders and four declension classes is presented in Tables A3.17.- A3.18. in Appendix 3.

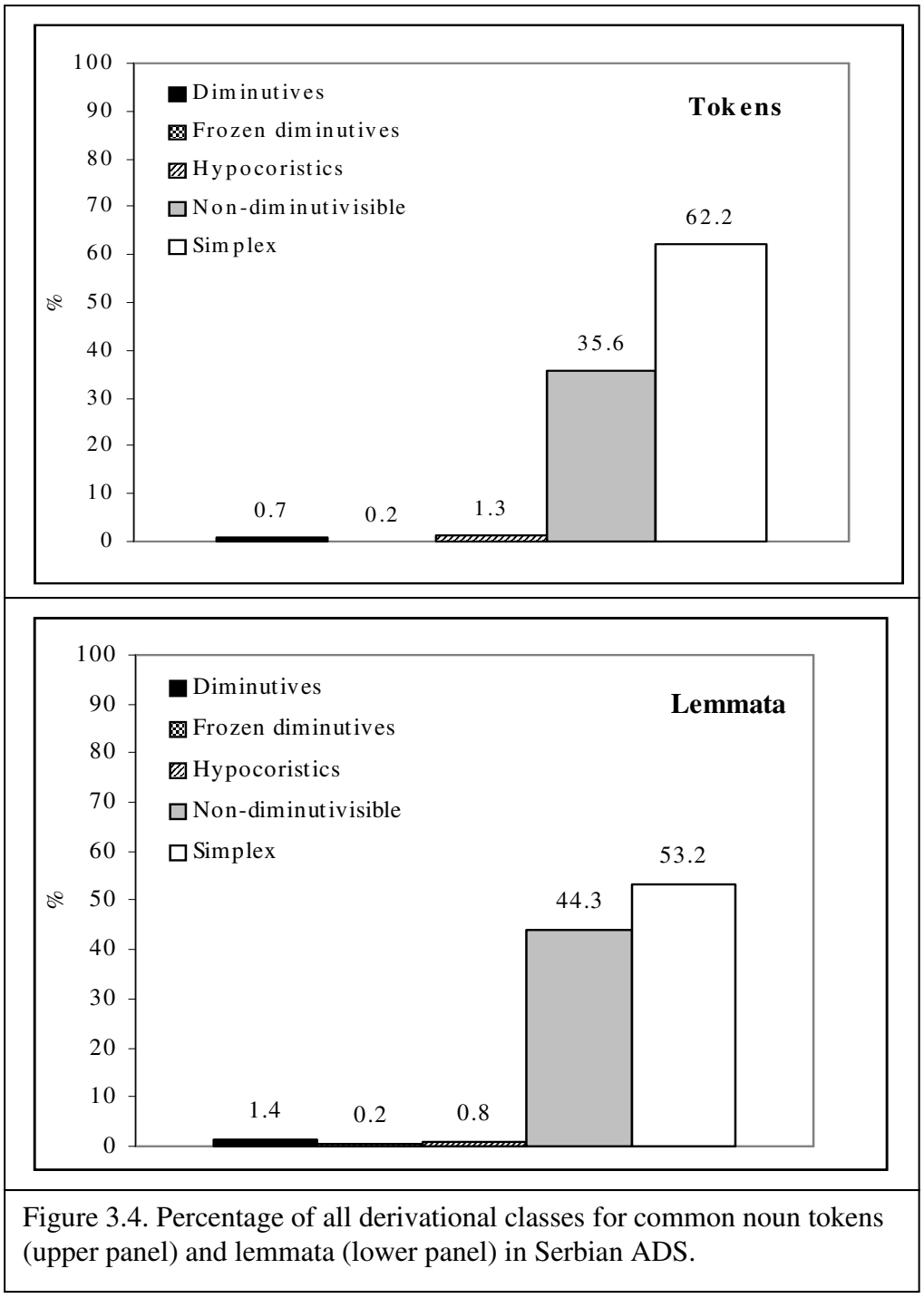


Figure 3.4. Percentage of all derivational classes for common noun tokens (upper panel) and lemmata (lower panel) in Serbian ADS.

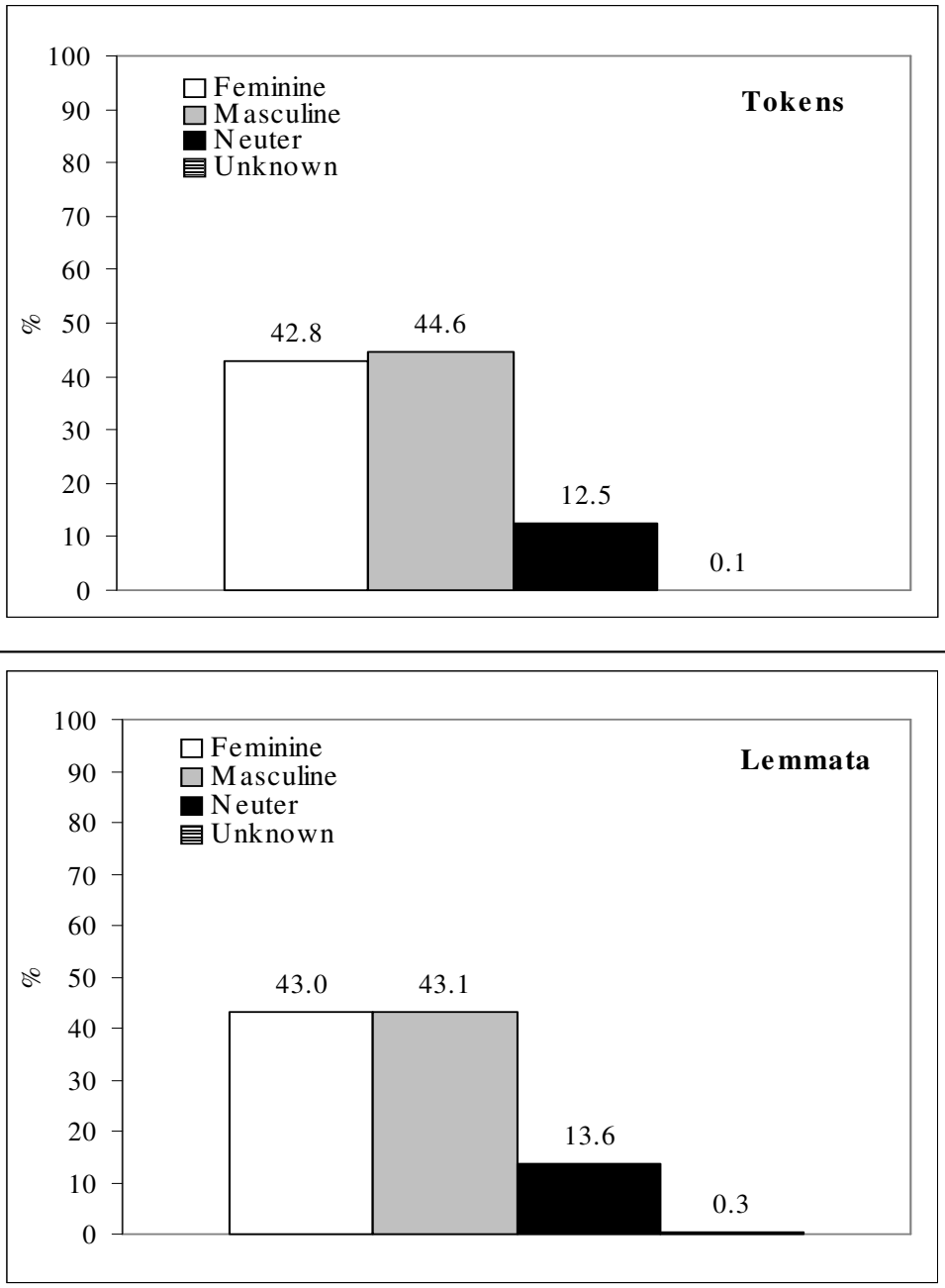


Figure 3.5. Percentage of all genders for common noun tokens (upper panel) and lemmata (lower panel) in Serbian ADS.

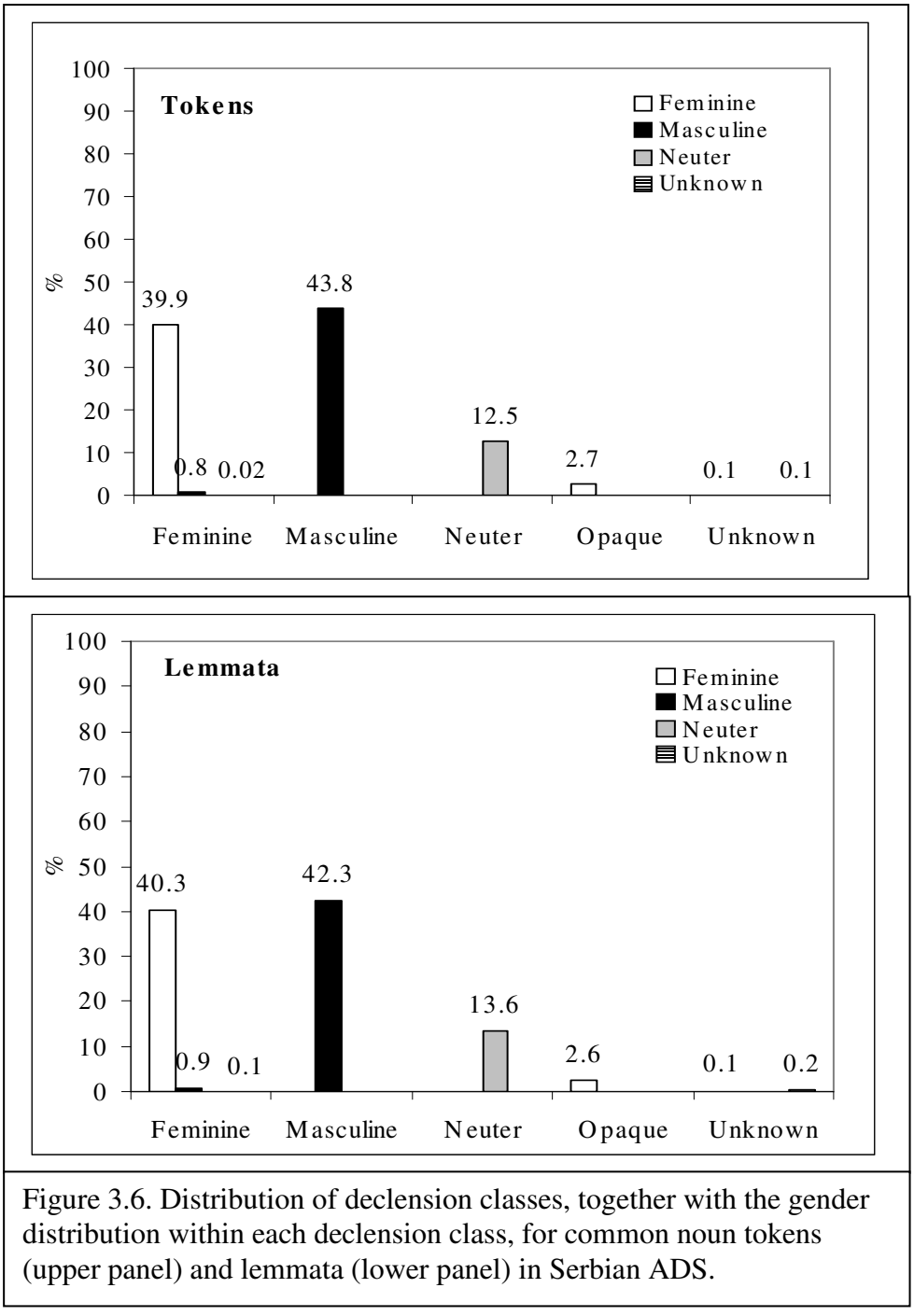


Figure 3.6. Distribution of declension classes, together with the gender distribution within each declension class, for common noun tokens (upper panel) and lemmata (lower panel) in Serbian ADS.

3.3.2. Distribution of diminutives in Serbian written language

3.3.2.1. Corpus description

The distribution of diminutives in written language was obtained from the *Frequency Dictionary of Serbian Contemporary Language* (Kostić, 1999). The dictionary contains 2,000,000 words, approximately 65,000 lemmata (36,000 nouns) and 240,000 different inflectional word forms. It was sampled from a collection of contemporary Serbian daily press articles and poetry, and represents one of the parts of *The Corpus of Serbian Language* (CSL), www.serbian-corpus.com (Kostić, 2001).

3.3.2.2. Noun sampling and coding

In order to estimate the percentage of diminutives in written language, I randomly selected a set of 2,000 noun lemmata[13], with an overall frequency of 35,389 noun tokens.

Nouns were manually coded for their derivational status, grammatical gender and paradigm class in the same way as for CDS and ADS samples.

[13] The frequency distribution of the 2,000 lemmata fitted the Zipfian hyperbolic function (Zipf, 1949) which means that the number of the most frequent words was twice as small as the number of the second most frequent words which in turn was two times less than a fourth of the most frequent words, etc. The same type of distribution was observed for the whole list of the nouns in the Frequency dictionary, indicating that the selected nouns constitute a representative sample of nouns in Serbian written language.

3.3.2.3. Results and discussion

The distribution of all derivational classes, gender and declension type for the noun tokens and lemmata for written language is presented in Figures 3.7.-3.9 and in more detail in Table A3.19.- A3.20 in Appendix 3.

A comparison of the distribution of derivational classes showed that the percentage of diminutives in written language is similar to ADS (0.6%), and roughly ten times less than in CDS (7%). In addition, the percentage of neuter nouns in written language was doubled in comparison to CDS and slightly higher than in ADS, probably due to the high incidence of abstract neuter nouns ending in the *–nje* suffix. Feminine nouns were also repeated a little bit more often than in ADS, but the frequency was still lower than the one observed in CDS. Ambiguously gender marked masculine nouns ending in *–a* which are declined following the feminine declension were very rare in written language (less than 1%) in contrast to the CDS sample with 10%. On the other hand, opaque feminine nouns which end in consonants were relatively more frequent in written language, with approximately 5% of lemmata compared to only 1% in CDS.

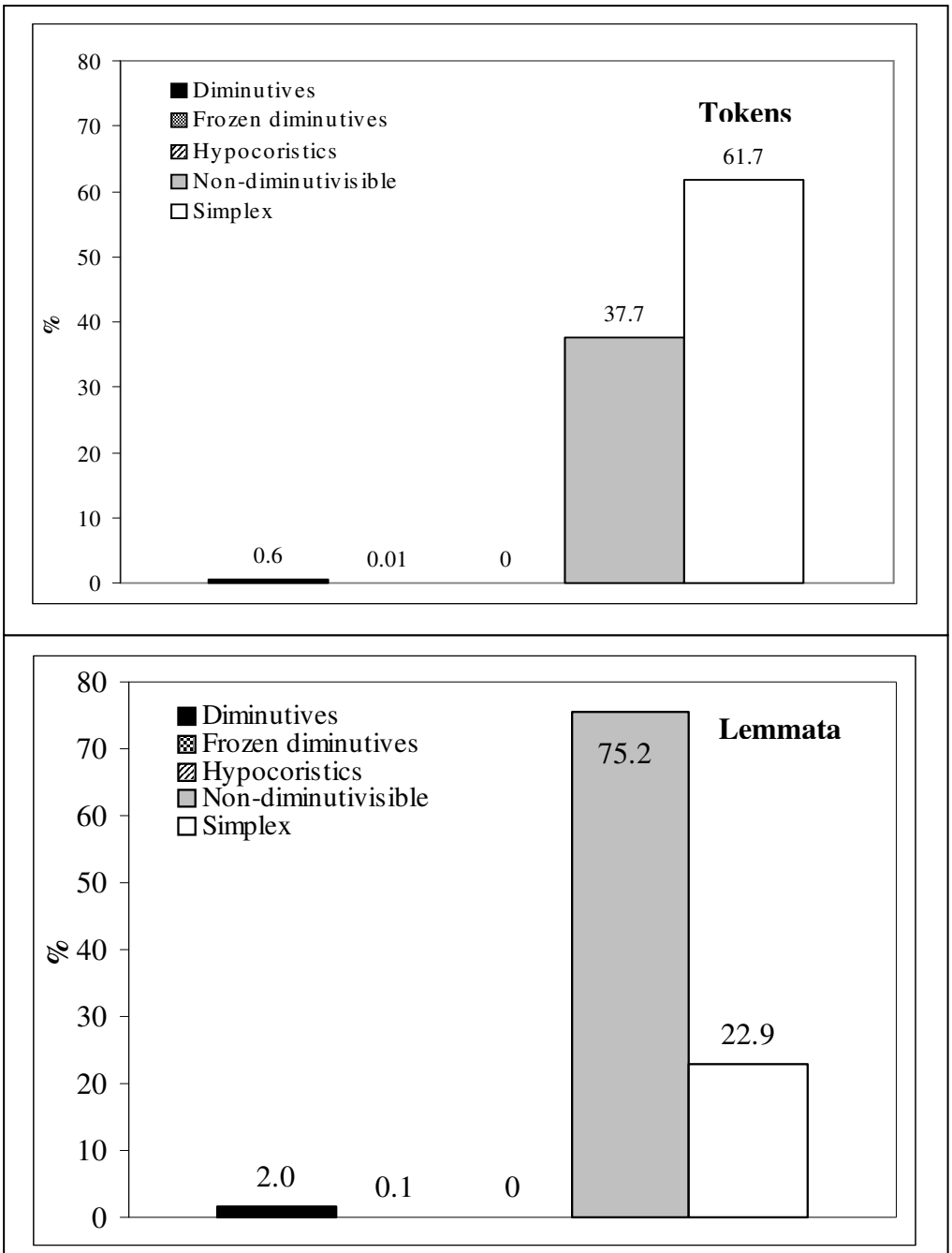


Figure 3.7. Percentage of all derivational classes for noun tokens (upper panel) and lemmata (lower panel) in Serbian written language.

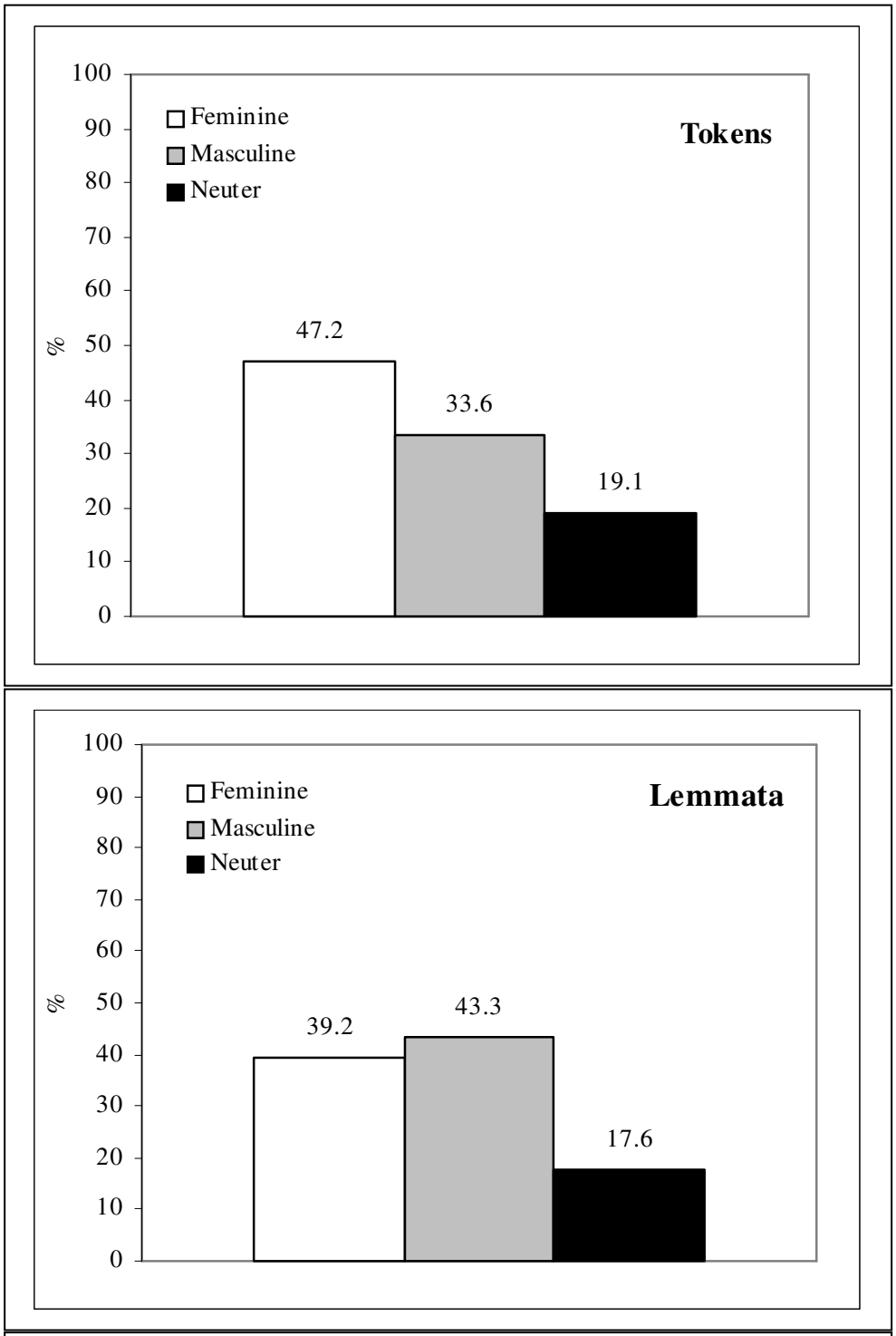


Figure 3.8. Distribution of gender classes for noun tokens (upper panel) and lemmata (lower panel) in Serbian written language.

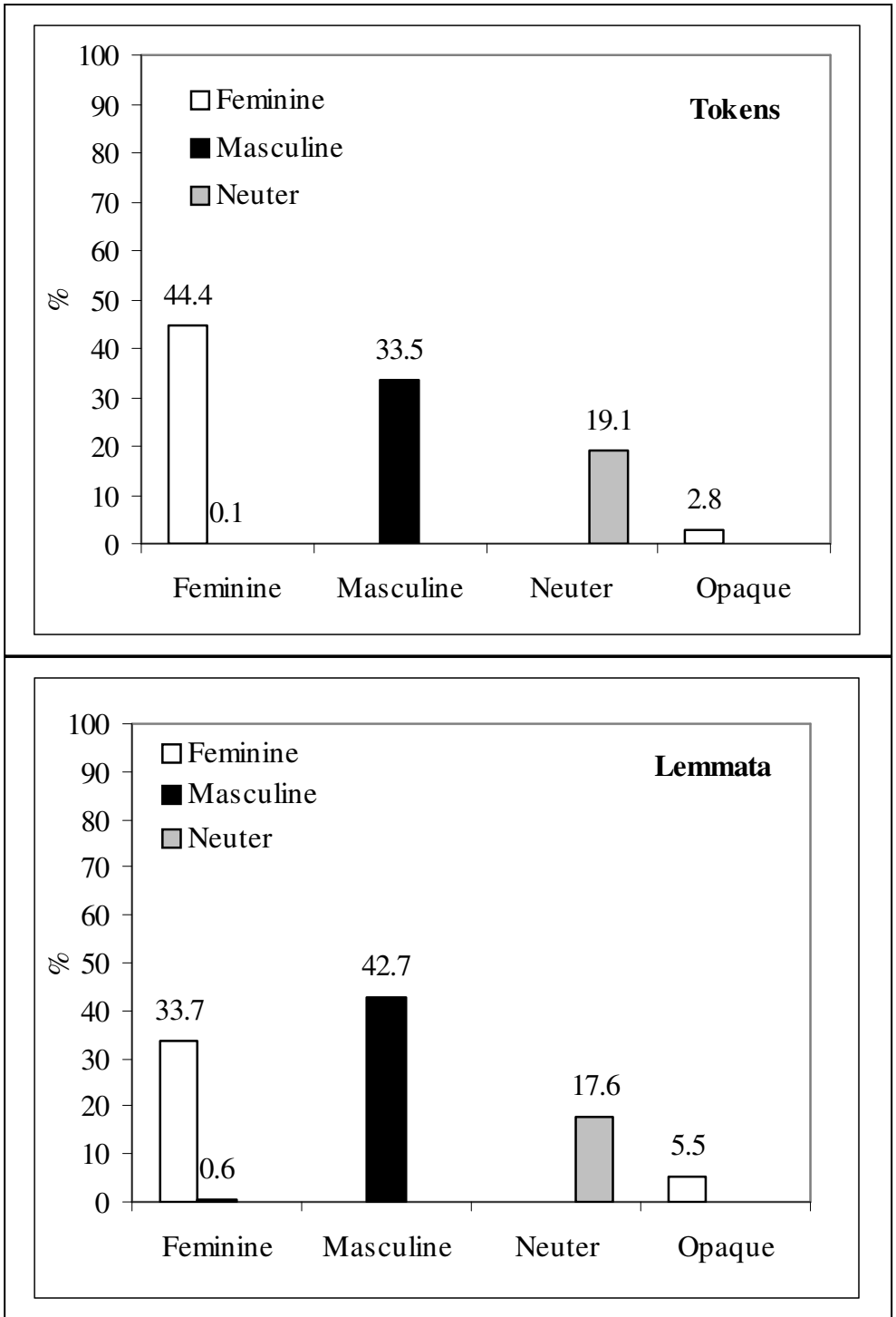


Figure 3.9. Distribution of declension classes, together with gender distribution within each declension class, for common noun tokens (upper panel) and lemmata (lower panel) in Serbian written language.

3.3.3. Discussion of the observed results on the distribution of diminutives in Serbian adult-directed speech and written language

The description of the adult baseline of diminutive production in Serbian indicated that Serbian parents tend to use approximately ten times more diminutives when addressing their children (7% of diminutive tokens), than when they talk to other adults (less than 1% of diminutive tokens). At the same time, the description of the distribution of gender and declension classes has shown some interesting shifts in Serbian CDS in comparison to the adult baseline. First, Serbian parents tend to use feminine gender and declension more frequently when addressing children which may increase overall phonological redundancy of the noun system given that feminine nouns always end in the same vowel *-a* in the nominative.

On the other side, the corpus analysis of these grammatical categories also revealed an increase in morphologically ambiguous nouns in CDS, mainly carried by hypocoristic masculine nouns ending with the typical feminine suffix *–a*, like *meda* 'bear HYP' or *deka* 'grandfather HYP'. This result indicates that the affective aspect of CDS might overwrite the facilitative dimension of this register for language learning.

Taken together, the corpus analysis of the adult baseline in Serbian showed that Serbian adults use similar number of diminutives when addressing other adults as in Russian ADS (1% for Serbian ADS vs. 3% for Russian ADS) (Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press). This suggests that the difference between Serbian and other Balto-Slavic languages in the production of diminutives exists only for the CDS register. I speculate that this is probably due to the combination of small socio-cultural and linguistic differences between these systems.

The last part of this chapter will attempt to illuminate some of the possible factors responsible for the observed overall low frequency of diminutives in Serbian CDS.

3.4. Factors which may influence the distribution of diminutives in CDS

The overview provided in the Introduction of this thesis showed that the production of CDS can be highly sensitive to factors like the age of children, the gender of speakers and addressees, etc. Given that the sample used to estimate the diminutive usage in Serbian CDS covered a wide age range, was produced in communication with equal number of girls and boys, and also represented a combination of utterance produced by both mothers and fathers, the corpus of Serbian CDS offers a unique opportunity to check whether these factors affected the frequency of diminutives in this sample of Serbian CDS.

In addition to this analysis, I will present the coding of diminutive suffixes used for derivation of other meanings in order to check whether the polyfunctionality of Serbian diminutive suffixes is another factor that may affect the diminutive production.

3.4.1. Analyses by age and gender of children

In their analysis of Peruvian mothers' production of diminutives, Melzi & King (2003) argued that children's age and gender might affect the overall distribution of diminutives in CDS. Previous quantitative cross-linguistic analyses of the distribution of diminutives in CDS showed that the diminutive usage reaches its peak around the children's second birthday, and that it starts decreasing between the age of three and five which should coincide with the children's full mastery of grammatical categories (Kempe et al, 2001; De Marco, 1998; Savickienė, 1998; Melzi & King, 2003, 2004; Stephany, 1997; Gleason et al., 1994).

The question as to whether the gender of the child affects diminutive usage in CDS was addressed in only two studies which obtained opposite results. For English, Gleason et al. (1994) showed that parents of two-year old girls tended to use more diminutives in contrast to parents of boys. This observed difference was explained by the fact that diminutives usage is not only related to CDS (Ferguson, 1978), but also to female speech in general (Daltas, 1985; Andrews, 1999, cited in Melzi & King, 2003). On the other hand, Melzi & King (2003) did not find a significant difference when Spanish speaking mothers were addressing their daughters in contrast to their sons.

In order to determine whether parental diminutive, frozen diminutive and hypocoristic usage in Serbian CDS changes as the children get older, I performed an analysis of the separate FREQ lists for the common nouns which were based on the combined utterances of mothers and fathers addressing a child across five age samples (40 lists all together). In addition to the analysis by age, I performed the same set of analyses for the boys and girls separately in order to determine whether the gender of children plays a role in the production of diminutives in Serbian CDS.

Overall, the mean percentage of dimutives calculated over 40 differenet samples was similar to the one observed for the cumulative frequency list (5.9% vs. 7%), and it ranged from 0% to 13.6% per sample. The mean percentage of diminutive, frozen diminutive and hypocoristic common noun tokens and lemmata, as well as the mean percentages for girls and boys separately are presented in Table 3.2 and Figures 3.10.-3.12 (and in more details in Tables A3.3.-A3.12 in Appendix 3).

Table 3.2. The mean percentage of diminutive, frozen diminutive and hypocoristic tokens and lemmata (with standard deviations in parentheses) for entire sample, as well as for boys and girls separately.					
Diminutives		**Frozen diminutives**		**Hypocoristics**	
Tokens	Lemmata	Tokens	Lemmata	Tokens	Lemmata
5.9 (3.4)	8 (3.7)	1.9 (1.0)	2.0 (0.1)	14.9 (5.2)	6 (1.8)
BOYS (N=4)					
4.1 (2.1)	6.2 (2.8)	1.8 (1.3)	1.8 (0.8)	13.5 (4.7)	6.1 (2.1)
GIRLS (N=4)					
7.7 (3.7)	9.7 (4.0)	1.9 (0.8)	2.4 (0.7)	16.4 (6.0)	6.0 (1.1)

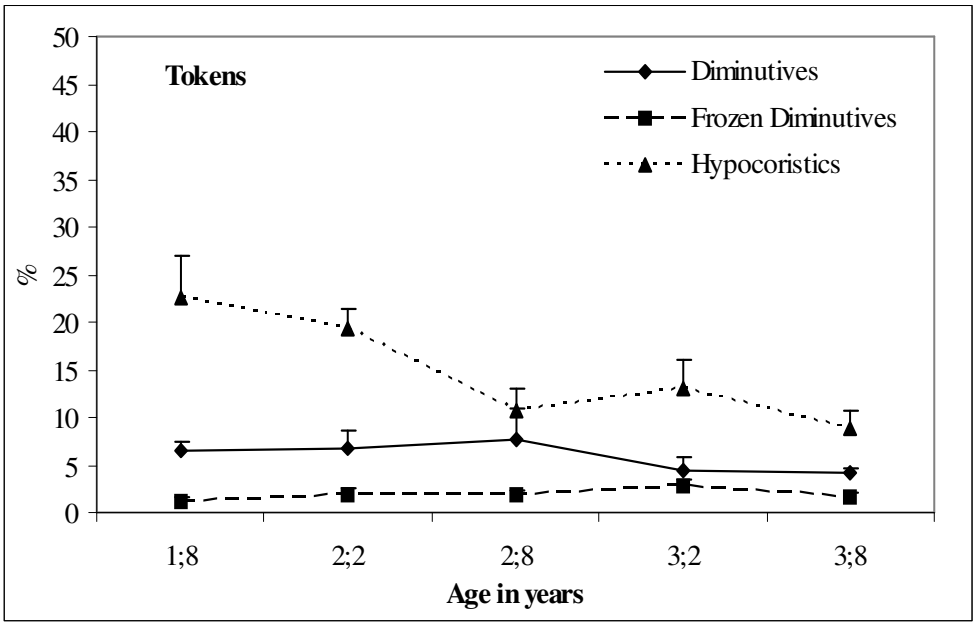
For the tokens, a (3) noun derivation: diminutive vs. frozen diminutive vs. hypocoristics x (5) (age: 1;8 vs. 2;2 vs. 2;8 vs. 3;2 vs. 3;8) as within-subjects factors and (2) child gender: boys vs. girls as between-subjects factor ANOVA revealed a main effect of derivation, $F(2,12) = 27.7$, $p<0.001$, $\eta^2=0.47$. Post-hoc tests using Fishers protected LSD revealed that parents used significantly more hypocoristics than 'frozen' diminutives, $t(7)= 6.5$, $p<0.05$ and real diminutives, $t(7)= 4.3$, $p<0.05$ and more diminutives than frozen diminutives, $t(7)= 3.7$, $p<0.05$. The analysis also showed a main effect of age, $F(4,24) = 4.0$, $p<0.05$, $\eta^2=0.06$ which indicated that the usage of terms expressing affection and endearment was significantly decreasing over age. Post-hoc tests using Fishers protected LSD revealed that this effect was mainly carried by a significant difference between the samples when the children were 1;8 and 3;8 years old, $t(7)= 3.5$, $p<0.05$, and 2;2 and 3;8 years old, $t(7)= 4.5$, $p<0.05$. The analysis also showed a significant two-way interaction between derivation and age, $F(8,48) = 4.5$, $p<0.001$, $\eta^2=0.1$. The ANOVA's for the three derivations separately with age (5 groups) as within-subjects factor revealed that the two-way interaction between derivation and age was carried by the hypocoristics, $F(4,24) = 5.3$, $p<0.05$, $\eta^2=0.45$, indicating that the significant decrease over five age groups was only present

for this derivation. Other interactions, as well as the main effect of child gender, were not significant, indicating that at least for this sample, parents addressed boys and girls with similar numbers of diminutive, hypocoristic and 'frozen' diminutive tokens. Similar effects were observed for the lemmata, where a 3 (noun derivation: diminutive vs. frozen diminutive vs. hypocoristics) x 5 (age: 1;8 vs. 2;2 vs. 2;8 vs. 3;2 vs. 3;8) within-subjects and (2) child gender: boys vs. girls as between-subjects factor ANOVA revealed a main effect of derivation, $F(2,14) = 11.0$, $p<0.05$, $\eta^2=0.34$. Post-hoc tests using Fishers protected LSD revealed that parents used significantly more hypocoristics than 'frozen' diminutives, $t(7)= 4.9$, $p<0.01$, more diminutives than frozen diminutives $t(7)= 4.2$, $p<0.01$, but with no significant difference between diminutive and hypocoristic lemmata. The analysis also yielded a main effect of age, $F(4,28) = 2.9$, $p<0.05$, $\eta^2=0.06$ which indicated that the number of lemmata expressing affection and endearment was significantly decreasing over age. Post-hoc tests using Fishers protected LSD revealed that this effect was carried by the difference between the samples for 1;8 and 3;2 years, $t(7)=2.9$, $p<0.01$; samples for the 1;8 and 3;8 years, $t(7)=3.4$, $p<0.01$ and samples for the 2;2 and 3;8 years, $t(7)=2.9$, $p<0.01$.

The analysis of the lemmata also showed a significant two-way interaction between derivation and age, $F(8,48) = 2.8$, $p<0.05$, $\eta^2=0.07$. The ANOVA's for the three derivations separately with age (5 groups) as within-subjects factor, revealed that the two-way interaction between derivation and age was carried by the hypocoristics, $F(4,24) = 4.7$, $p<0.01$, $\eta^2=0.4$. Other interactions were not significant. The main effect of child gender was not significant as well, confirming that at least for this sample, parents used a similar number of lemmata for the production of terms of endearment and affection when addressing boys and girls.

In sum, these analyses showed that Serbian CDS is consistent with previous data on the decrease of terms for endearment and affection in CDS as children get older. This effect was mainly carried by hypocoristics which is the most prominent derivation in Serbian CDS. The lack of a significant effect of age for diminutives might be due to several factors like the overall low frequency observed for all age samples, but also because a significant decrease in diminutives might happen in the later stages of development, around the age of five (as the Peruvian mothers, Melzi & King (2003)), for which Serbian data could not be obtained.

With respect to the children's gender as a potential factor which may influence diminutive usage in this sample of Serbian CDS, the results did not show a difference between parental strategies in addressing boys and girls. Of course this result should be treated with caution due to the relatively small number of subjects. In order for us to establish whether there is a difference in addressing boys and girls in Serbian CDS in general, we need to extend our analyses not only to a broader age span and more parent-child interactions, together with different conversational situations both between parents and their children, but also between children and their peers, other family members or adults from outside the family.



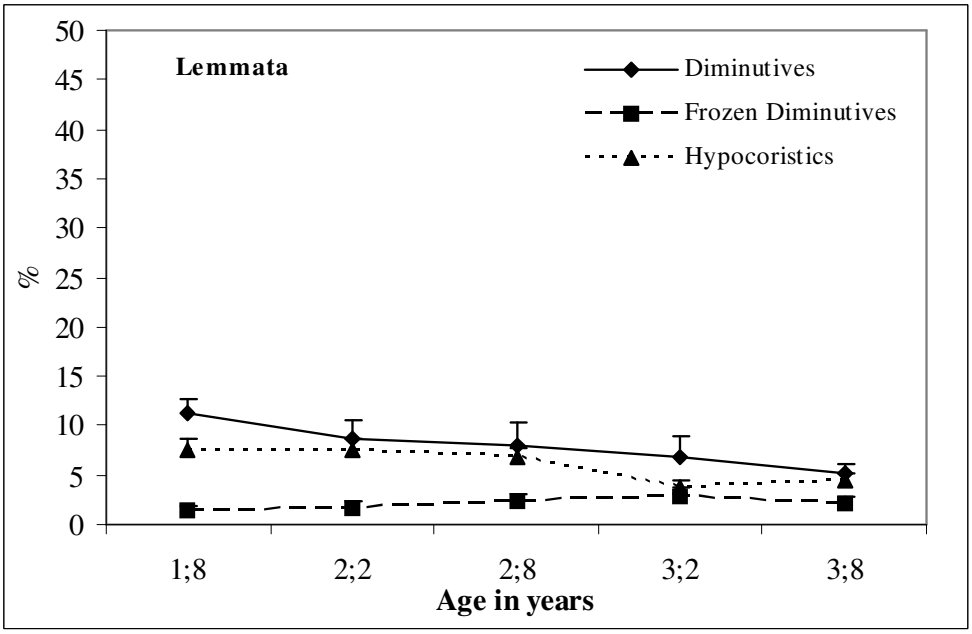
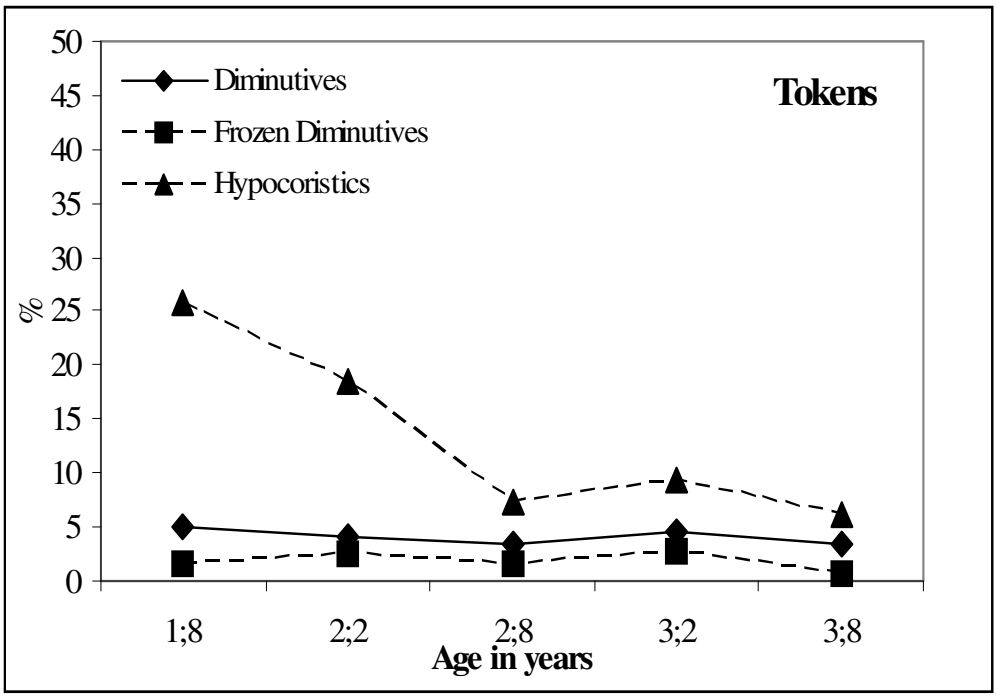


Figure 3.10. Mean percentage (and 1 S.E.M.) of diminutive, frozen diminutive and hypocoristic common noun tokens (upper panel) and lemmata (lower panel) across five ages.



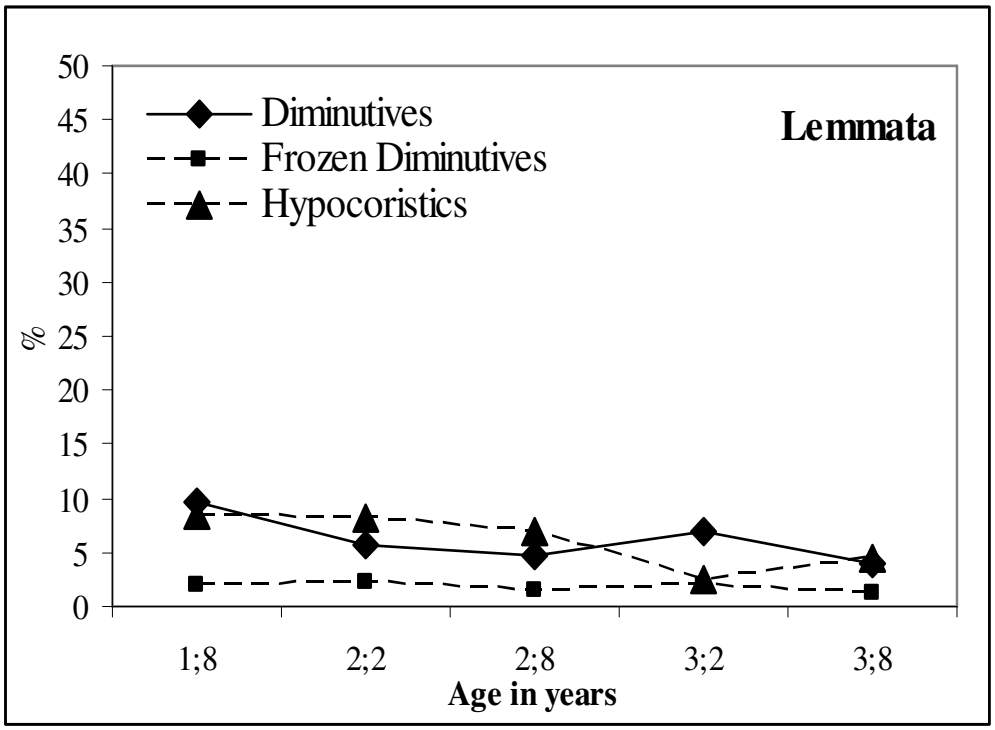


Figure 3.11. Distribution of diminutive, frozen diminutive and hypocoristic common noun tokens (upper panel) and lemmata (lower panel) over five age groups, for the boys.

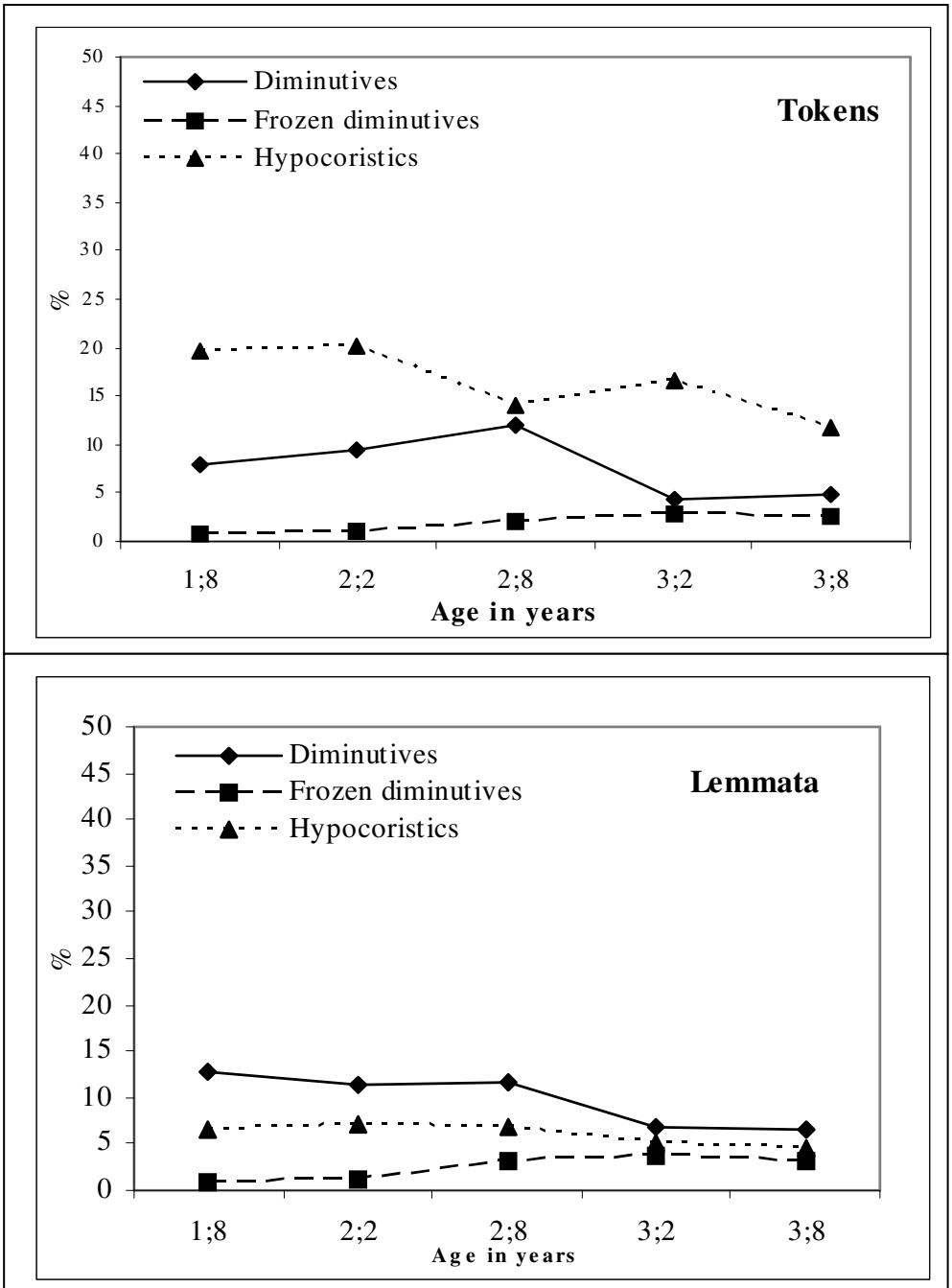


Figure 3.12. Distribution of diminutive, frozen diminutive and hypocoristic common noun tokens (upper panel) and lemmata (lower panel) over five age groups, for the girls.

3.4.2. Differences between mothers and fathers

Previous estimates of the diminutive distribution in CDS for other languages were mainly based on the utterances of mothers addressing their children. The only study contrasting diminutive production of mothers and fathers was Gleason et al. (1994) which showed a slight but not significantly more frequent use of diminutives for English speaking mothers in contrast to the diminutive usage of fathers. Given this lack of data, it would be interesting to see whether Serbian mothers and fathers differ in the production of terms for affection and endearment and whether this potential difference may have influenced the overall low frequency of diminutives in Serbian CDS.

In order to determine the frequency of diminutives and hypocoristics in Serbian CDS for mothers and fathers separately, utterances of mother B.G. and father A.G. of the boy M.G. were analysed when the child was 1;8, 2;2, 2;8, 3;2 and 3;8 years old. The analyses were constrained to this couple of parents due to the high discrepancy in the overall amount of data avalible for mothers and fathers in the other couples (see the description at the beginning of the chapter). This resulted in having only one pair of parents which participated in all sessions to the same extent.

The overall cumulative output of common nouns for this parental couple was very similar, with 1,333 tokens and 433 lemmata for the mother vs. 1,192 tokens and 367 lemmata for the father. Lemma/token ratio was around 0.3 for both parents, showing that both parents had a similarly rich lexicon.

The cumulative distribution of all derivational classes is depicted in Figure 3.13, and it shows that this parental couple produced a similar percentage of diminutives and hypocoristics both at the level of tokens and lemmata. Figures 3.14.

and 3.15. represent the parents' usage of endearment forms over five age samples which indicates that both the father and the mother were decreasing the number of diminutives and hypocoristics as their son was getting older. For more detailed description of distribution of derivational, gender and declension classes for the mother and the father see Tables A3.13.-A3.16. in Appendix 3. Moreover, the distributional patterns observed for this parental couple are very similar to the overall distribution in the cumulative sample for the eight couples of mothers and fathers, presented in the first part of this chapter.

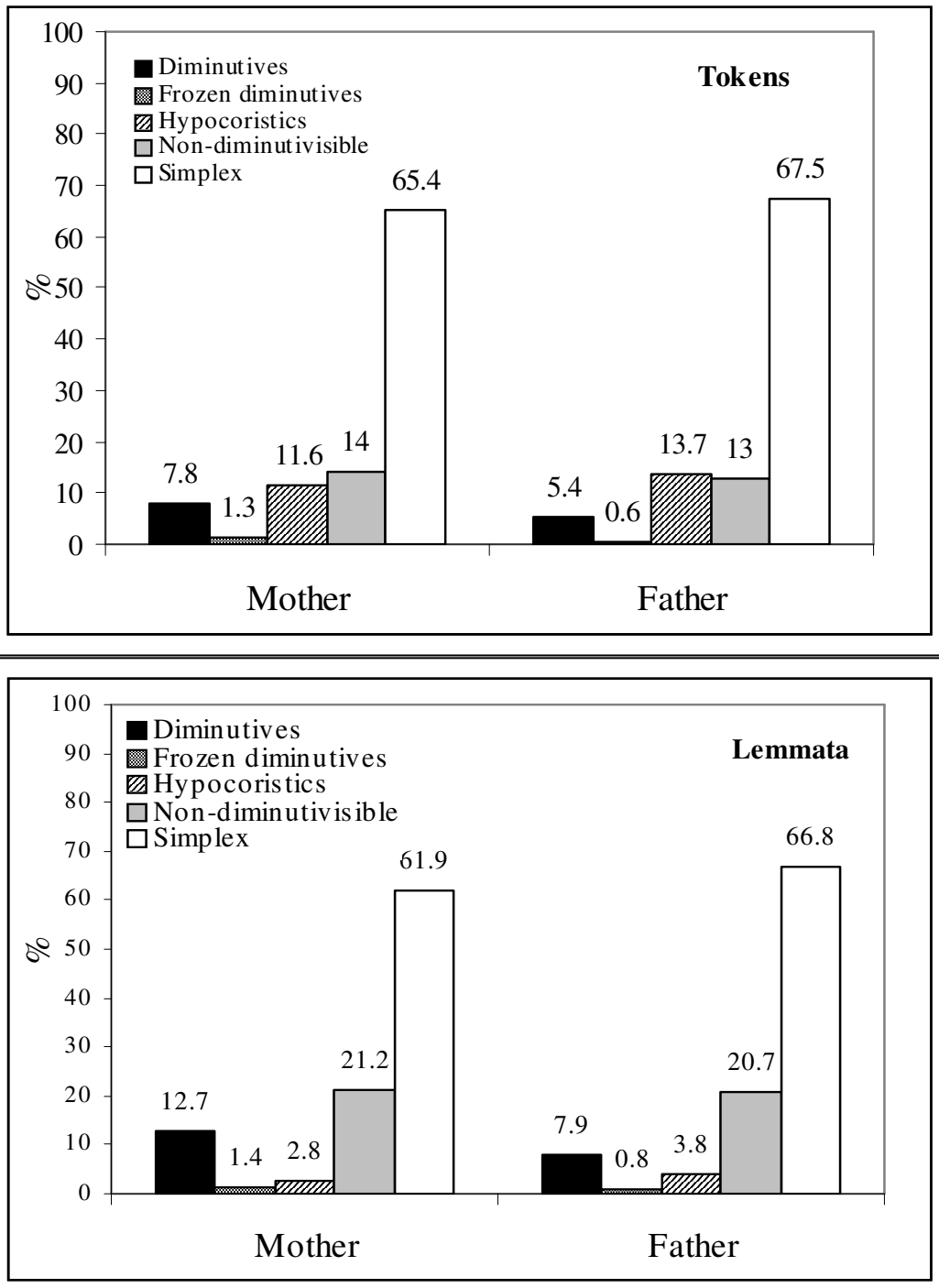
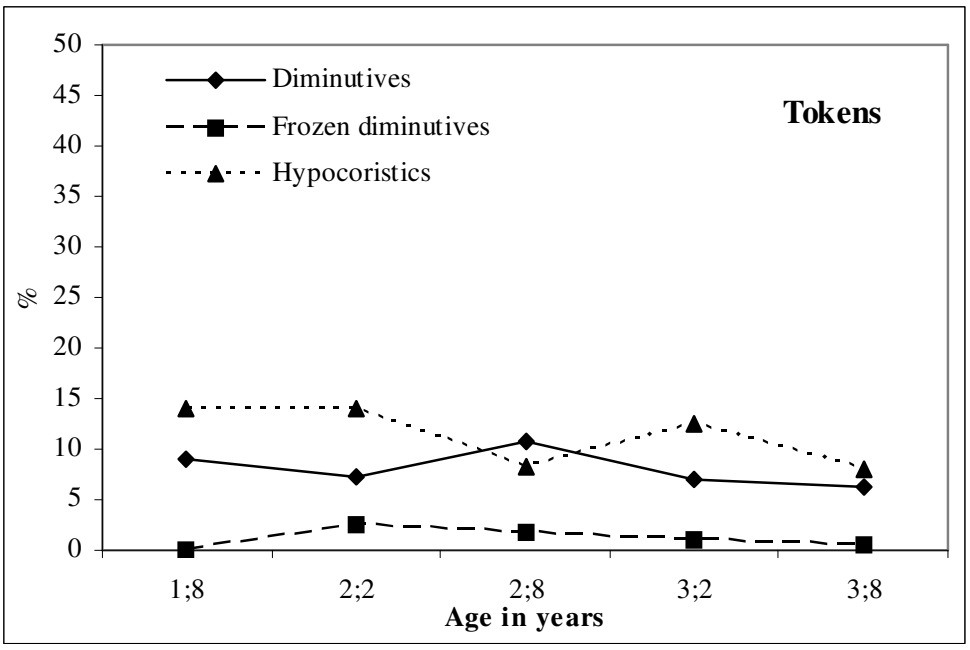


Figure 3.13. Overall distribution of derivational classes for common noun tokens (upper panel) and lemmata (lower panel), for mothers and fathers separately.



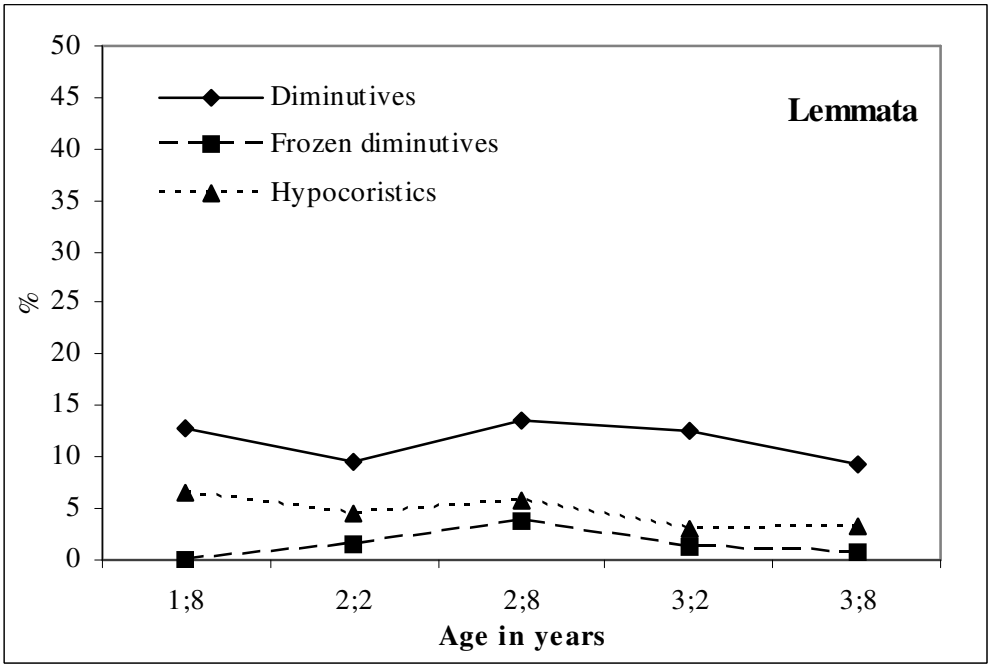
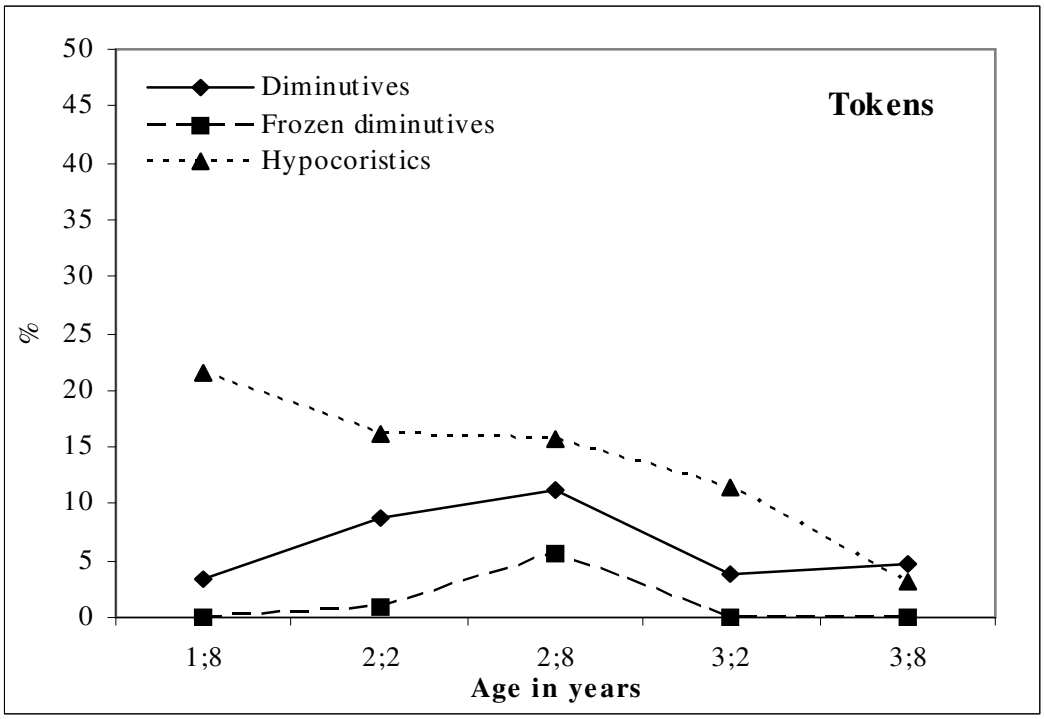


Figure 3.14. Distribution of diminutive, frozen diminutive and hypocoristic common noun tokens (upper panel) and lemmata (lower panel), across five age samples for the mother.



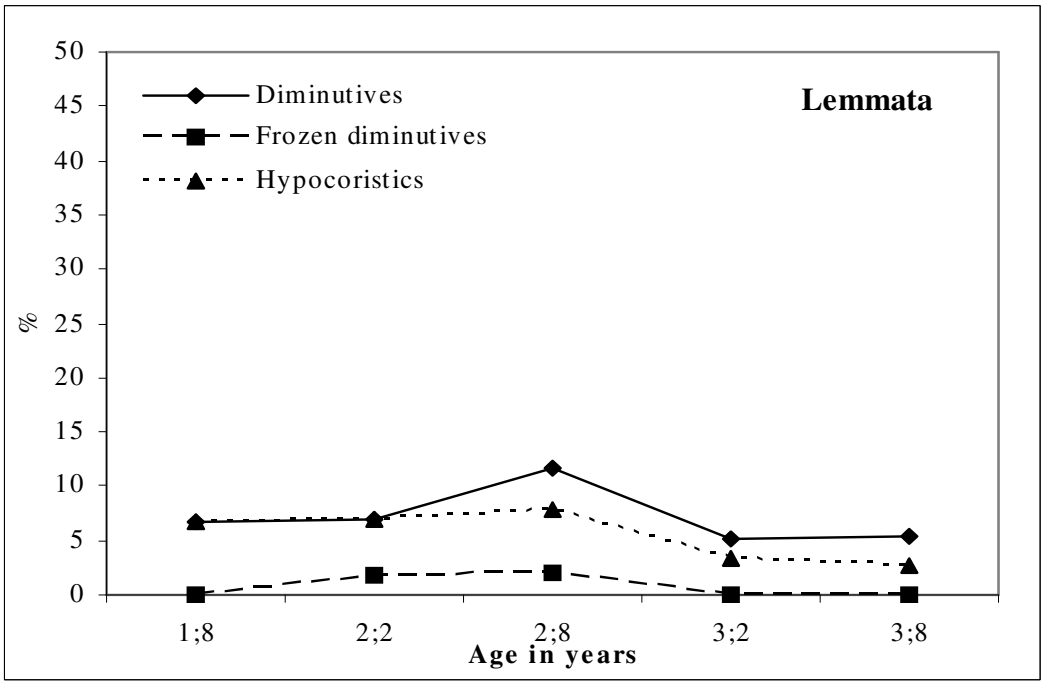


Figure 3.15. Distribution of diminutive, frozen diminutive and hypocoristic common noun tokens (upper panel) and lemmata (lower panel), across five age samples for the father.

3.4.3. Diminutive productivity and poly-functionality of diminutive suffixes in Serbian child-directed speech, adult-directed speech and written language

The description of morphological characteristics of Serbian diminutive suffixes provided at the beginning of this chapter showed that these suffixes (mainly *–ica* and *–ić*) can be used for the derivation of other meanings. This difference between Serbian and other Slavic languages potentially makes the diminutive suffixes in Serbian phonologically and semantically less prominent candidate for the expression of affection and endearment, because these suffixes are therefore functionally ambiguous. As a consequence this ambiguity may contribute to the low-frequency of diminutive derivations in Serbian CDS.

In this part of the chapter, I will provide a quantitative description of poly-functional usage of diminutive suffixes in Serbian CDS, ADS and written language in order to see whether the frequency of this cluster of words will increase if we include instances where diminutive suffixes were used for the derivation of other meanings. This group of suffixes will henceforth be called the 'diminutive-like' suffixes. For the CDS register, I have coded all nouns in cumulative frequency lists for each parental couple across the five age samples (40 samples). This count of poly-functional suffixes showed that the average number of words ending like diminutives doubles in the parental utterances both for tokens and lemmata (tokens: 7% of diminutives vs. 14% of all diminutive-like suffixes together, and lemmata: 11.7% of diminutives vs. 18% of all diminutive-like suffixes together). Figure 3.16. depicts the distribution of the cumulative percentage of all diminutives-like suffixes in Serbian CDS.

Detailed results for the cumulative frequency of diminutive and diminutive-like suffixes across different derivational classes for Serbian CDS are presented in Table A3.21. in Appendix 3.

A similar distribution of poly-functional suffixes was observed for ADS and written language, where the cumulative percentage of all diminutive-like suffixes increased up to 14-16% for noun tokens and up to 12% for noun lemmata which is also depicted in Figure 3.16.

Detailed results for the cumulative frequency of all diminutive-like suffixes across different derivational classes for Serbian ADS and written language are presented in Tables A3.22-A3.23 in Appendix 3.

This quantitative description showed that diminutive suffixes are used for the diminutive derivation as productively as for the derivation of any other meaning in all three registers which indicates that polyfunctionality of the suffix might play a role in the relative low frequency of diminutives in Serbian CDS in comparison to other Slavic CDS's.

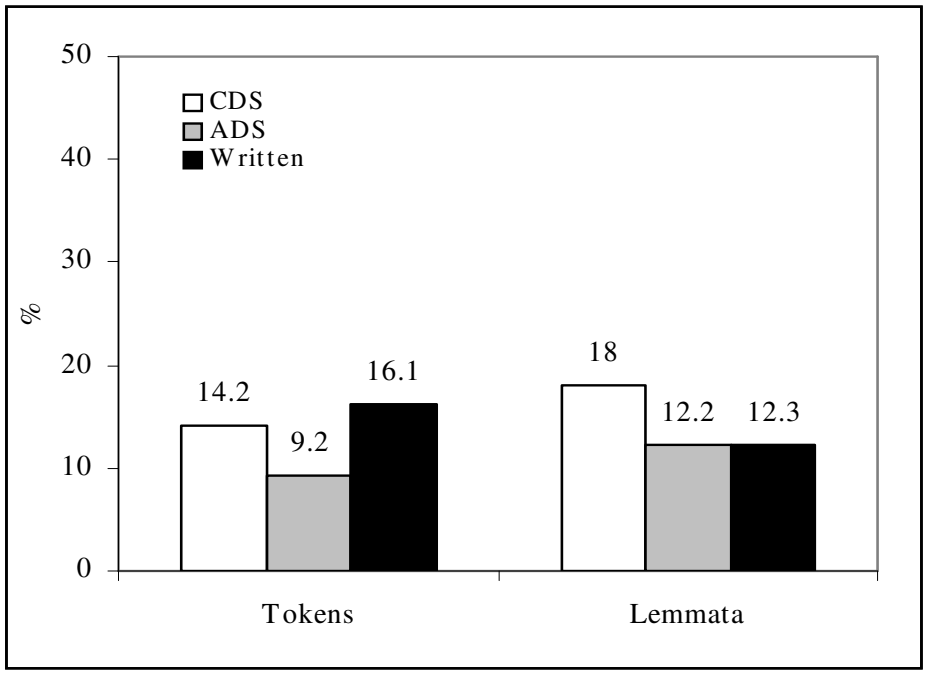


Figure 3.16. Cumulative percentage of all diminutive-like suffixes across three registers: CDS, ADS and written language.

3.5. General discussion of the results

In this chapter, I presented the distribution of diminutives in Serbian CDS. The corpus analyses of Serbian CDS showed that Serbian parents produced around 7% of diminutive tokens and 11% of diminutive lemmata out of all common nouns. This percentage was considerably lower than the frequency observed for Russian, Polish and Lithuanian CDS's with 20-50% of nouns in diminutive form. Nevertheless, Serbian parents increased their diminutive usage ten times more in comparison to Serbian ADS and written language, with less than 1% of common noun tokens and 1-2% of noun lemmata in ADS and written language used in diminutive form. Diminutive usage did not decrease as children got older within an age range of 1;8 to 3;8 years. Effect of the age was observed only with the hypocoristics. Also, the gender of the children and the gender of the adult interlocutors (mother vs. father) did not affect the overall diminutive frequency. Similarly to Russian, Serbian parents also used about 5% of noun stems in word-play alternations of simplex and diminutive forms. A similar observation has been reported for Lithuanian (Savickienė, 1998).

In addition to this analysis, I coded nouns for their gender and declension classes, in order to see whether: a) the distribution of diminutives is different for different grammatical categories and b) there are some distributional changes in gender and declension classes in general. This analysis is important as the observed distributional shifts in CDS may influence children's acquisition of gender and case marking which will be experimentaly tested in the later chapters of this thesis.

For diminutives, the analysis showed that in all three registers, diminutives were commonly derived from feminine nouns, indicating that the *–ica* suffix is the most frequent diminutive suffix in Serbian. Furthermore, Serbian CDS exhibited distributional changes for both gender and declension classes in comparison to ADS and written language. For gender classes, parents tended to repeat more feminine nouns, with 10% more noun tokens in feminine gender in the CDS register. The distribution of declension classes was also shifted towards the feminine declension, with approximately 60% of noun tokens requiring the feminine declension in CDS in contrast to 40-45% in ADS and written language.

What remains open is the question as to what accounts for the difference in diminutive frequency between Serbian CDS and the other morphologically similar languages, like Russian and Polish. Without a detailed cross-linguistic comparison of patterns of productivity and diachronic development, I can only speculate on potential reasons for the observed discrepancy in diminutive production in CDS. This difference may be due to a combination of several factors including cultural and language specific differences. Below I will describe four potential factors.

1. One possible explanation might be related to the fact that Serbian has other derivational forms suitable for connoting affection and endearment, like the hypocoristics. Although this derivation is not as productive as the diminutives, since hypocoristics are formed mainly for proper nouns, some animals and kinship terms, they still provide a clear marker of endearment which is frequent in CDS, especially because animals and kinship terms are typical child-directed conversational domains. Indeed, the corpus analysis revealed that hypocoristics are used significantly more frequently than diminutives, with approximately 18% of common noun tokens produced as hypocoristics. Hypocoristics are used also in Russian

CDS, but much less frequently than in Serbian CDS, with only 3% of common noun tokens (Ševa et al, in press), indicating that for some reasons Serbian parents prefer to use this derivation more than diminutives when addressing their children.

2. The low frequency of diminutives in Serbian CDS might also be related to the polyfunctionality of Serbian diminutive morphemes as described above. It is possible that this feature makes Serbian diminutives less prominent candidates for the expression of affection and endearment, given that suffixes like the feminine diminutive suffix *–ica* may also be used for the derivation of other meanings. The corpus analysis showed that the suffix *–ica* is used as productively for other derivations as for diminutive formation, indicating that this factor might play a significant role in the low frequency of diminutives in CDS.

3. The third possible reason for the difference between Serbian CDS and other Slavic CDS registers is related to the question as to which words drive the imitative reinforcement of diminutive usage which progressively builds up to the high percentage of diminutives in CDS? Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press argue that one of the reasons for diminutive usage in Russian CDS is the disambiguation of opaque feminine nouns ending in a palatalised consonant which in their diminutive form end in the transparent feminine ending *–ka.* Some of these words like *mysh* FEM 'mouse (FEM)' vs. *myshka* 'mouseDIM (FEM)', are semantically and pragmatically closely related to CDS and almost exclusively used in diminutive form in this register (Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press; Seva et al., in press). This set of 'frozen' word forms might serve as a base or critical mass for the later imitative reinforcement of the diminutive usage between parents and children which ultimately results in 40-50% of diminutives in Russian CDS. In Serbian, on the other hand, most of the

concrete feminine nouns that are opaque in Russian are transparently marked masculine nouns, e.g. Russian *mysh* 'mouse (FEM)' vs. Serbian *miš* 'mouse (MASC)'. This language difference, plus the fact that Serbian feminine opaque nouns are practically not used in Serbian CDS (less than 1%), might be another reason for the lack of a critical mass of 'frozen' word forms in Serbian CDS, necessary to boost the high frequency of diminutive forms. Moreover, in Serbian this set of 'frozen' word forms comprises mainly hypocoristics, like *meda* 'bearHYP', *zeka* 'rabitHYP', etc. The high frequency of hypocoristics in Serbian CDS indicates that parents probably enhanced the hypocoristic usage through imitative mechanisms described by King & Melzi (2003, 2004). However, given the relatively limited number of lemmata which can be put in this form, hypocoristic usage in Serbian CDS cannot reach a frequency similar to the high frequency of diminutives in Russian CDS.

4. I also cannot exclude the possibility that there may be discrepancies in: a) the reliability and b) the representativeness of the speech samples underlying our estimates.

a) The *reliability* of the data is related to the difference in the sample size, where estimates for Serbian CDS were based on a roughly ten times bigger sample than Russian estimates. In order to check whether the number of diminutives decreases with an increase in the number of words, I counted the diminutives in samples of the same size like the samples used in the Russian corpus study (Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press). For these reduced samples, estimates were based on the first 100 nouns produced by four mothers in conversations with their children, two boys and two girls, at the same ages (20 and 34 months of age). The results are depicted in Table 3.3., and are showing that the

observed frequency for the smaller sample is in the same range as the frequency of diminutives in the bigger sample for Serbian. This lends more credibility to my claim that the frequency of diminutives differs dramatically between Russian and Serbian CDs, with an average of 45% noun tokens and 50% of noun lemmata in Russian and 7% noun tokens and 10% of noun lemmata in Serbian. Given that the suffixes used for diminutivisation in Serbian can also carry a different function, I have counted the occurrence of all diminutive-like suffixes in the reduced corpus, and obtained a frequency of 10.5% of noun tokens and 15% of noun lemmata which was still below the frequency of diminutives in Russian.

Thus, the differences in frequency of diminutives observed for those two languages are not due to the different size of samples I have used in this study.

Table 3.3. Percentage of diminutives and all diminutive-like suffixes in CDS as function of child age (standard deviations are given in parentheses).

	Percentage of diminutives in CDS	
	20 months	34 months
	Russian	
Noun lemmata	58.6 (11.3)	41.4 (9.2)
Noun tokens	54.7 (19.3)	35.0 (6.9)
	Serbian	
Noun lemmata	9.8 (4.8)	9.8 (3.6)
Noun tokens	7.8 (5.4)	7.0 (3.4)
	Serbian (all diminutive-like suffixes)	
Noun lemmata	13.2 (14.7)	16.8 (5.2)
Noun tokens	9.8 (5.1)	11.3 (5.6)

b) The criterion of *representativeness* is related to the type of material included in the analysis. The estimates both for Serbian and Russian CDS stem from a limited number of mothers speaking in slightly different situations. Previous research of the nature of parent-child interactions showed qualitative and quantitative features of interactions being dependent on the conversational situation and presence of other interlocutors (Hoff-Ginsberg, 1991; Lanvers, 2004). Here, the Russian mothers audio-recorded their interactions in the absence of a third person (Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press) whereas the Serbian mother-child interactions were video-taped (Anđelković et al., 2001) which made the presence of another adult necessary. This may have slightly discouraged the Serbian mothers from full use of CDS and, thus, inflated the cross-linguistic differences in diminutive frequency. These methodological difficulties underscore the importance to control for the representativeness of corpus data outside of the corpora itself, for example through the use of experimental methods for the elicitation of diminutive usage in addressing an (imaginary) child under controlled laboratory conditions. Such elicited production of CDS in controlled experimental conditions will also provide the opportunity to determine whether the observed cross-linguistic differences in the frequency of diminutives are statistically significant.

Controlled laboratory elicitation of CDS was tried in a first pilot study of this kind (Ševa, Hadjiconstantinou, Kempe, 2005) in which 46 native adult speakers of Serbian and Greek were tested. Greek was contrasted to Serbian because we could obtain a large enough sample of native Greek participants at the University of Stirling. In addition to this purely technical reason, Greek was also interesting because it is a typologically different language from Serbian and those two

languages have a different diminutive distribution in CDS, with Serbian having a low frequency (7%) and Greek a high frequency (32%, Stephany, 1997). Participants performed a simple Map Task (Brown, Anderson, Shillcock, & Yule, 1984) addressing an imaginary child and an imaginary adult, to test diminutive production under controlled conditions (see Appendix 4 for more details). The results confirmed a significant increase of diminutive usage in CDS for both languages. For Serbian, the mean percentage of diminutive nouns was 2.9% in ADS vs. 10.3% in CDS. For Greek, the mean percentage of diminutive nouns was 7.9% in ADS and 37.3% in CDS. Note that the obtained percentages of diminutives in CDS were very close to estimates from previous corpus analyses (32% for Greek, Stephany, 1997 and 7% for Serbian).

Thus, using controlled elicitation, this study was able to replicate the cross-linguistic differences in frequency of diminutives suggesting that the corpus analysis presented in this thesis represents reliable estimates (see Appendix 4 for more details).

In sum, in order to determine which of the above listed factors is responsible for the variability in diminutive production across languages, we need further empirical cross-linguistic testing.

Having established that Serbian CDS shows a reduced frequency of diminutives compared to Russian, Polish and Lithuanian, I will now proceed to examine the effect of these observations on the acquisition of noun morphology. Serbian seems to be the perfect candidate for those analyses, given the overall morphological similarities between all Slavic languages and a vast discrepancy in diminutive production in Serbian CDS and Russian and Polish CDS. Specifically, I will test whether a similar diminutive advantage as observed for Russian, Polish and

Lithuanian can be shown for Serbian children, or whether the facilitating effect of diminutives is reduced in this language. These questions will be addressed in further experimental studies with Serbian children, described in more detail in Chapter 4.

Chapter 4

4. Experimental studies with Serbian children on the facilitating effect of diminutives in the acquisition of noun gender and case marking

The previous studies, described in more detail in the Introduction, have shown that diminutives facilitate noun morphology acquisition in languages with complex morphological systems like Russian, Polish or Lithuanian. The beneficial effect of diminutives for those languages was attributed to the high frequency of diminutives in the input and the consequent decrease of the morpho-phonological complexity at the ends of words.

This chapter will address the question as to whether a diminutive advantage for gender agreement and case marking in novel nouns will be observed in Serbian, a language morphologically similar to Russian and Polish, but with considerably lower frequency of diminutives in CDS. The Russian gender (Kempe et al., 2003; Ševa et al., in press) and case marking (Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press; Kempe et al. in preparation) studies will be replicated with Serbian children, in order to look in greater detail at whether frequency is the main factor responsible for the diminutive advantage obtained for Russian and Polish. If the diminutive advantage observed in Russian and Polish is predominately due to the high frequency of diminutives in CDS, the potential benefit of diminutives should not be observed or should at least be attenuated in Serbian.

Additionally, the experimental studies with Serbian children represent the first experiments conducted in this language where three year old children were tested for their knowledge of morphology in a production task. In previous research on the acquisition of the Serbian noun system (Johnston & Slobin, 1979; Jocić,

1980; Mimica, 1988; Kostić & Vidosavljević, 1995) which tested children's performance in comprehension based tasks, it was shown that acquisition of noun inflectional morphology starts between 1st and 2nd year and ends between 3rd and 4th year. In contrast to that, experiments with the production task of familiar and novel nouns examine to what extent children are really capable of generalising grammatical knowledge to novel items.

4.1. Gender agreement in Serbian children

As described in Chapter 2, the gender of nouns in Serbian and other Slavic languages is marked by the appropriate suffixes at the ends of nouns, but also by endings of pronominal words or past tense forms of verbs which always have to agree in gender, case or number with the noun. This feature of the Slavic languages was used in designing the experimental procedure for Kempe et al. (2003), where gender agreement was elicited with several questions to describe familiar and novel animal names with pronominal words and past tense forms of verbs. Responses in which the children produced the appropriate agreement between nouns and pronominal words/past tense verbs, were counted as indicators for correct gender categorisation of the presented nouns. In a replication with Russian children, Ševa et al. (in press) used a similar methodology, but slightly modified the procedure so that only adjective responses were elicited from the children. This change of the procedure took place mainly because results from Kempe et al., (2003) suggested that children produce fewer agreement errors with pronouns or verbs, but also because more constrained tasks allow for easier cross-linguistic comparisons. The tasks used in Kempe et al. (2003) and Ševa et al. (in press) were both an adapted

version of the 'wug' test (Berko-Gleason, 1958), where children's grammatical knowledge was tested with novel nouns presented in familiar context. For example, in Ševa et al. (in press), children were first introduced to one familiar noun and the adjectives which had to be used for the description of nouns and implicit gender categorisation, e.g. *eto pauk. pauk – khoroshij ili plokhoj?* 'This is spiderNOM. Is spiderNOM goodMASC or badMASC?' After the children had given their response, the experimenter presented the first test picture of a familiar or a novel animal, labelled with a familiar or a novel noun, for example, *slon* FAMILIAR 'elephant' vs. *krufa* NOVEL, accompanied by the utterance: *eto slon/krufa. pauk khoroshij. a slon/krufa?* 'This is elephant/krufaNOM. spiderNOM is goodMASC. And what about elephant/krufaNOM?'

A similar procedure will be used for testing noun-adjective agreement marking in Serbian in order to directly compare Serbian and Russian children.

4.1.1. Method

Participants: 22 children (mean age 3;7, range from 3;0 to 4;1) were tested in various kindergartens in the Belgrade region. All children were acquiring the Belgrade variety of Serbian. An additional 3 children were tested but excluded because they did not complete the task.

Materials: Sixteen colour photographs of familiar animals and 16 colour photographs of unfamiliar animals were selected from *Faszination Tier & Natur* published continuously by Meister Verlag GmbH, München, IMP B.V. The unfamiliar animals were selected for their unusual appearance making sure that their real habitat was distant from Europe. Eight of the nouns denoting the familiar

animals were masculine, and eight were feminine. In addition, I created 16 Serbian pseudo-word labels for the unfamiliar animals. Eight of the novel names for the unfamiliar animals ended in a consonant resembling the dominant word form of Serbian masculine nouns, and eight ended in the suffix *-a*, resembling the dominant form of Serbian feminine nouns. All 32 nouns were transparently marked for gender. All nouns were diminutivised for presentation in the diminutive condition. No neuter nouns were included as it is impossible to find a matching number of Serbian neuter nouns denoting animals. Thirteen pictures from the Serbian set were identical to the pictures used with Russian children in Ševa et al., in press. Of the identical pictures, ten Serbian nouns had the same gender as their Russian translations. Eight of the nouns denoting familiar animals were masculine, and eight were feminine. The Serbian pseudo-nouns were identical to the Russian pseudo-nouns or as similar to their Russian counterparts as Serbian phonotactics permits (e.g. *farzjak* (Russian) vs. *farzak* (Serbian)). In addition, I selected four familiar animals for practice, and four more to provide examples of the adjective production template (see Procedure for details). All nouns and their diminutive derivations are listed in Table A5.1. in Appendix 5.

The derivational status of nouns (simplex vs. diminutive) and noun gender (masculine vs. feminine) were varied as within-subject factors. The 16 familiar and 16 unfamiliar nouns were distributed across two lists in such a way that each noun appeared as simplex in one list, and as diminutive in the other. Each list contained an equal number of simplex and diminutive, familiar and unfamiliar nouns. Half of the children were presented with list 1, and the other half with list 2. Children were quasi-randomly assigned to the two lists of 32 items, matching for sex and age. Items from each list were randomly assigned to four blocks of 8 trials.

I also selected two antonymous adjective pairs that were used to prompt the children to talk about the animals. These pairs were *dobar-loš* (masc.) vs. *dobra-loša* (fem.) 'good-bad' and *lep-ružan* (masc.) vs. *lepa-ružna* (fem.) 'beautiful-ugly'. The adjective endings served as indicators for correct or erroneous gender agreement.

Procedure: Each child was tested individually by a female native speaker of Serbian[14] in a room adjacent to the main activity room of the kindergarten. The entire procedure took about 20 minutes to complete.

The experiment comprised three phases: (1) a Practice phase, to engage the child in labelling and describing the animals; (2) a Template phase, to introduce a specific pair of adjectives to be used to describe the subsequently presented 8 test items; (3) a Test phase, to elicit use of gender-marked adjectives as descriptions of animals (i.e., adjective or adjective-noun production).

The children were first shown the four practice pictures depicting familiar animals, labelled by the experimenter. The children were instructed to repeat the labels. The experimenter then provided a simple statement about the animal like *Medved je velik. Ponovi*. 'BearNOM is big. Can you repeat this?'. The children were then shown one template picture and told *Ovo je pauk. Je li pauk dobar ili loš*? 'This is spiderNOM. Is spiderNOM goodMASC or badMASC?'. After the child answered the question, the experimenter presented the first target picture, accompanied by the utterance: *Ovo je slon. Pauk je dobar. A slon?* 'This is elephantNOM. SpiderNOM is goodMASC. And what about elephantNOM?'.

This elicitation form avoided the experimenter's use of gender agreement, and gave the children the opportunity to pick one of the members of the adjective

[14] The author of the thesis was the experimentator in all three studies presented in following parts of the thesis.

pair. The same adjective pair was used for eight consecutive test nouns, after which the experimenter introduced a new template noun, along with the other antonymous adjective pair. This procedure of introducing a template picture (with one of the two adjective pairs) followed by eight test trials was repeated four times for a total of 32 test trials. Alternation of adjective pairs and order of template gender were counterbalanced across participants. Instances of erroneous gender agreement as reflected in the adjective endings were recorded as the dependent variable.

4.1.2. Results and Discussion

Twenty-four items (3.4%) were coded as missing values because the children failed to produce an answer, they produced grammatical but non-targeted neuter responses or the experimenter accidentally revealed the noun gender. Agreement errors per child averaged 7.4 %, and ranged from 0 to 25 % (S.D. = 6.6 %).

The agreement error percentages, corrected for the number of missing values per subject and condition, are presented in Table 4.1. A 2 (noun familiarity: familiar vs. unfamiliar) x 2 (derivational status: simplex vs. diminutive) x 2 (gender: feminine vs. masculine) within-subjects ANOVA revealed a main effect of familiarity, $F(1,21) = 20.3$, $p<0.001$, $\eta^2=0.11$ which indicated that children made more errors with unfamiliar than with familiar nouns, a main effect of derivational status, $F(1,21) = 11.8$, $p<0.05$, $\eta^2=0.03$, due to fewer errors for diminutive nouns than for their simplex counterparts, and a main effect of noun gender, $F(1,21) = 12.4$, $p<0.05$, $\eta^2=0.10$, due to fewer errors for masculine than for feminine nouns.

Table 4.1. Gender agreement error percentages and standard deviations (in parentheses) as a function of noun familiarity, derivational status, and gender.					
	Simplex nouns			Diminutive nouns	
	Feminine	Masculine		Feminine	Masculine
Familiar nouns	4.6 (12.5)	0.0 (0.0)		2.6 (8.7)	0.0 (0.0)
Novel nouns	30.3 (28.5)	7.2 (12.1)		12.5 (24.1)	1.1 (5.3)

The analysis also yielded an interaction between noun familiarity and gender, $F(1,21) = 8.3$, $p<0.05$, $\eta^2=0.04$, suggesting that, like in Russian, the familiarity effect was predominantly carried by feminine nouns and an interaction between familiarity and derivational status, $F(1,21) =6.8$, $p<0.05$, $\eta^2=0.02$, suggesting that in Serbian children, the familiarity effect was somewhat more pronounced in novel simplex nouns. With the exception of the interaction between familiarity and derivational status, the results were almost identical to Russian (Ševa, et al., in press). Children performed better with familiar nouns compared to novel nouns, and with diminutive nouns compared to simplex nouns. As in Russian, performance in masculine nouns was near ceiling resulting in a more pronounced familiarity effect for feminine nouns.

The source of the masculine advantage is not clear: One possibility is that it is due to the relatively high frequency of gender ambiguous words both in CDS and in children's speech (words like *tata* 'daddy', *meda* 'teddy bear' or proper names ending in *–a*) which end like feminine nouns, but take masculine gender adjectives. Such words exist in Russian as well, and Russian children also exhibit superior gender agreement performance for masculine nouns. It is also possible that the masculine advantage is due to the shorter and, thus, morphologically less complex

masculine adjectives (e.g. *lep, loš*) which are often one syllable shorter than their feminine counterparts (e.g. *lepa, loša*). Preliminary work on gender agreement in Lithuanian (Savickienė et al., in preparation), where masculine and feminine adjectives are of equal length, confirms the diminutive advantage but does not show a masculine advantage in gender agreement performance. Thus, gender agreement in Serbian and Russian masculine nouns might be aided by the fact that the adjectives are phonologically less complex and, thus, easier to produce.

4.1.3. Joint analysis and discussion of the Serbian and Russian experiments

The comparisons effect sizes for Russian (Ševa, et al., in press) and Serbian experiments revealed that the noun familiarity accounted for 3% of variance in Russian and for 11% in Serbian, derivational status accounted for 8% of variance in Russian and 3% in Serbian, and gender accounted for 2% of variance in Russian and 10 % in Serbian. Given that the corresponding effect sizes for Russian in Kempe et al. (2003) were 6% for familiarity, 2% for derivational status, and 12% for gender, respectively, it seems that the differences between Russian and Serbian do not exceed n`ormal fluctuations found across different studies within a language. In short, the results of this experiment support the existence of a diminutive advantage in Serbian which is similar to Russian. The only difference to Russian was that the familiarity effect was more pronounced in the simplex nouns.

In order to obtain a more accurate estimate of cross-linguistic differences, I performed a 4-way ANOVA with noun familiarity, derivational status and gender as within-subjects factors and language as a between-subjects factor. This analysis confirmed all the effects found in the Russian and Serbian experiments. I found a

main effect of familiarity, $F(1,44) = 24.2$, $p < 0.001$, $\eta^2 = 0.07$, a main effect of gender, $F(1,44) = 15.2$, $p < 0.001$, $\eta^2 = 0.09$, a main effect of derivational status, $F(1,44) = 19.3$, $p < 0.001$, $\eta^2 = 0.03$, as well as significant interactions between familiarity and gender, $F(1,44) = 15.7$, $p < 0.001$, $\eta^2 = 0.04$, and a significant interaction between familiarity and derivational status, $F(1,44) = 4.4$, $p < 0.05$, $\eta^2 = 0.01$ which is depicted in Figure 4.1. The only effect involving the factor of language was the interaction between familiarity, gender and language, $F(1,44) = 4.4$, $p < 0.05$, $\eta^2 = 0.01$. This interaction is depicted in Figure 4.2. It indicates that the Serbian children had more difficulty with novel feminine nouns.

The two experiments with Serbian and Russian children provide a very stringent cross-linguistic comparison of gender-agreement performance in Russian and Serbian children. In both languages, children performed better with familiar nouns than with novel nouns, and with masculine nouns than with feminine nouns. Most importantly, children in both languages showed superior gender-agreement performance with diminutive compared to simplex nouns. The only subtle difference between languages concerned the feminine novel nouns which proved to be slightly more difficult for Serbian children than for Russian children. There are two possible explanations for this effect. One would be that the relatively high percentage of hypocoristic forms of Serbian masculine animal nouns and kinship terms ending with *-a* (e.g. *medved* 'bearSIM' vs. *meda* 'bearHYP') additionally obscures the gender distribution which can then mislead children in the gender-agreement task.

Although hypocoristics like *meda* 'bearHYP' or *zeka* 'rabbitHYP' exist in Russian too, e.g. *mishka* 'bearHYP' or *zajka* 'rabbitHYP', these forms tend to be more productive and hence, more frequent, in Serbian. In fact, a comparison of the

corpora of CDS described above yielded a token frequency of hypocoristics of 20.3% in Serbian and 2.9% in Russian.

One the other hand, it is possible that the gender effect is just a consequence of the selection of the novel nouns for Serbian which had been modelled after the Russian novel nouns. A comparison of the individual feminine novel nouns revealed that especially the pseudo-words *timza* and *mompa* elicited more errors in Serbian than in Russian. It is possible that these two items constitute slightly less acceptable non-words in Serbian than in Russian. Still, this minor difference between the languages does not affect the main finding, namely that there is a diminutive advantage for gender-agreement production of comparable magnitude in Russian and in Serbian, despite the fact that the frequency of diminutives in Serbian CDS is markedly lower.

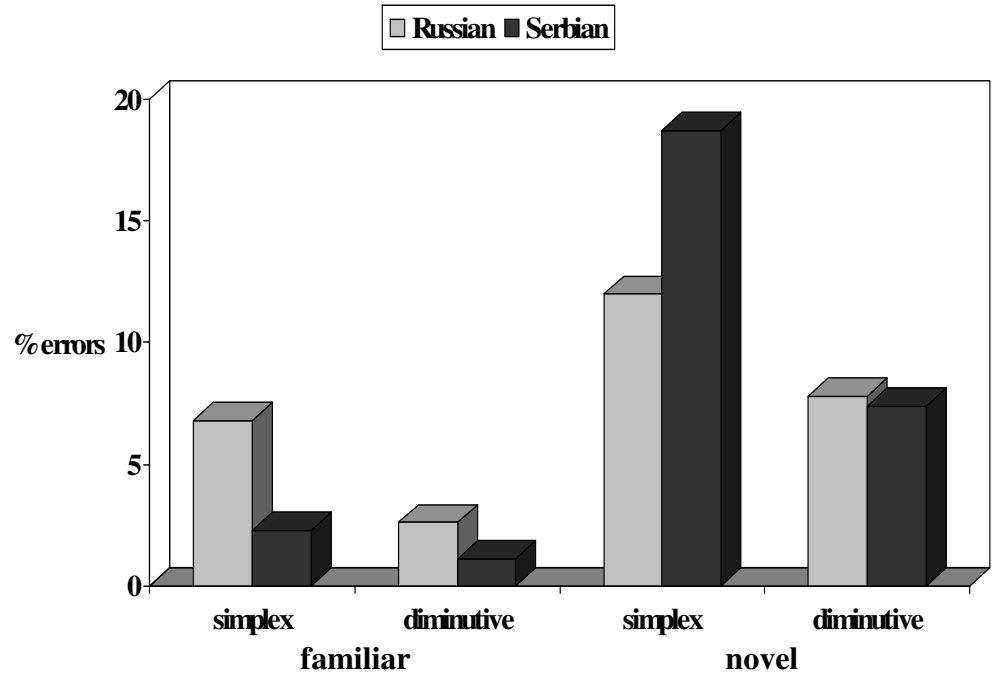


Figure 4.1. Percent gender-agreement errors in Russian and Serbian children as a function of noun familiarity and derivational status.

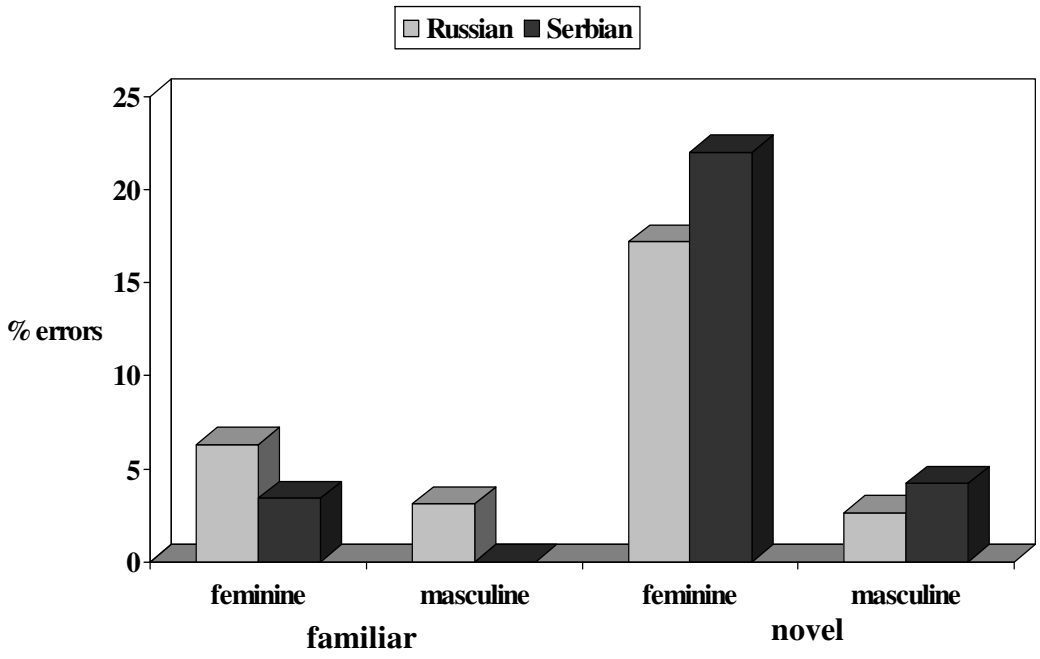


Figure 4.2. Percent gender-agreement errors in Serbian and Russian children as a function of noun familiarity and noun gender.

4.2. Case marking in Serbian children

Results from the gender agreement experiment with Serbian children addressed the question as to whether the frequency of diminutives is responsible for the observed diminutive advantage in noun morphology acquisition. Before undertaking further explorations of possible learning mechanisms behind this effect, it would be interesting to see whether the observed diminutive advantage for Serbian gender marking can be extended to other domains of Serbian noun morphology, like case marking. The second part of this chapter will address this question by replicating the Russian experiments for case marking (Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press; Kempe et al., in preparation).

In order to stay as close as possible to the paradigm of the Russian experiment, I used the same relations of movement used in the Russian study: **moving from an object** and **moving towards an object.** The relation of moving from an object is expressed with the *od+genitive* construction in Serbian which is highly frequent, and one of the first to be acquired. The second relation of moving towards an object is expressed with three prepositional phrases in Serbian: *do+genitive*, *prema+dative* and *k(a)+dative*. Based on previous research on the acquisition of locative constructions in highly inflected languages like Serbian (Johnston & Slobin, 1979), this particular relation should present problems for children, because it can be expressed with three almost synonymous constructions in Serbian: a) for marking the direction and the end of the movement – *do + genitive*, b) for marking the direction of movement – *prema + dative* and, like in Russian, *k(a) + dative*. In addition to the spatial meaning, some of these prepositional phrases can be used for expressing temporal or relational meanings. A

more detalied corpus analysis of the four different registers children's speech, CDS, ADS and written language showed that in both children's speech and CDS the most frequent construction used for marking the relation of moving towards an object is *do + genitive* (see Table 4.2.). *Prema + dative* and *ka + dative* were infrequent both in CDS and in children's speech, with a slight advantage for the *prema + dative* construction. For the purpose of this study I used the construction *prema + dative* in order to be able to present different constructions in the case marking task (*od + genitive* vs. *prema + dative* as oppossed to *od + genitive* vs. *do + genitive*) which require the same genitive case, and because it will allow for further explorations of the similarities and differences in error types in comparison to Russian. However, the existence of alternative constructions for the expression of movement towards an object in Serbian constitutes a feature that is different between the two languages making an exact replication difficult.

Table 4.2. Percentage of tokens of the four prepositions out of the total number of words across four different registers: child speech, CDS, ADS and written language.				
	od + genitive	*do + genitive*	*prema + dative*	*k(a) + dative*
children's speech[1]	0.16	0.02	0.0	0.0
CDS[2]	0.16	0.03	0.008	0.001
ADS[3]	0.39	0.25	0.02	0.003
written language[4]	0.7	0.3	0.07	0.06

[1] Estimates based on the five age samples (20, 26, 32, 38, 44 months) for 8 children from the *Serbian Corpus of Early Child Language* (Anđelković et al., 2001).
[2] Estimations based samples for parents of 8 children when children were 20, 26, 32, 38, 44 months from the *Serbian Corpus of Early Child Language* (Anđelković et al., 2001).
[3] Estimates based on the *Conversational corpus of Serbo-Croatian language* (Savić & Polovina, 1989).
[4] Estimates based on the *Frequency dictionary of contemporary Serbian language* (Kostić, 2001).

4.2.1. Method

Participants: Twenty-four Serbian-speaking children (10 girls and 14 boys), aged 2;10 – 4;11 years (mean age 3;8 years), were recruited in various kindergartens in Belgrade (Serbia).

Materials: I selected six masculine and six feminine nouns and the corresponding objects. All the objects were familiar to small Serbian children. I selected four additional familiar nouns and objects for practice purposes. Furthermore, I created 12 unfamiliar nouns, six of which ended in *–a* thus resembling the form of feminine nouns, and six ending in a consonant thus resembling the form of masculine nouns. All novel nouns were bi-syllabic with stress on the first syllable. I selected another 12 novel objects which were highly unusual objects not readily nameable by Serbian children or adults, and assigned the novel words to the novel objects. In addition, I used a toy elephant which served as protagonist in the game. All familiar

and novel Serbian words are listed in their diminutive and simplex forms in Table A6.1. in Appendix 6.

The nouns and their diminutive derivations were distributed across two lists in such a way that each noun appeared as simplex in one list, and as diminutive in the other (see Table A6.2. in Appendix 6 for more details). Each list contained an equal number of simplex and diminutive, familiar and unfamiliar nouns. Half of the children were presented with one list, and the other half with other list. I created four pseudo-randomisations of items per list with the only restriction that each list had to start with a familiar noun. Children were quasi-randomly assigned to the eight lists, matching for sex and age.

Procedure: Each child was tested individually by a female native speaker of Serbian in a room adjacent to the main activity room of the kindergarten. Children were first introduced to the protagonist of the game, the little elephant. Then the experimenter took one practice object and labelled it by saying: *Ovo je forzak. Ponovi.* 'This is the forzak. Repeat.' Next, the experimenter showed how the elephant was moving towards or moving away from the object, and asked one of the questions designed to elicit a case-marked response: *Kuda ide slon? (Odakle ide slon?*) 'Where is the elephant going? (Where is the elephant coming from?)'. Children were prompted to produce two common prepositional phrases (*prema* + *dative* for the 'going to' relation and *od*+*genitive* for the 'moving from' relation). When the child gave a response containing a case-marked noun, the experimenter moved on to the next practice item, and subsequently to the target items. If the child failed to give a response, the experimenter modelled the correct response for the practice items, and encouraged the child to repeat it. Once the child was able to form the response on their own, the experimenter proceeded with the target items. In this experiment, the

experimenter had to model the targeted responses in 24% of trials, by supplying the correct preposition, mainly for the *prema + dative* construction, because the children kept producing the more frequent and familiar synonymous constructions like *do + genitive, kod + genitive,* or they would just alternate between the two constructions elicited in this experiment: *prema + dative,* instead of *od + genitive*, or vice versa. In every other respect, the procedure was identical to the Russian case experiment (Kempe et al., in preparation).

4.2.2. Results, analysis of error types and discussion

4.2.2.1. Results and discussion

The children produced the following types of responses:

a) 85.3% targeted answers (*od + genitive* and *prema + dative*); b) 10.6% non-grammatical answers counted as real errors; c) 2.2% grammatical non-targeted answers which mainly appeared when the children failed to produce the *prema + dative* construction, even after the experimenter prompted the examples. In this case, the children produced *do + genitive*, *kod + genitive*, or *od + genitive*. These items were excluded from further analysis in order to stay consistent with the way of coding in the gender marking experiment where I excluded the grammatical non-targeted neuter answers. In addition, 24 items (1.9%) were coded as missing values because the children failed to produce an answer, or the experimenter accidentally revealed the case of the noun. Case marking errors per child averaged 11%, and ranged from 0 to 46% (S.D. = 13 %). Since the children in this study encompassed an age range of over two years, I performed a median split by age (3;9 years) to

explore whether there were any age effects in producing correct case inflections for the novel nouns. The mean age for younger group was 3;4 years, and the mean age for older group was 4;0 years. The case marking errors percentages, corrected for the number of lost trials and non-targeted grammatical responses per subject and condition over two age groups are presented in Table 4.3. A 2 (age: below median vs. above median) x 2 (noun familiarity: familiar vs. novel) x 2 (derivational status: simplex vs. diminutive) x 2 (gender: feminine vs. masculine) x 2 (case: genitive vs. dative) omnibus ANOVA revealed significant main effects of: age, $F(1,22)=7.67$, $p<0.05$, with the older children making fewer errors (4.9%) in comparison to the younger group (18.3%), familiarity, $F(1,22)=5.85$, $p<0.05$, $\eta^2=0.02$, with better performance on familiar than on novel words and a main effect of derivation, $F(1,22)=10.85$, $p<0.05$, $\eta^2=0.01$, with diminutives being easier to decline in comparison to simplex forms of nouns. The analysis also revealed a two-way interaction between familiarity and derivation, $F(1,22)=12.3$, $p<0.01$, $\eta^2=0.02$ indicating that the effect of derivation was mainly carried by novel nouns. The rest of the three-, four- and five-way interactions are presented in Table 4.4.

Taken together, the case-marking experiment replicated the diminutive advantage observed for gender agreement which indicates that diminutives have a facilitating effect on both gender and case acquisition in Serbian. Also, this production experiment confirmed previous assumptions based on comprehension tasks that children's case marking errors are at ceiling by the age of four in the complex morphological systems like Serbian. The three-, four- and five-way interactions indicated that this effect was mainly carried by the older children performing better in inflecting masculine dative and feminine genitive nouns.

The rest of this chapter will provide more detailed analyses of error types, as well as comparisons with results from the Russian case-marking experiment (Kempe et al., in preparation).

Table 4.3. Case marking error percentages and standard deviations (in parentheses) for two age groups separately as a function of noun familiarity, gender, derivational status, and case.

	Masculine				Feminine			
	Diminutive		Simplex		Diminutive		Simplex	
	Dative	Genitive	Dative	Genitive	Dative	Genitive	Dative	Genitive
Younger children								
Familiar nouns	23.6(33.7)	12.5(31.1)	19.4(38.8)	11.1(29.6)	22.2(38.5)	5.5(19.2)	16.7(38.9)	8.3(20.7)
Novel nouns	25.0(45.2)	4.2(14.4)	30.5(36.1)	13.9(22.3)	19.4(38.8)	11.1(21.7)	30.5(41.3)	38.9(42.2)
Older children								
Familiar nouns	0.0(0.0)	0.0(0.0)	0.0(0.0)	0.0(0.0)	4.2(14.4)	2.8(9.6)	5.5(13.0)	8.3(15.1)
Novel nouns	11.1(21.7)	0.0(0.0)	5.5(13.0)	11.1(26.0)	0.0(0.0)	0.0(0.0)	29.2(24.7)	0.0(0.0)

Table 4.4. The significant three-, four- and five-way interactions for Serbian case marking experiment.

Interaction	Significant effects	The effects explanations
gender*familiarity*age	$F(1,22)=4.5$, $p<0.05$, $\eta^2=0.01$	This effect was mainly carried by the older children performing better than the younger children in the novel feminine nouns condition.
gender*case*age	$F(1,22)=5.1$, $p<0.05$, $\eta^2=0.01$	This effect was mainly carried by the older children performing better than the younger children in the dative of masculine nouns condition.
gender*familiarity*case*age	$F(1,22)=6.7$, $p<0.05$, $\eta^2=0.01$	This effect was mainly carried by the older children performing better than the younger children in the genitive of novel feminine nouns condition.
gender*derivation*case*age	$F(1,22)=6.9$, $p<0.05$, $\eta^2=0.01$	This effect was mainly carried by the older children performing better than the younger children in the genitive of simplex feminine nouns condition.
gender*familiarity*derivation* case*age	$F(1,22)=6.0$, $p<0.05$, $\eta^2=0.01$	This effect was mainly carried by the older children performing better than the younger children in the genitive of simplex novel feminine nouns condition.

4.2.2.2. Error type analysis

Based on the error analysis for Russian case marking (Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press; Ševa et al., in press; Kempe et al., in preparation), the erroneous responses were divided into five groups:

1) *nominative* answers – items where children produced the nominative case instead of the targeted form, e.g. *od* + *nominative* [od + forzak-Ø] instead of *od* + *genitive* [od + forzak-a];

2) *wrong case-right paradigm* – items in which children produced an incorrect case inflection appropriate for the gender of the noun, i.e. the masculine genitive ending *–a* instead of the required dative ending *–u* as in '*prema forzak-a*' instead of *'prema forzak-u*;

3) *right case-wrong gender* – instances in which children produced a case inflection appropriate for the opposite gender, i.e. the feminine genitive ending *–e* in conjunction with a masculine noun as in '*od forzake*' instead of *'od forzaka';*

4) *wrong case – wrong gender* – instances in which children produced a wrong case inflection for the opposite gender, i.e. the feminine dative ending *–i* in conjunction with a masculine noun as in "*od forzak-i*" instead of "*od forzak-a*";

5) *other* – any other type of answers which could not be classified into any of the previous categories.

Table 4.5. lists the mean error percentages for each error type as well as the significant results of the omnibus ANOVA with 2 (familiarity: familiar vs. novel) x 2 (derivational status: simplex vs. diminutive) x 2 (gender: feminine vs. masculine) x 2 (case: genitive vs. dative as within-subject factors) and 2 (age: below median vs. above median) as between-subjects factor. On average, children produced

significantly more ‘nominative’ than ‘right case wrong gender’ errors, $t(23)=3.4$, $p<0.01$ and more ‘wrong case right gender’ than ‘right case wrong gender’ errors, $t(23)=2.3$, $p<0.05$. The difference between ‘nominative’ and ‘wrong case right gender’ type of errors was not significant, $t(23)<1$.

Separate ANOVA’s for each type of error revealed that the ‘wrong case right gender’ error type was mainly carried by the ‘dative of novel feminine simplex nouns’ condition. The ‘nominative’ errors were more pronounced with the simplex than with diminutive nouns and with novel than with familiar nouns. This analysis also showed an interesting 3-way interaction between gender, case and age for nominative errors which indicated that for feminine nouns, younger children were more prone to decline the noun, in contrast to masculine nouns where they just repeated the citation (nominative) form, especially in the dative masculine condition.

Table 4.5. Mean error type percentage and standard deviations (in parentheses), and the results of the omnibus ANOVA.

Error type	Mean (%)	Significant effects
nominative	4.0 (4.7)	**Familiarity:** $F(1,22)=7.5$, $p<0.05$, $\eta^2=0.03$ **Derivation:** $F(1,22)=8.1$, $p<0.01$, $\eta^2=0.02$ **Familiarity * Derivation:** $F(1,22)=6.8$, $p<0.05$, $\eta^2=0.01$ **Gender*Case:** $F(1,22)=8.6$, $p<0.01$, $\eta^2=0.02$ **Gender*Case*Age:** $F(1,22)=8.6$, $p<0.01$, $\eta^2=0.02$
wrong case right gender	6.5 (12.7)	**Gender*Case:** $F(1,22)=12.4$, $p<0.01$, $\eta^2=0.007$ **Gender*Familiar*Derivation*Case:** $F(1,22)=5.1$, $p<0.05$, $\eta^2=0.005$
right case wrong gender	0.5 (1.3)	
wrong case wrong gender	0.0 (0.0)	
others	0.0 (0.0)	

4.2.3. Joint analysis of the Russian and Serbian case marking experiments

In order to test whether the observed diminutive advantage found in the case marking experiment with Serbian children is in the same range as for Russian children, I performed an ANOVA with 2 (familiarity: familiar vs. novel) x 2 (derivational status: simplex vs. diminutive) x 2 (gender: feminine vs. masculine) x 2 (case: genitive vs. dative) as within-subject factors and 2 (age: below median vs. above median) and 2 (experiment: Russian vs. Serbian) as between subject factors. The between-subject effects of the experiment were not significant, $p > 0.4$ which indicates that overall Serbian and Russian children performed similarly across the two experiments. The analysis revealed main effects of familiarity, $F(1,44)=19.9$, $p<0.001$, $\eta^2=0.024$, and derivation, $F(1,44)=16.75$, $p<0.001$, $\eta^2=0.016$. The two-way interaction between familiarity and derivation, $F(1,44)=16.75$, $p<0.001$, $\eta^2=0.006$ showed that the effect of derivation was mainly carried by the novel simplex nouns in both languages.

Furthermore, the main effect of age was significant, $F(1,44)=4.7$, $p<0.05$, $\eta^2=0.036$ with the older children performing better (9.7% of errors) than the younger children (16.2% of errors). The two-way interaction between the factors experiment and age, $F(1,44)=5.5$, $p<0.05$, $\eta^2=0.042$, showed that the age effect was mainly carried by an improvement in performance in the Serbian older children. The overall performance of the Serbian and Russian children as a function of familiarity and derivation is depicted in Figure 4.3. for the younger children and Figure 4.4. for the older children.

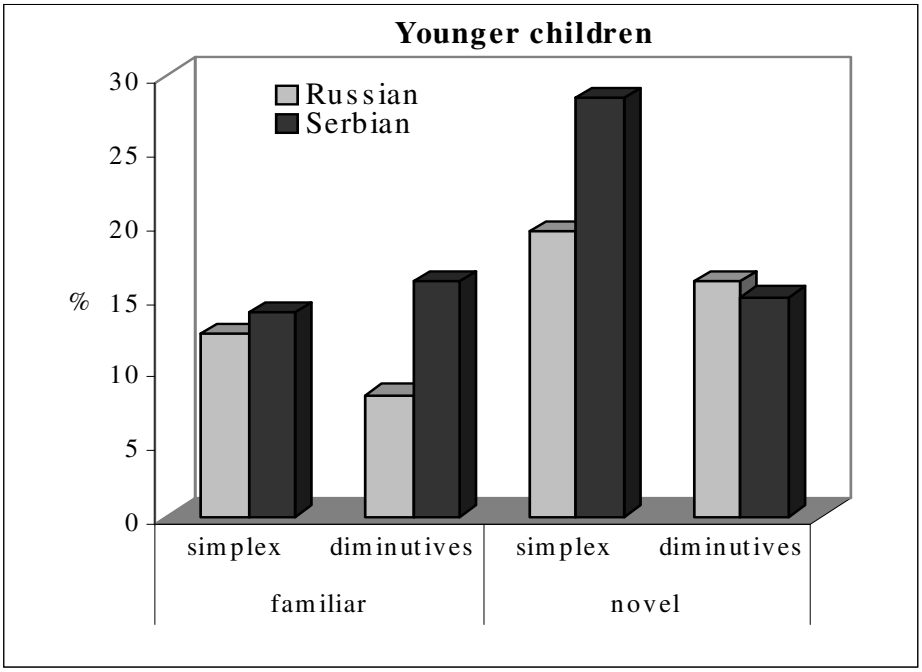


Figure 4.3. Mean percentage of errors in Serbian and Russian case marking experiment as a function of familiarity and derivation for younger group of children.

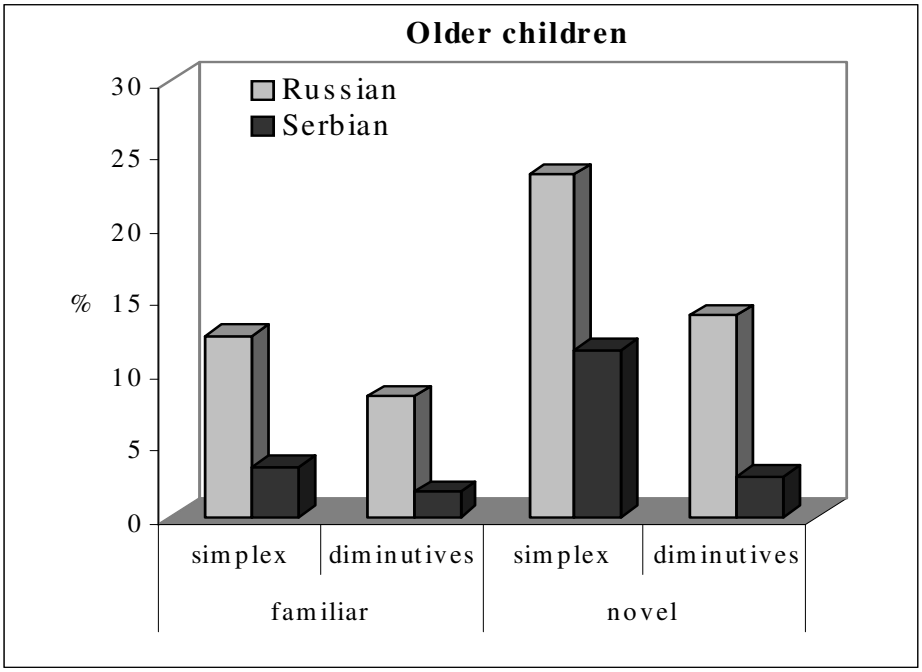


Figure 4.4. Mean percentage of errors in Serbian and Russian case marking experiment as a function of familiarity and derivation for older group of children.

The other significant interactions are presented in Table 4.6. The significant interactions which showed a difference between two experiments are depicted in Figures 4.5-4.8.

Table 4.6. Overview of the interactions with short explanation of the effects in the joint analysis of the Serbian and Russian case experiment.

Interaction	**Statistics**	**Explanations of the effects**
gender*derivation	$F(1,44)=6.9$, $p<0.05$, $\eta^2=0.007$	This effect was mainly carried by the difference between the simplex and diminutive feminine nouns being more pronounced than between the simplex and diminutive masculine nouns.
gender * case * age	$F(1,44)=10.1$, $p<0.01$, $\eta^2=0.007$	This effect was mainly carried by older children performing better than younger children in the 'dative of masculine nouns' condition.
gender*familiar*case*age	$F(1,44)=5.4$, $p<0.05$, $\eta^2=0.004$	This effect was mainly carried by older children performing better than younger children in 'genitive of novel feminine nouns' condition.
gender * familiar * derivation * case * age	$F(1,44)=6.4$, $p<0.05$, $\eta^2=0.003$	This effect was mainly carried by older children performing better than younger children in 'genitive of simplex novel feminine' nouns.
gender*familiarity	$F(1,44)=8.0$, $p<0.01$, $\eta^2=0.007$	This effect was mainly carried by the higher difference between familiar and novel masculine nouns than between familiar and novel feminine nouns.
gender*familiarity*exp.	$F(1,44)=11.2$, $p<0.01$, $\eta^2=0.013$	This effect was mainly carried by Serbian children performing better than Russian children in the 'novel masculine nouns' condition.
gender*case*exp.	$F(1,44)=5.9$, $p<0.05$, $\eta^2=0.004$	This effect was mainly carried by Serbian children performing better than Russian children in the 'genitive of masculine nouns' condition.
familiar * case*exp.	$F(1,44)=5.5$, $p<0.05$, $\eta^2=0.005$	This effect was mainly carried by Serbian children performing better than Russian children in the 'genitive of novel nouns' condition.
derivation*case*exp.	$F(1,44)=4.1$, $p<0.05$, $\eta^2=0.003$	This effect was mainly carried by Serbian children performing better than Russian children in the 'genitive of diminutive nouns' condition.

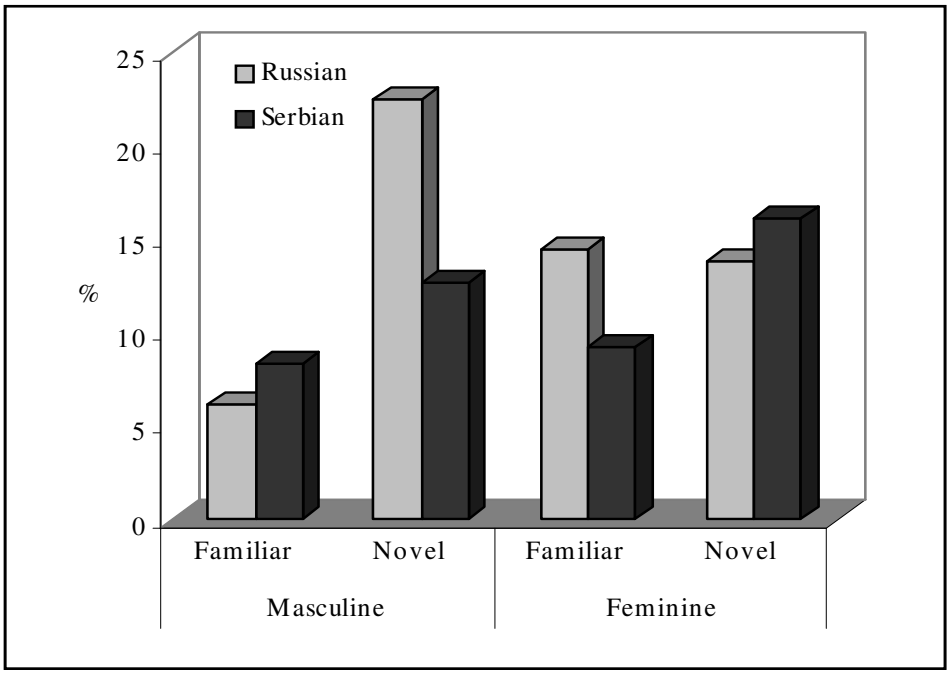


Figure 4.5. Mean percentage of errors in the Serbian and Russian case marking experiments as a function of familiarity and gender.

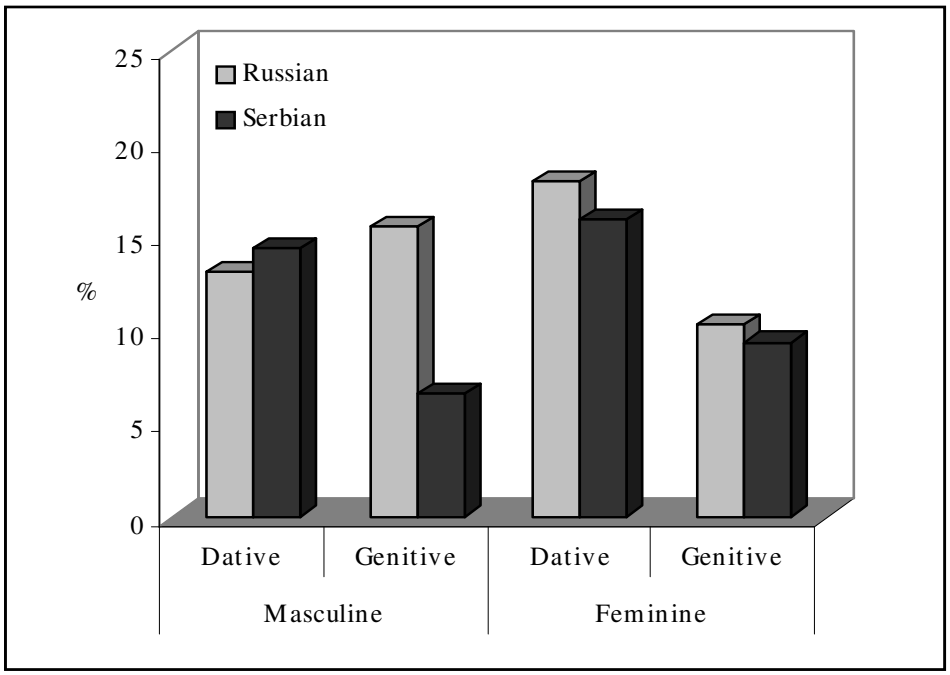


Figure 4.6. Mean percentage of errors in the Serbian and Russian case marking experiments as a function of case and gender.

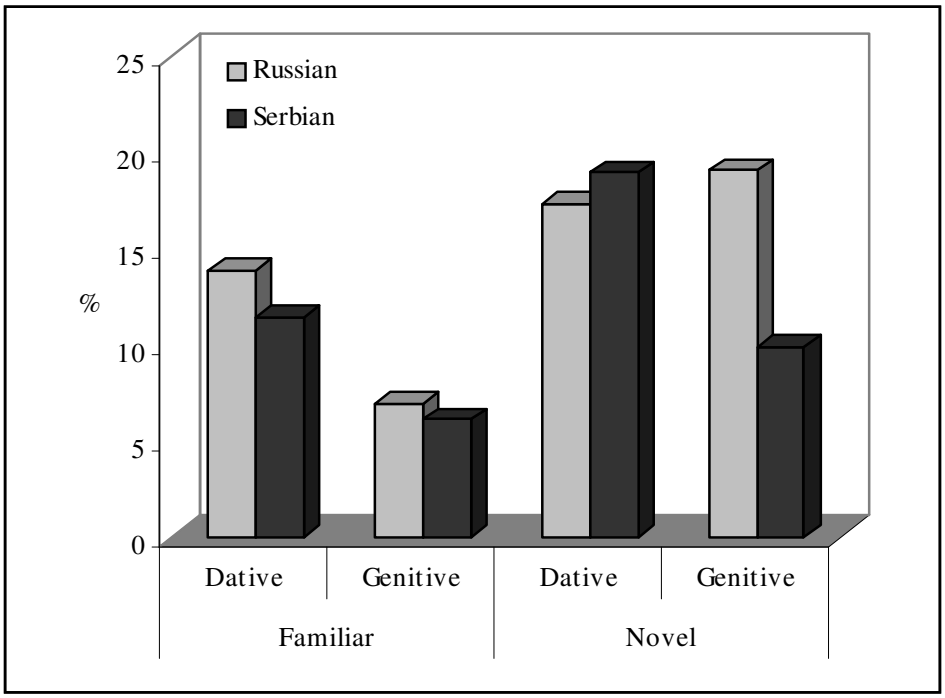


Figure 4.7. Mean percentage of errors in the Serbian and Russian case marking experiments as a function of familiarity and case.

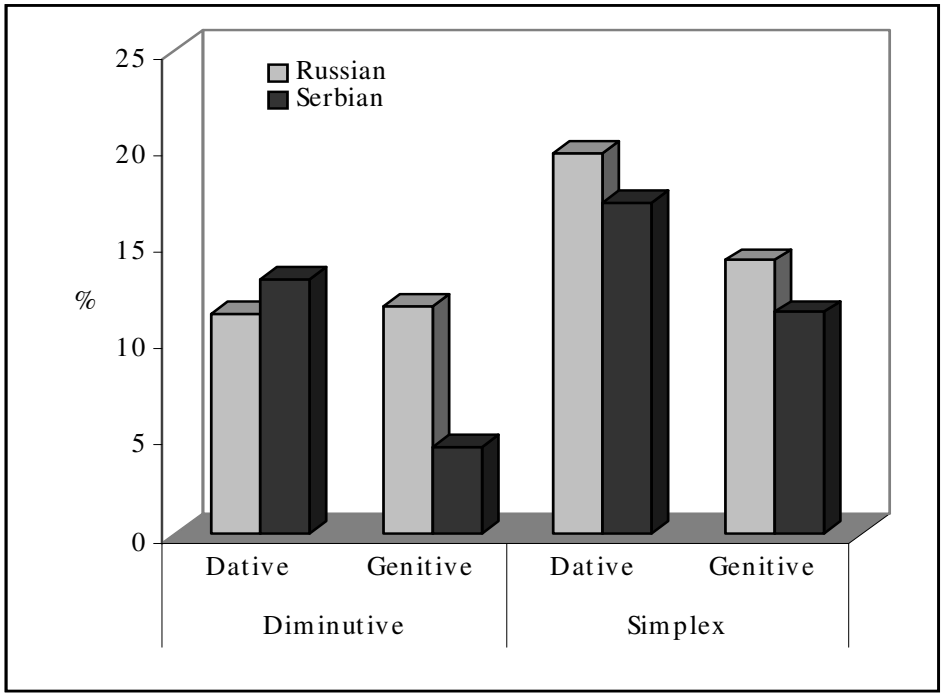


Figure 4.8. Mean percentage of errors in the Serbian and Russian case marking experiments as a function of derivation and case.

Taken together, Serbian and Russian children performed very similarly in the case marking experiment, with better performance on familiar compared to novel nouns, and diminutive compared to simplex nouns. The main differences were related to the older Serbian children performing better than the older Russian children. Also, the Russian children had more problems with masculine novel nouns and with genitive of masculine nouns in comparison to Serbian children. This effect was mainly carried by *right case wrong gender* errors, where Russian children used feminine instead of masculine suffixes. "Feminisation" of masculine nouns was a marginal type of error with Serbian children in this experiment.

4.3. General discussion on the results for Serbian children

From the estimation of diminutive frequencies in Serbian CDS and other Slavic CDS's, and from the two experiments on gender agreement and case marking with Serbian and Russian children reported above, the following picture emerges: Russian and Serbian are two languages with very similar inflectional systems, but a marked difference in the frequency of diminutives in CDS. Despite this difference, Russian and Serbian children acquire noun morphology faster with diminutive than with simplex nouns, as evidenced by superior gender-agreement and case marking performance with diminutive nouns. In fact, the Russian and Serbian results both for gender agreement and case marking were remarkably similar demonstrating the merits of careful cross-linguistic comparisons.

What remains open is the question as to why there is a sizable advantage for morpho-syntactic processing of diminutives even if their frequency in the input is low. In other words, why are the low-level generalisations extracted as successfully in Serbian as they are in Russian or Polish given the difference in frequency in the input that children encounter? One possible reason is that Serbian children may be sensitive to the overall distribution of all diminutive-like suffixes which comprise 14% of noun tokens and 18% of noun lemmata. Perhaps this slightly higher frequency of all diminutive-like suffixes is sufficient for generalisations of novel nouns to occur. Another possible reason is that in each of these languages, diminutives are distinguished from simplex nouns by their salient word endings, and they are the earliest acquired derivations in children's speech, and the most common derivation in both children's speech and in CDS. It seems that phonological homogeneity among this word cluster and morpho-phonological distinctiveness

from other words might be factors as important for low-level generalisations as the type and token frequency of derived forms. The following chapter will provide a more detailed analysis of the relationship between the various factors responsible for the observed diminutive advantage for gender agreement and case marking in languages with complex noun morphology.

Chapter 5

5. How does the diminutive advantage in learning grammatical gender emerge? Comparing the performance of Serbian children and neural networks.

The strong diminutive advantage for both gender agreement and case marking in Serbian suggests that it is not just the frequency of diminutives in the input nor the degree of regularisation (in Serbian, gender marking in diminutives is as regular as in simplex nouns) that drives the facilitating effect of diminutives. Instead, it seems that it is the properties of diminutives that increase morpho-phonological similarity at the ends of words that are responsible for the beneficial effects on noun morphology acquisition. In the last part of the thesis, I will report on a set of studies using artificial suffixes to further explore how the diminutive advantage emerges given an increased amount of morpho-phonological homogeneity among word endings.

In the first part of the chapter I will present data from an experimental study with Serbian children in which they were introduced to novel nouns with or without artificial suffixes. These suffixes were specially designed to play the role of pseudo-diminutives. More specifically, this experiment will track the emergence of the pseudo-diminutive advantage for gender marking in the same gender-agreement task used for Russian (Kempe et al., 2003; Ševa et al., in press) and for the Serbian experiment presented in Chapter 4. I will then present a series of neural network simulations of the obtained data, designed to capture the learning mechanism which could explain the pattern of results observed in the experimental studies.

5.1. Experimental study with artificial suffixes: gender agreement in Serbian like simplex and pseudo-diminutive nouns

5.1.1. Method

Participants: 24 Serbian-speaking children age 3;6-5;4, mean age 4;4 years were recruited in various kindergartens in Belgrade (Serbia).

Materials: I created 32 unfamiliar nouns, 16 of which ended in *-a* thus resembling the form of feminine nouns, and 16 ending in a consonant thus resembling the form of masculine nouns. All novel nouns were bi-syllabic with stress on the first syllable. I selected pictures of 16 novel animals and 16 novel objects which were highly unusual and not readily nameable by Serbian children or adults, and assigned the novel words to the novel animals and objects. I also constructed two pseudo-noun suffixes: *-upa* for feminine nouns and *-uf* for masculine nouns *which* resembled the Serbian diminutive suffixes *-ica* and *-ić* but do not exist in Serbian and thus were not familiar to the children. These suffixes were then used for the derivation of what could be taken as analogue of diminutives. The nouns and their pseudo-derivations were distributed across two lists in such a way that each noun appeared as simplex in one list, and as pseudo-diminutive in the other. Nouns were divided into four groups (8 nouns per group). The order of groups and lists was counterbalanced over four sessions. Pictures were presented randomly in each session. All novel words are listed in their pseudo-diminutive and simplex forms in Table A7.1. in Appendix 7.

I selected another four familiar nouns and corresponding pictures of familiar animals and objects for practice purposes. Two antonymous adjective pairs were

used to prompt the children to talk about the animals. These pairs were *lep-ružan* (masc.) vs. *lepa-ružna* (fem.) 'beautiful-ugly' and *dobar-loš* (masc.) vs. *dobra-loša* (fem.) 'good-bad'. The adjective endings served as indicators for correct or erroneous gender agreement. In addition, I used a toy elephant which served as protagonist in the phase of the experiment designed to increase the childrens' exposure to the nouns using other constructions.

5.1.2. Procedure

The children were tested individually in four different sessions, 2-7 days apart, by a female native speaker of Serbian in a room adjacent to the main activity room of the day care centre.

Each session comprised three blocks utilising different tasks: (1) In the Practice block, children engaged in the process of labelling and describing objects and animals, and were introduced to a specific pair of adjectives to be used to describe the subsequently presented 4 test items; (2) In the Test block, the experimenter elicited the use of gender-marked adjectives as descriptions of objects and animals by prompting the production of adjectives or adjective-noun phrases; (3) In the Additional Exposure block, children were given more opportunities to familiarise themselves with the novel nouns using the case-marking elicitation methodology described in Chapter 4.

First, the children were shown one template (practice) picture, i.e. the spider, and told: *Ovo je pauk. Da li je pauk dobar ili loš?* 'This is spiderNOM. Is spiderNOM goodMASC or badMASC?'. The practice pictures were used to introduce the children to the activity, and to encourage them to produce whole

sentences. Then the experimenter presented the first test picture, accompanied by the utterance: *Ovo je krufa. Pauk je dobar. A krufa?* 'This is krufaNOM. SpiderNOM is goodMASC. And what about krufaNOM?'. This elicitation form was the only way to avoid the experimenter's use of gender agreement, and gave the children the opportunity to pick one of the members of the adjective pair. The same adjective pair was used for four consecutive test nouns, after which the experimenter introduced a new template noun, along with the other antonymous adjective pair. Alternation of adjective pairs and order of template gender were counterbalanced.

The third block was used as an additional exposure phase, giving the children more opportunities to familiarise themselves with the nouns from the second block, but in a different type of task (case-marking task). This task was presented last because grammatical gender and case paradigms are related in Serbian (each gender is associated with a different case-marking paradigm – see Chapter 2). By presenting the case-marking task last, I eliminated the possibility that children implicitly detected the gender of the novel nouns. In this last block, I used the same elicitation paradigm as in the Russian and Serbian case-marking experiments, described in Chapter 4 (Kempe, Brooks, Mironova, Pershukova, & Fedorova, in press; Kempe et al., in preparation). Childrens' responses were prompted by a toy elephant walking towards or away from each object to produce *od*+genitive 'from+genitive' or *prema*+dative 'towards+dative' constructions with the same set of nouns as in the test phase.

In total, each session constituted of 24 trials containing novel nouns, so that the children could hear and repeat every novel noun three times. Note that during the Test block, the children did not receive corrective feedback when they produced

non-targeted adjective-noun gender-agreement in order to keep exposure conditions identical across children.

5.1.3. Results and discussion

For each trial of the Test block, I transcribed the child's first instance of adjective-noun gender agreement. Cases of non-targeted gender agreement as reflected in the adjective endings were recorded as the dependent variable. Children produced the following types of answers:

a) targeted answers: *Krufa je lepa*. 'KrufaFEM is beautifulFEM' and *Forzak je lep*. 'ForzakMASC is beautifulMASC'

b) two types of answers which were non-targeted gender agreement responses:

1) low frequent neuter responses (*To je lepo.* 'ThisNEUT is beautifulNEUT'.) which appeared only in the first two sessions and which were grammatically unmarked answers to the pronominal subject from the experimenter's item introduction sentence **Introduction**: *Ovo je forzak*. **Answer**: *To je lepo.* 'ThisNEUT is forzak. ThatNEUT is beautifulNEUT';

2) real agreement errors (*Krufa je lep.* 'KrufaFEM is beautifulMASC.' or *Forzak je lepa.* 'ForzakMASC is beautifulFEM.').

Some items were coded as missing values because the children failed to produce an answer, or the experimenter accidentally revealed the noun gender (three in the first session, zero in the second session, one in the third and fourth session). The average percentage of non-targeted answers per child over four sessions computed as proportions of completed trials, corrected for the number of lost trials per subject and condition are presented in Table 5.1.

Table 5.1. Mean percentage of non-targeted answers per child with (N=24) or without (N=16)[15] neuter gender answers counted as errors. Standard derivations are given in parentheses.		
Number of subjects/Session	N=24	N=16[14]
Session 1	18.1 (13.6)	20.1 (14.8)
Session 2	16.1 (14.0)	18.8 (15.8)
Session 3	14.2 (16.3)	17.2 (18.8)
Session 4	11.5 (16.0)	11.7 (15.5)

I performed a 4 (session) x 2 (derivational status: simplex vs. pseudo-diminutive) x 2 (gender: feminine vs. masculine) within-subjects ANOVA on the proportions of non-targeted answers (neuter responses plus real errors). The analysis yielded a main effect of noun gender, $F(1,23) = 11.0$, $p < 0.01$, $\eta^2=0.12$ which indicated that the children committed more errors with feminine than with masculine nouns in all four sessions (see Figure 5.1), as well as a significant two-way interaction between session and noun derivation, $F(1,23) = 5.0$, $p < 0.01$, $\eta^2=0.02$ (see Figure 5.2). None of the other interactions was significant indicating

[15]Additionally, in order to present the results from this experiment consistently to the Serbian experiments with gender agreement and case marking (see Chapter 4), I repeated same within-subject ANOVA, but in this case only with real agreement errors over four sessions computed as proportions of completed trials, corrected for the number of lost trials and non-targeted grammatical answers (neuter gender responses) per subject and condition. Since the number of neuter answers was slightly higher than in the previous two experiments, probably due to the increase in complexity of the stimuli, eight subjects had to be excluded from the overall analysis. Nevertheless, the mean percentage of errors over the four sessions for the remaining 16 subjects did not change dramatically from the analysis for 24 subjects (see Table 5.1).
The 4 (session) x 2 (derivational status: simplex vs. pseudo-diminutive) x 2 (gender: feminine vs. masculine) within-subjects ANOVA on the proportions of real agreement errors showed the same main effect of noun gender, $F(1,15) = 8.5$, $p < 0.05$, $\eta^2=0.15$ which indicated that children committed more errors with feminine than with masculine nouns in all four sessions, as well as a significant two-way interaction between session and noun derivation, $F(3,45) = 4.7$, $p < 0.01$, $\eta^2=0.03$. Similar to the analysis with 24 subjects, none of the other interactions was significant.
Given that all main effects and interactions were the same as for the analysis with 24 subjects, the rest of the analysis will be performed on the larger sample of subjects.

that the obtained decrease in the number of non-targeted answers for pseudo-diminutive nouns after Session 1 was present both in feminine and masculine nouns (see Figure 5.3).

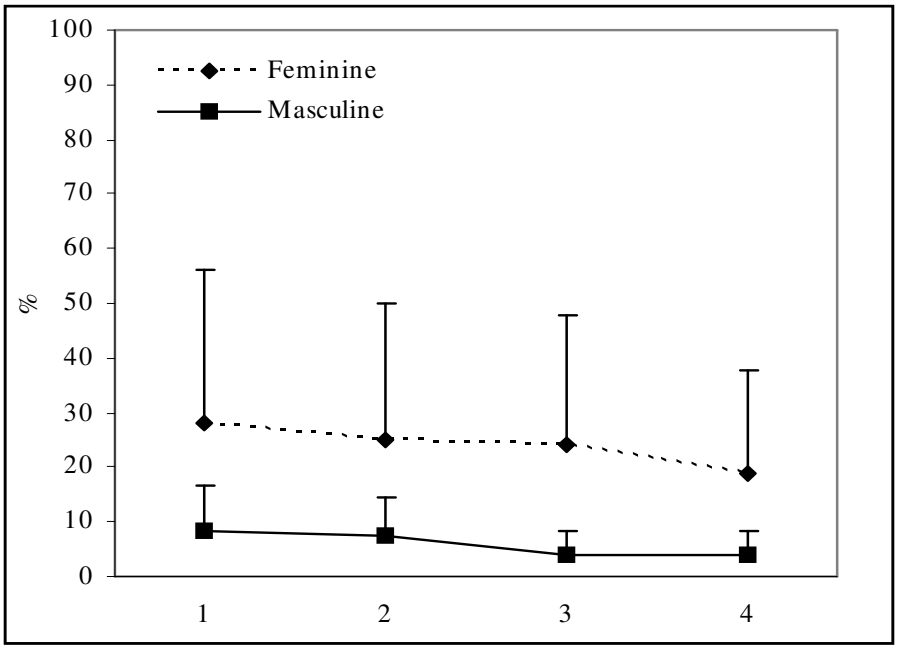


Figure 5.1. Mean percentage of non-targeted answers (and 1 S.E.M.) over four sessions and two genders.

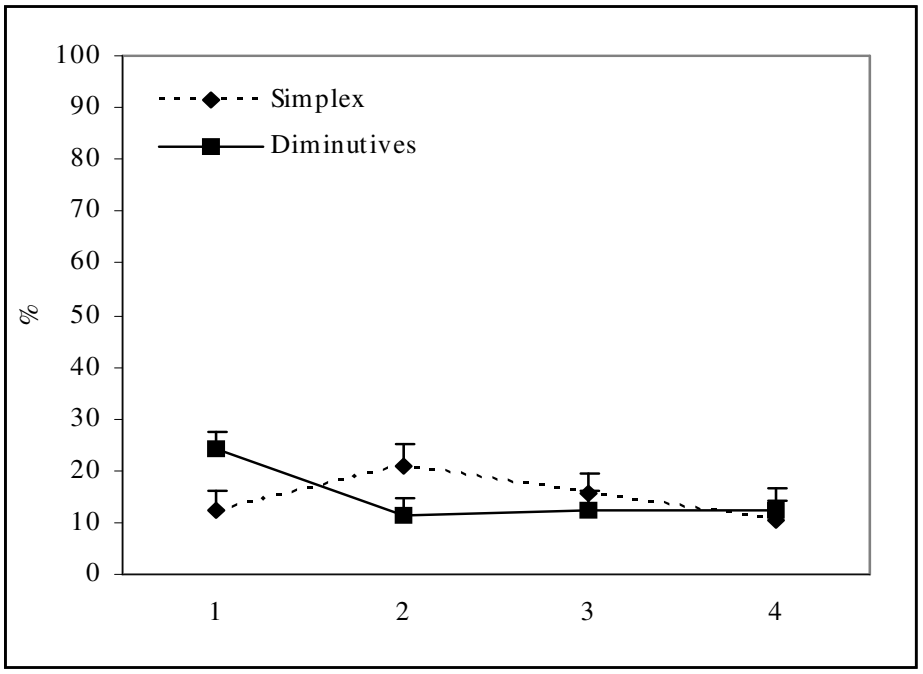


Figure 5.2. Mean percentage of non-targeted answers (and 1 S.E.M.) over four sessions and two derivations.

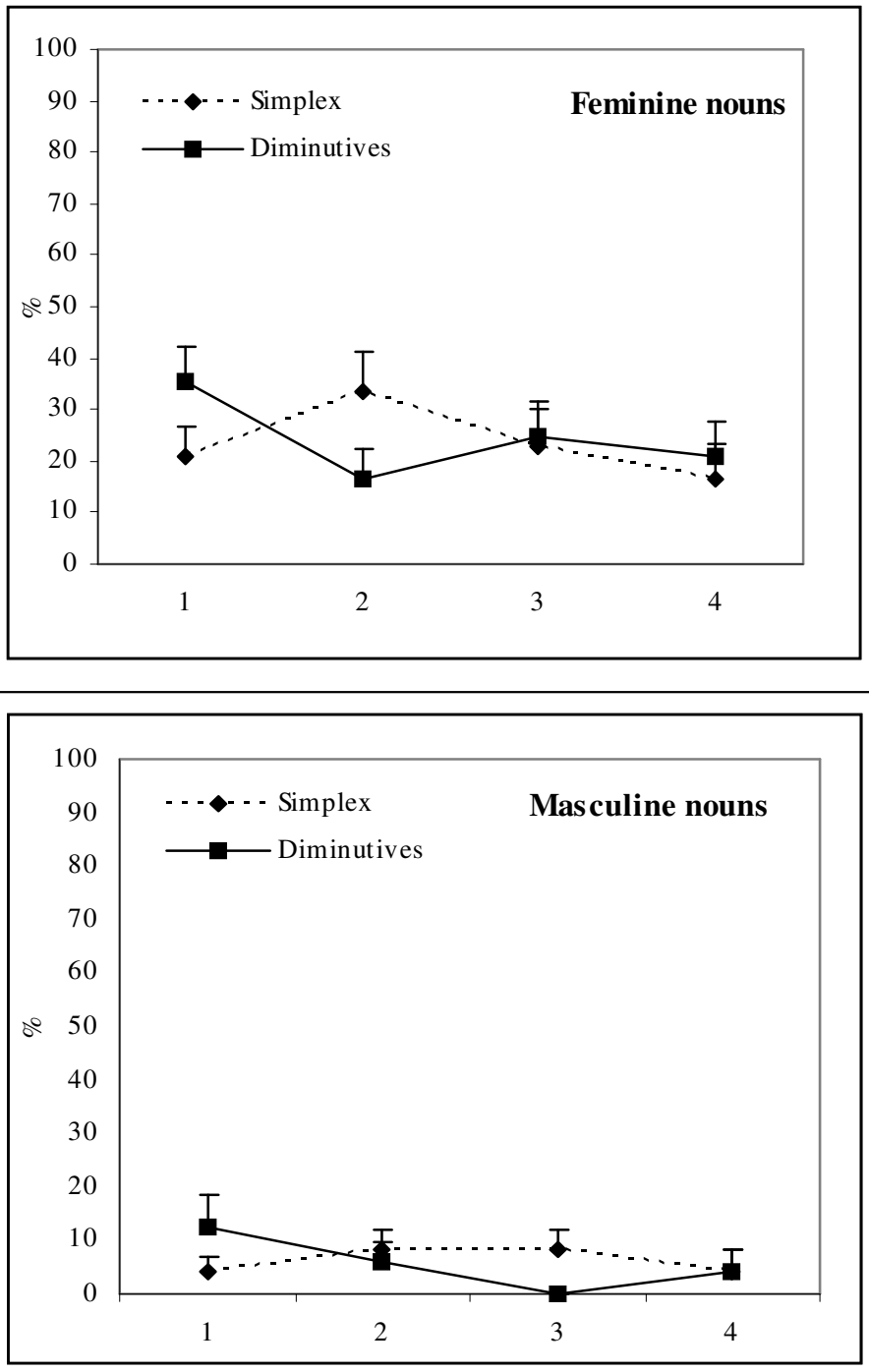


Figure 5.3. Mean percentage of non-targeted answers (and 1 S.E.M.) over four sessions and two derivations feminine nouns (upper panel) and masculine nouns (lower panel).

Separate ANOVAs with gender and noun derivation as within-subjects factors were conducted for each session to qualify the interaction between session and derivation.

For Session 1, this analysis revealed a main effect of gender, $F(1,23) = 7.2$, $p < 0.05$, $\eta^2=0.14$, with better performance for masculine nouns, and a main effect of noun derivation, $F(1,23) = 8.3$, $p < 0.01$, $\eta^2=0.05$, indicating that the children performed better with simplex nouns than with pseudo-diminutives. For Session 2, I found a main effect of gender, $F(1,23) = 6.2$, $p < 0.05$, $\eta^2=0.13$, with better performance for masculine items, and a main effect of noun derivation, $F(1,23) = 4.4$, $p < 0.05$, $\eta^2=0.04$, this time due to better performance with pseudo-diminutives compared to simplex nouns. For Sessions 3 and 4, I found only a significant main effect of gender, $F(1,23) = 6.9$, $p < 0.05$, $\eta^2=0.19$ and $F(1,23) = 4.5$, $p < 0.05$, $\eta^2=0.11$, due to better performance for masculine nouns. In sum, these analyses suggest a dramatic change in the effect of pseudo-diminutives: while children performed worse on these items compared to the simplex nouns in Session 1, the effect was reversed in Sessions 2.

Separate 2 (gender) x 4 (session) ANOVAs for the simplex nouns revealed only a main effect of gender, $F(1,23) = 8.9$, $p < 0.01$, $\eta^2=0.11$ confirming the masculine advantage. For the pseudo-diminutives, the 2 (gender) x 4 (session) ANOVA revealed a main effect of session, $F(3,69) = 4.1$, $p < 0.01$, $\eta^2=0.04$, as well as an effect of gender, $F(1,23) = 10.3$, $p < 0.01$, $\eta^2=0.13$. Bonferroni-corrected t-tests comparing performance between all sessions confirmed that for the pseudo-diminutives the improvement in performance between Session 1 and 2 was significant, $t(23) = 3.4$, $p < 0.05$. These analyses suggest that the apparent increase in errors for the simplex nouns between Session 1 and 2 was not significant, while

the decrease in errors for pseudo-diminutives was. Thus, performance in the simplex nouns remained mainly unchanged while performance in pseudo-diminutives improved rapidly after only one session of exposure.

In sum, as in the previous studies, the experiment with artificial pseudo-diminutive suffixes showed that the children committed fewer agreement errors with masculine nouns than with feminine nouns. This confirms earlier findings of a masculine advantage in gender agreement for Russian and Serbian, despite the fact that in the present study, the children were almost a year older (4;4 years) than in the previous Russian (mean age: 3;9 years) and Serbian studies (mean age: 3;7 years) (Ševa et al., in press).

Crucially, this study indicates that a pseudo-diminutive advantage emerged already at Session 2. While the children at first committed many errors with the unfamiliar derivations, they soon seemed to treat these nouns as a phonologically more homogeneous cluster of words compared to the simplex pseudo-words which facilitated correct gender agreement within this cluster. Thus, local/low-level generalisations do not take a long time to emerge nor do they seem to require a lot of exposure to the particular word cluster. In other words, morpho-phonological homogeneity is as crucial a factor as frequency in the process of local/low-level generalisations, and may be sufficient to trigger the process, even if a highly homogeneous cluster of words is not very frequent in the input. This can help to explain why the diminutive advantage in Serbian is of similar magnitude as in Russian despite markedly lower frequency of diminutives in the input of Serbian children. The fact that the high frequency of diminutives in Russian CDS did not provide an additional advantage for the Russian children suggests that phonological homogeneity might be the more important factor.

The following part of this chapter will explore in greater detail which learning mechanisms may underlie the diminutive advantage observed in this experiment.

5.2. Neural-network simulations of the experimental data

The experimental study described in the first part of this chapter showed that novel morpho-phonologically similar words cluster together relatively fast into sufficiently compact groups, and that children used inflectional changes more reliably with words within such a cluster, showing that gender categories of novel words within such a cluster emerged after just a few exposures (the pseudo-diminutive advantage emerged after only one set of presentations). The second part of this chapter will present a series of neural network simulations designed to capture the pattern of results from the experiment. The simulations will shed light on the underlying mechanisms to explain the obtained experimental data.

5.2.1. The general role of neural networks in research on language acquisition

The past two decades of research on language acquisition were marked to a great extent by advances in the field of computational simulations of language-related phenomena. Probably the most important study which established a foundation for this research and caused a lot of controversy was Rumelhart & McClelland's (1986) neural network simulation of the U-shaped learning of English past tense verb forms. This simulation showed that the process of abstraction and generalisation of grammatical categories does not need to be based on symbolic rules and that non-linear change observed in most developmental processes can be explained through an associative learning mechanism and extraction of statistical regularities of the domain to be learned.

From that point on, neural networks were used for the exploration and understanding of the following language related developmental problems (Seidenberg, 1992; Brent, 1996; Elman et al., 1996; Munakata & McClelland, 2003; Elman, 2005):

1) explaining the non-linear shape of change in development and developmental disorders;

2) exploring the nature of language representation and processing mechanisms (symbolic vs. non-symbolic);

3) exploring how much can be learned by extracting distributional patterns from the input.

In a recent overview of neural network simulations of language acquisition, Elman (2005) evaluated this work as very positive, stating that:

…."Connectionist models now offer alternative hypotheses for many important developmental phenomena and in several cases appear to provide a richer and more accurate account of those phenomena than hypotheses from behavioural work"…

At the same time he emphasised that the modelling work on language development so far has captured the "low-lying fruit" of single behaviours and that questions which should be addressed in the future are: a) exploring other domains of development, like social, emotional, physical and moral development, etc.; b) modelling multi-tasking; c) modelling more realistic and detailed behaviours; d) modelling cascading effects over time.

In sum, today's connectionism represents a good methodological tool for testing the effectiveness of algorithms proposed by traditional symbolic accounts on cognition in general. At the same time, it complements alternative models of cognitive processing, like dynamic systems theory which argue that cognition is an

emergent phenomenon, grounded in low-level, simple and non-symbolic processes (Smith & Samuelson, 2003).

From a methodological perspective, in order to be used as a tool for testing theoretical assumptions and for the replication of experimental behavioural data, all neural network models have to be organised and designed around the following criteria (Christiansen & Chater, 2001; Plunkett & Elman, 1997; Chater & Vitányi, 2002; Chater, 1996):

1) **Task veridicality:** Tasks presented to network have to be as close as possible to the tasks presented to human participants.
2) **Input representativeness:** There should be a match between the information available to the model and the person.
3) **Simplicity:** Choices among potential models of finite data should represent an optimal trade-off between model complexity (the simpler, the better) and accuracy of a model's fit to the data.
4) **Data contact:** Success of the model is determined based on the match with the experimental data**.**

The two most common architectures of networks used in modelling of language acquisition processes are:

1) **Feed-forward networks (FFN)** – multilayered networks with a continuous flow of information from the input over one or more layers of hidden units towards output units.
2) **Simple recurrent networks (SRN)** or Elman Networks– a version of feed-forward networks which contain additional layers of context units. In contrast to FFN models, where information flows from hidden units to output units only, in SRN models the state of the hidden units is copied into

the context units. When the next set of data is propagated from the input layer to the hidden layer, the values for the previous activation state of the hidden units are copied back from the context units into the hidden units. Thus, this provides a simple form of recurrence that captures memory effects and that can be used to train networks to perform sequential tasks over time (Elman, 1991).

The task which the children were performing in the experiment with artificial suffixes can be treated as a version of a categorisation task. Since feed-forward networks are mainly used for solving these types of problems, and in order to meet the simplicity principle, the first set of simulations will use the FFN architecture.

All simulations in this chapter were carried out by using the T-learn simulator (Windows version 1.0.3, ftp://ftp.crl.ucsd.edu/pub/neuralnets/tlearn/; Plunkett & Elman, 1997).

5.2.2. Feed-forward network with whole words

5.2.2.1. Network architecture

The first neural network was designed as a feed-forward model. The system had a set of input units fully connected to a hidden layer which was fully connected to an output layer (see Figure 5.4). The 240 input units corresponded to a two or three-syllabic representation of the words. Syllables were represented using a vowel centred CCVCC frame (C=consonant, V=vowel, *r*, *l*, *n* were coded as vowels in cases where they had a vocalic function; Stanojčić & Popović, 2003). Phonemes within each syllable were represented as a sequence of 16 binary phonetic features (Table A8.1. in Appendix 8). This input coding scheme is similar to the one used in previous simulations of the processing of Serbian nouns (Mirković, MacDonald & Seidenberg, 2005).

The output layer had 2 units representing the gender of nouns ("1 0" for feminine and "0 1" for masculine). The hidden unit layer had 20 units.

5.2.2.2. Materials

The network was trained and tested with two sets of words: a) A set of 120 real Serbian nouns, obtained from *The Serbian Corpus of Early Child Language* (Anđelković et al., 2001) was created. Half of the nouns were feminine and half masculine. Several nouns, both in the feminine and the masculine group, were ambiguously marked for gender, with 2 feminine nouns (1.7%) ending in a consonant, and 5 masculine nouns (4.2%) ending in *–a*. Words were sampled in

this way in order to resemble as closely as possible the noun gender distribution in Serbian CDS (following the criterion of input representativeness).

b) The set of Serbian novel words with artificial suffixes used in the experimental study with Serbian children, previously described in the first part of this chapter, was used. The nouns were divided into four groups in exactly the same way as in the experiment (see Appendix 7).

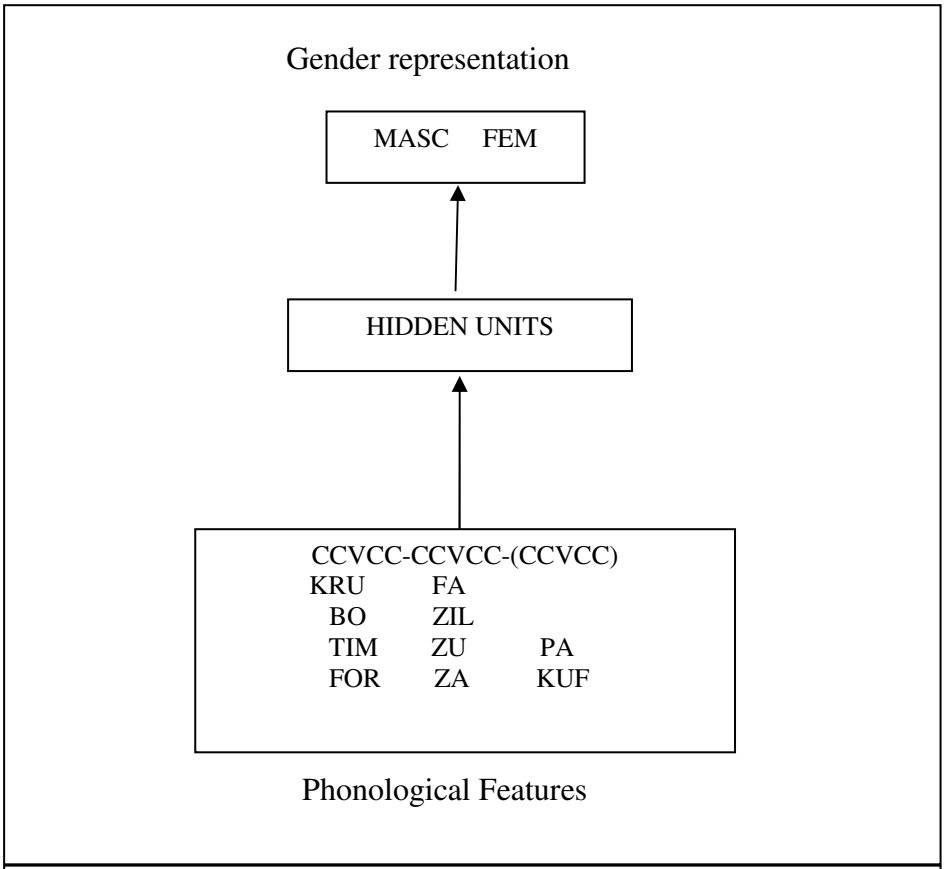


Figure 5.4. Feed-forward network used to simulate gender categorisation in the experiment with artificial suffixes.

5.2.2.3. Training and testing procedure

To capture the variability introduced by the 24 children from the experiment with artificial suffixes, I used 24 groups of feed-forward networks (FFNs), where each group (henceforth: subject-network) represented one child and consisted of five FFNs (following the data contact and task veridicality criteria). The networks were trained and tested for noun gender learning, i.e. they had to classify nouns into the two gender categories. Performance on this categorisation task was measured as the activation of one of the output units: for the feminine nouns the resulting activation of the output units was: "1 0" and for the masculine nouns: "0 1". Out of the five sub-networks within one Subject-network, the first network (henceforth: **Gender-learning network**) was used to pre-train the subject-network with the set of real Serbian words in order to match the children's state of grammatical (more specifically gender) knowledge as closely as possible. This allowed for a more realistic comparison of the network performance on novel words with artificial suffixes at later stages of the simulation with the children's performance in the experiment. The other four networks represent the four experimental sessions from the experiment (henceforth: **Session 1, Session 2, Session 3, Session 4 networks**). For a more detailed presentation of the architecture of the Subject networks see Figure 5.5. Order and counterbalancing of the presentation of nouns were identical to the experiment.

The order of presentation of words was randomised. Each network had a different set of initial weights which were randomised within the interval [0.05; -

0.05][16]. The learning rate for all networks was set to 0.005[17], and the momentum to 0.9.[18] The Gender-training networks were trained for 25 epochs. The number of epochs for the gender-training networks was set to a relatively low number so as not to over-train the network. In other words, the training of the network was stopped just before the point where the difference between targeted and the output activations started reaching asymptote. In this way, I attempted to match the state of the network's "knowledge" of the Serbian noun gender system with the state of knowledge of Serbian children. After training the gender-training network, the weights from this network were transferred to the Session 1-network which was exposed to the first 8 novel nouns in simplex and pseudo-diminutive form. The Session 1-network was then tested for its performance after 50 epochs of training. The weights from the Session 1-network were then frozen and transferred to the Session 2-network which went through the same procedure as the Session 1-network. The same steps were repeated for the Session 3 and 4 networks, so that the Session 4-network contained the accumulated "knowledge" of gender classification from all previous Session-networks as well as from the Gender-training network. Order of training and weight transfer are presented in Figure 5.5.

[16] This interval is recommended by connectionist modellers, as the small initial weights do not commit later stages of learning to the states from the beginning of the simulation (Plunkett & Elman, 1997).

[17] The learning rate parameter ranges from 0 to 1 and it is user-designated in order to determine how much the link weights and node biases can be modified based on change direction and change rate. The higher the learning rate the faster the network is trained. At the same time, the network is more likely to end in local minimum. A local minimum is a point at which the network stabilises on a solution which is not the most optimal global solution. Thus, the learning rates should be set at the lowest possible value.

[18] Momentum also ranges from 0 to 1. This parameter is used to prevent the system from settling into a local minimum. A history of change rate and direction are maintained and used, in part, to push the solution past local minima. A momentum rate set at the maximum of 1.0 may result in training which is highly unstable and thus may not achieve even a local minimum, or the network may take an inordinate amount of training time. If set at a low of 0.0, momentum is not considered and the network is more likely to settle into a local minimum. Thus the momentum should be set at the highest possible level.

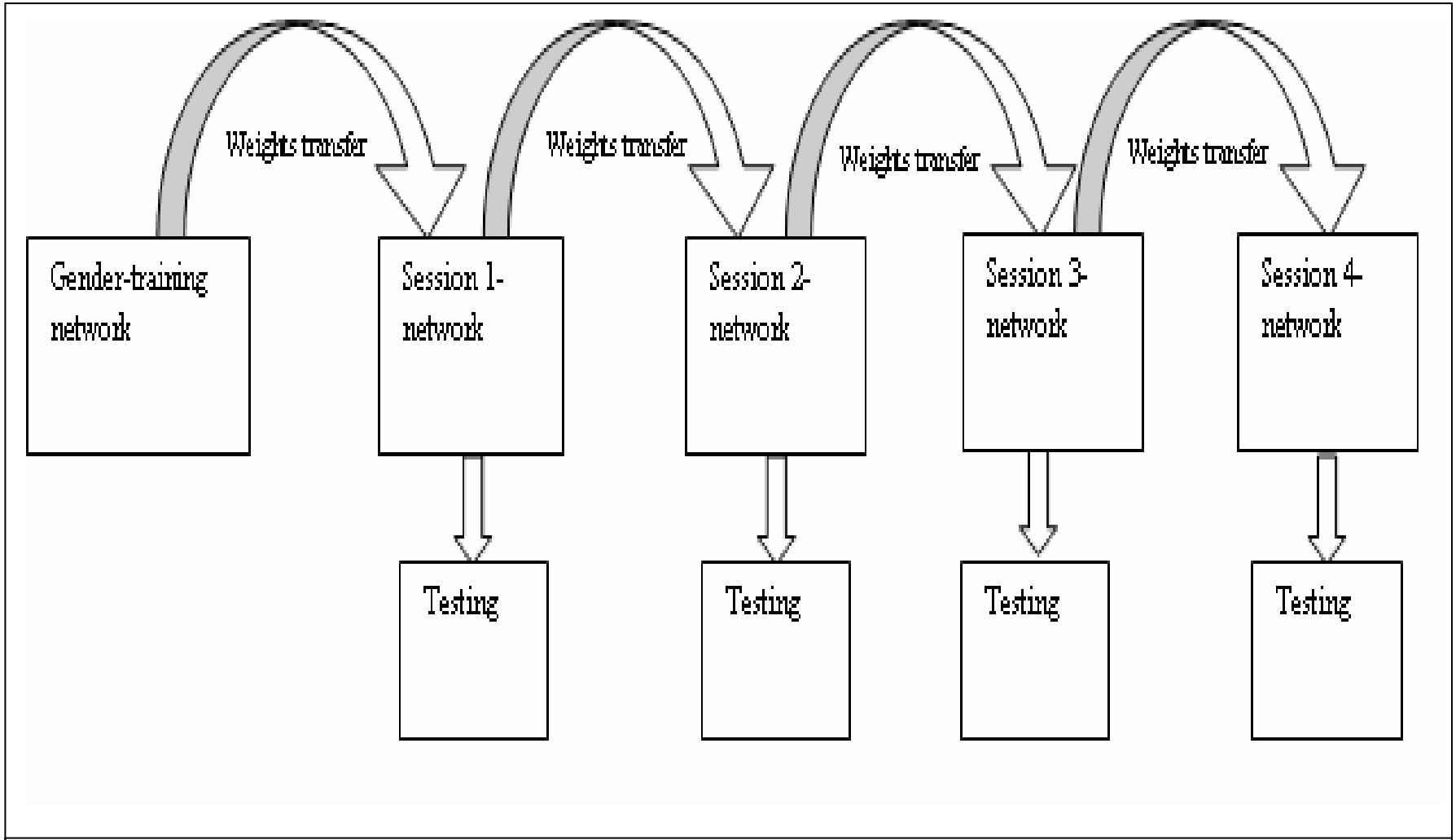


Figure 5.5. Training and testing procedure for the Gender-training network and all Session-networks.

5.2.2.4. Results and discussion

The performance of the Session-networks was measured as the difference between the output activation and the targeted activation value for the output units, e.g. the activation for the word *timza* (feminine gender) was 0.93 for the first unit instead of the targeted 1 and 0.07 for the second unit instead of the targeted 0. Since the sum of the activation of the two units was always around 1, I used the absolute difference between the real activation and the targeted value for the first output unit. In this example, the difference was 0.07 and this value was counted as the dependent variable.

The average difference between output activation and target activation over four sessions for the 24 different subject-networks is presented in Table 5.2.

Table 5.2. Mean difference between output and targetted activations values per subject-network. Standard derivations are given in parentheses.

Session 1	0.09 (0.009)
Session 2	0.08 (0.006)
Session 3	0.07 (0.007)
Session 4	0.06 (0.007)

As with the children, I performed a 4 (session) x 2 (derivational status: simplex vs. pseudo-diminutive) x 2 (gender: feminine vs. masculine) within-subjects ANOVA on the mean absolute difference from the targeted activation values. The analysis yielded three main effects: a main effect of session, $F(3,69) = 63.6$, $p < 0.001$, $\eta^2=0.38$ which indicated that the network performance improved

overall over the four sessions (Figure 5.6.), a main effect of derivation, $F(1,23)=25.3$, $p<0.001$, $\eta^2=0.02$, with better performance with pseudo-diminutive than with simplex words, from the first session onwards, and a main effect of gender, $F(3,69)=6.7$, $p<0.05$, $\eta^2=0.002$, with better overall performance for masculine than for feminine nouns. The analysis also revealed a two-way interaction between derivation and gender, $F(1,23)=289.1$, $p<0.001$, $\eta^2=0.26$, with feminine nouns leading to better performance on the pseudo-diminutives compared to the simplex nouns, and masculine nouns leading to better performance on simplex than on diminutive nouns. The other significant interaction was a three-way interaction between session, derivation and gender, $F(3,69)=13.5$, $p<0.001$, $\eta^2=0.03$, mainly carried by feminine simple nouns in the first session which appeared to be the most complex set of words for the network (see Figure 5.7. and 5.8.)[19].

[19] S.E.M.s are to small to be visible on this scale which has been chosen to maintain comparability with the SRN and the children's data.

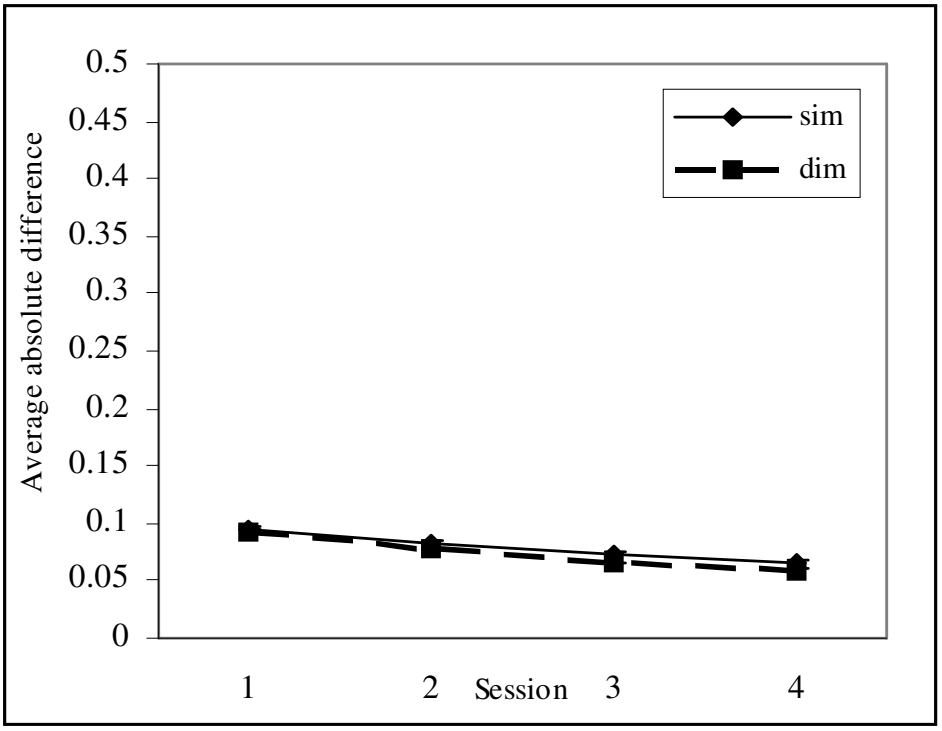


Figure 5.6. Average absolute difference (and 1 S.E.M.)[18] between output and targeted activations over four sessions and two derivations for the whole-word FFN model.

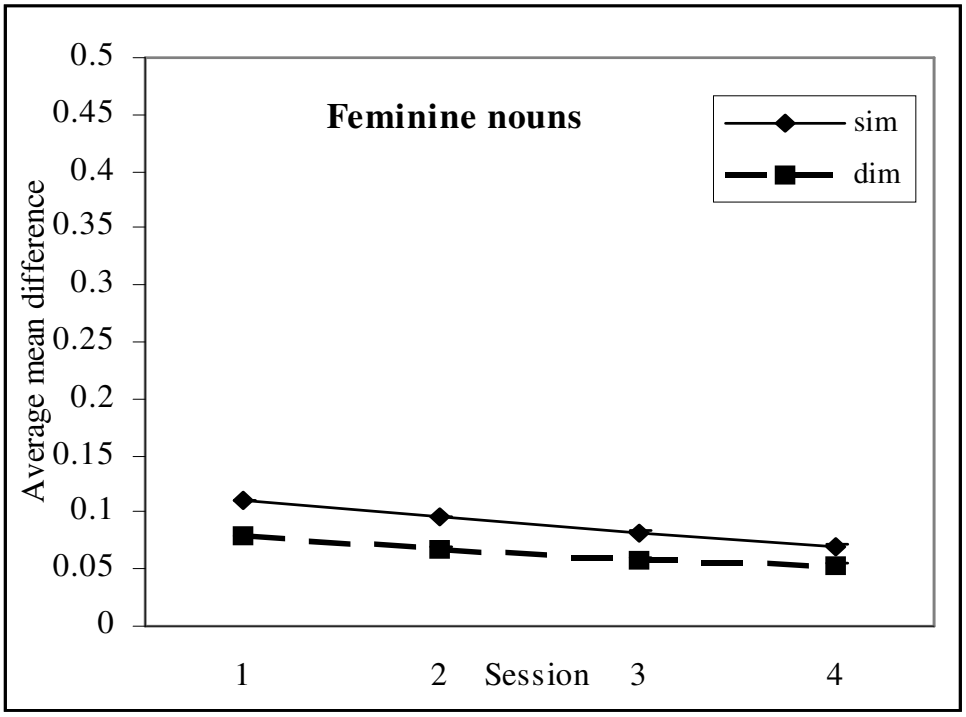


Figure 5.7. Average absolute difference between output and targeted activations over four sessions and two derivations for feminine nouns separately (and 1 S.E.M.)[18] for the whole-word FFN model.

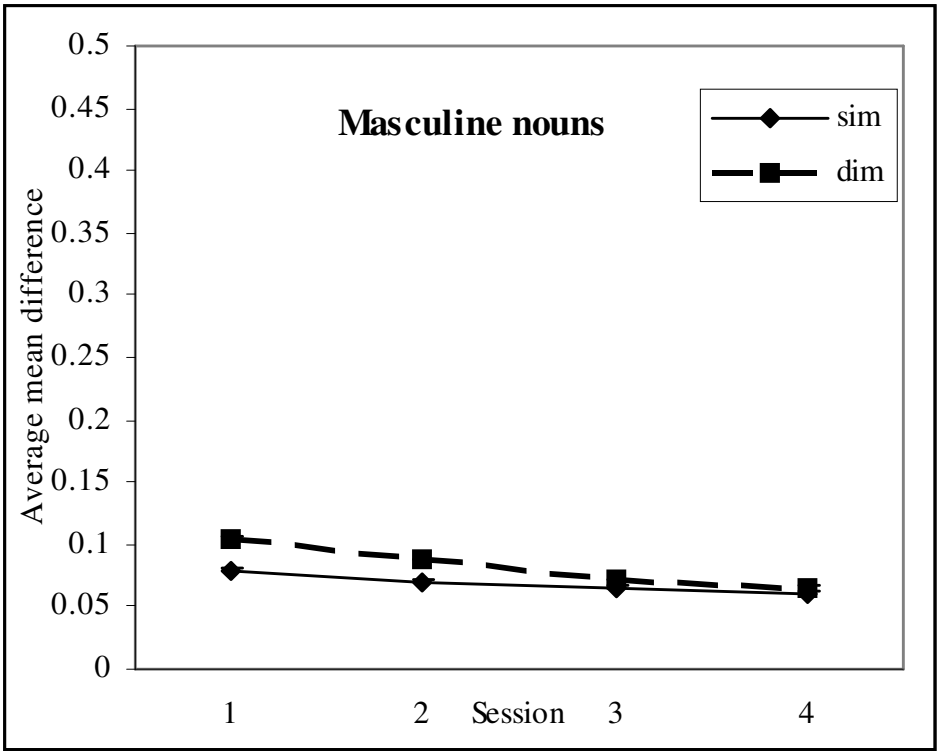


Figure 5.8. Average absolute difference between output and targeted activations over four sessions and two derivations for masculine nouns separately (and 1 S.E.M.)[18] for the whole-word FFN model.

Taken together, these analyses showed that the feed-forward networks were sensitive to the morpho-phonological similarities between words, with very fast emergence of a pseudo-diminutive advantage right from the first session. Moreover, additional analyses showed that the effect of derivation was carried mainly by masculine pseudo-diminutive nouns over all four sessions which was similar to the effect observed with Serbian children, who performed worse with feminine nouns in general. However, the observed crossover between simplex and pseudo-diminutive nouns with the children in Session 2 did not occur with this model. So despite the fact that the FFN performed better with pseudo-diminutive nouns, the overall

pattern obtained with this kind of architecture did not match the pattern observed in the Serbian children.

Probably the most crucial difference between the experimental study and the FFN model was the absence of a temporal component in the presentation of words. In order to check whether the observed difference between children and FFN models is due to this aspect of the network architecture and the way the words were presented to the model, I designed a similar set of 24 subject-networks using a simple recurrent network architecture (SRN). In contrast to the FFN models, where words are presented holistically, in SRN models words are presented sequentially over time. In such models, the weights for every unit (phoneme, syllable or word) are included into the build-up of the representation of the next unit, e.g. the representation of the last syllable in a three-syllabic word is built not only on the activations for that specific unit, but also on the activations for the previous two syllables. This characteristic makes SRN models perceptually more plausible than feed-forward models which may improve the match to the results from the experiment.

5.2.3. Simple recurrent network

5.2.3.1. Network architecture

Each network had a set of input units fully connected to a hidden layer which was fully connected to an output layer and to a set of context units. The model is shown in Figure 5.9.

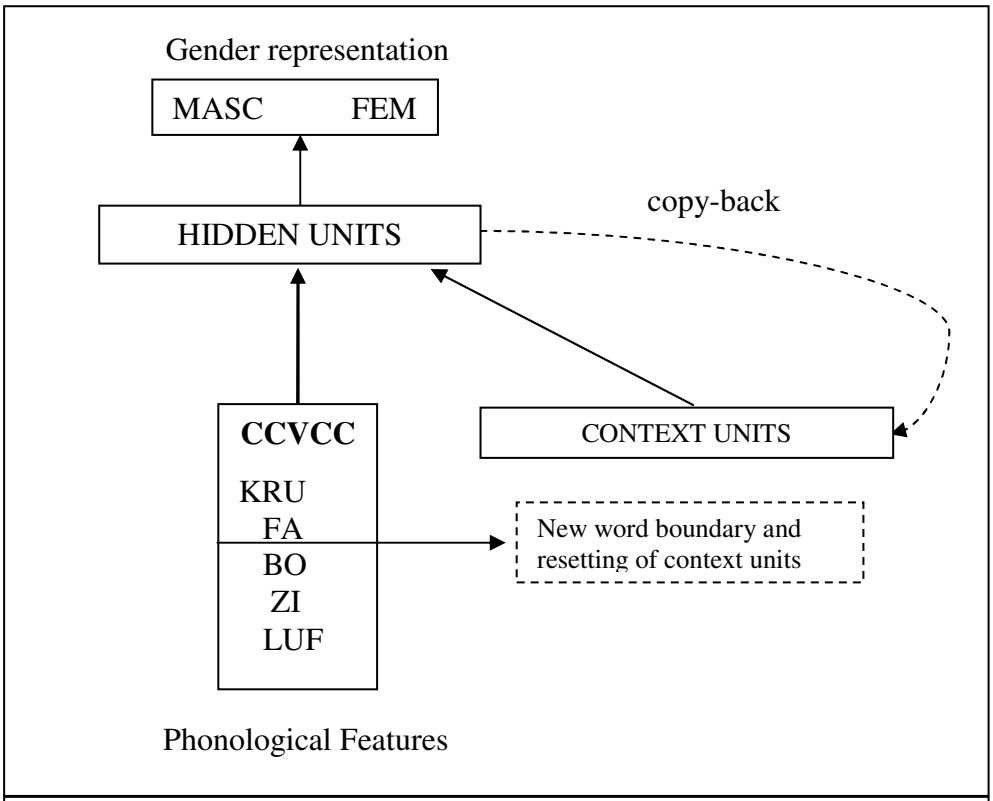


Figure 5.9. SRN model of gender categorisation. The solid lines indicate trainable weights, whereas the dashed line denotes the copy-back weights (which are always 1).

In this model, words were divided into syllables which were vowel centred within a CCVCC frame. Thus, the 80 input units corresponded to the phonological structure of a syllable (Appendix 8).

The output layer had 2 units representing the gender of nouns ("1 0" for feminine and "0 1" for masculine). The hidden and the context unit layer had 20 units. The context units were reset after every second or third syllable marking the end of bi- and three-syllabic words, respectively.

5.2.3.2. Materials

The networks were trained and tested with the same set of words as the feed-forward whole word network.

5.2.3.3. Training and testing procedure

The SRN model had the same task, to categorise nouns into two gender categories. As with the FFN whole word model, I created 24 groups of simple recurrent networks (SRNs), where each group (henceforth: subject-network) represented one child and consisted of five SRN networks (one for pre-training the network with real gender nouns and four for training and testing of novel simplex and pseudo-diminutive nouns). The order of presentation of the networks, the counter-balancing of different groups of nouns, and the weight transfer from one network to another were organised in the same way as with the feed-forward network (see Figure 5.5.).

Since the words were split into syllables, data and output files contained 256 vectors for the first pre-training network which represented 120 words and 20 vectors for the remaining four networks which represented 8 novel words each. Additionally, each network had a different set of initial weights, randomised within an interval of [0.05; -0.05]. The learning rate for all networks was set to 0.005, and the momentum to 0.9. The gender-training networks were trained for 100 epochs. As for the whole-word FFN model, number of epochs for the gender-training network was set to a relatively low number in order not to over-train the network, i.e. training of the network was stopped just before the point where the difference between targeted and output activations started to reach asymptote. The session-networks were tested after 200 epochs of training.

5.2.3.4. Results and discussion

Similar to the FFN model, network performance was measured as the absolute difference between the targeted values for the first output unit ("1" for feminine and "0" for masculine nouns) and the real output activation values. As described previously, the words in the SRN model were divided into two or three syllables, with the last syllable carrying cumulative information from all previous syllables. Given this feature of SRN networks, I analysed only the activation values obtained after the presentation of the last syllable, since this value could be treated as the final activation at the end of processing a word. Thus, the dependent variable in the SRN model was the average absolute difference between targeted and output values for the first output unit on the last syllable of the word which was the second or third syllable depending on the length of the words. Table 5.3 presents the mean

absolute differences between output and targeted values for the last syllables of the words.

Table 5.3. Mean absolute differences between targeted and output values per SRN network. Standard derivations are given in parentheses.	
Session 1	0.16 (0.101)
Session 2	0.13 (0.052)
Session 3	0.15 (0.101)
Session 4	0.14 (0.105)

As in the experimental study with Serbian children and the whole-word FFN simulations, I performed a 4 (session) x 2 (derivational status: simplex vs. pseudo-diminutive) x 2 (gender: feminine vs. masculine) within-subjects ANOVA on the mean absolute differences between the targeted and the output values. The analysis yielded a significant two-way interaction between session and noun derivation, $F(3,69) = 3.1$, $p < 0.05$, $\eta^2=0.02$, indicating that a pseudo-diminutive advantage emerged during Session 3 (see Figure 5.9.). Separate 2 (derivation) x 4 (session) ANOVAs for the two genders revealed only a significant two-way session x derivation interaction for the masculine nouns, $F(1,23)=3.78$, $p<0.05$, $\eta^2=0.04$ which was in contrast to the children's performance, where a crossover between simplex and pseudo-diminutive nouns occurred for both genders (see Figures 5.10. - 5.12.).

Separate ANOVAs with gender and noun derivation as within-subjects factors were conducted for each session to qualify the interaction. For Session 1, the analysis only revealed a marginally significant effect of derivation, $F(1,23)=3.89$,

p=0.06, η^2=0.02, indicating superior processing of simplex nouns. For Session 2, none of the effects were significant. For Session 3, there was a significant effect of derivation, F(1,23)=5.19, p<0.05, η^2=0.04, indicating superior processing of pseudo-diminutive nouns. Finally, for Session 4, the only significant effect was an effect of gender, F(1,23)=5.86, p<0.05, η^2=0.12, with better performance for masculine nouns, similar to the children's performance in the last session.

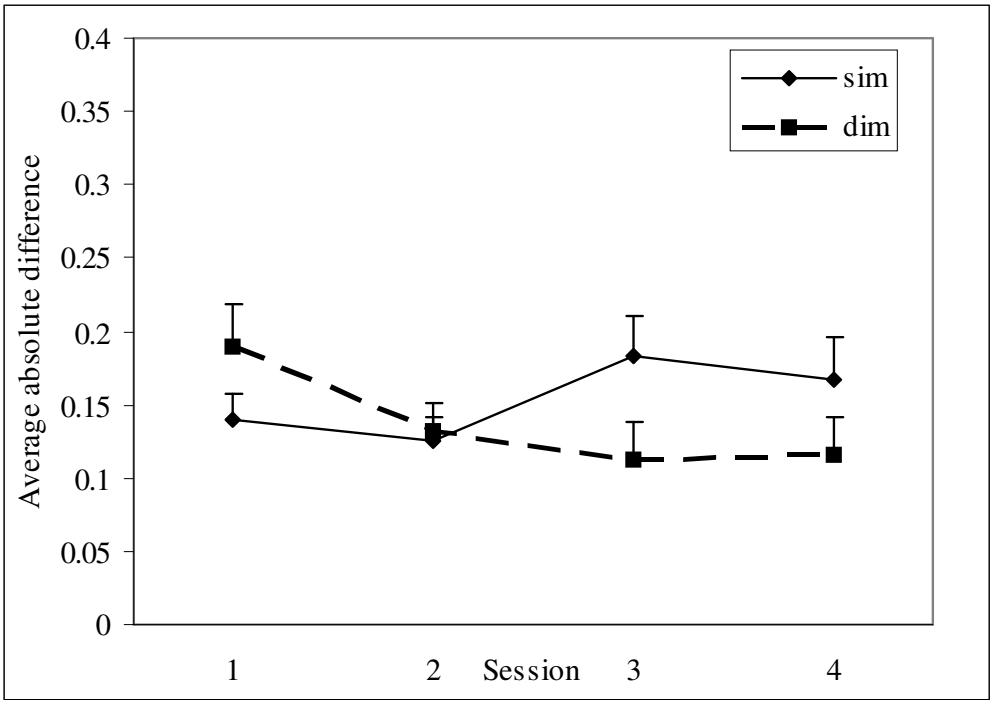


Figure 5.10. Average absolute difference (and 1 S.E.M.) between output and targeted activations over four sessions, two derivations and two genders for the SRN model.

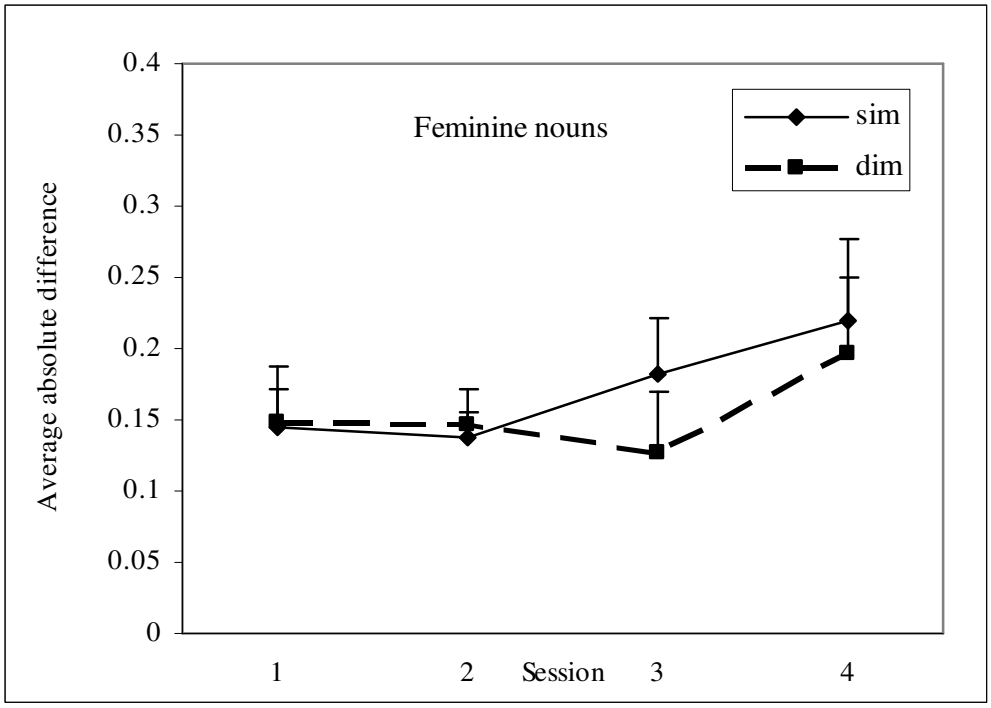


Figure 5.11. Average absolute difference (and 1 S.E.M.) between output and targeted activations over four sessions and two derivations for the feminine nouns separately in the SRN model.

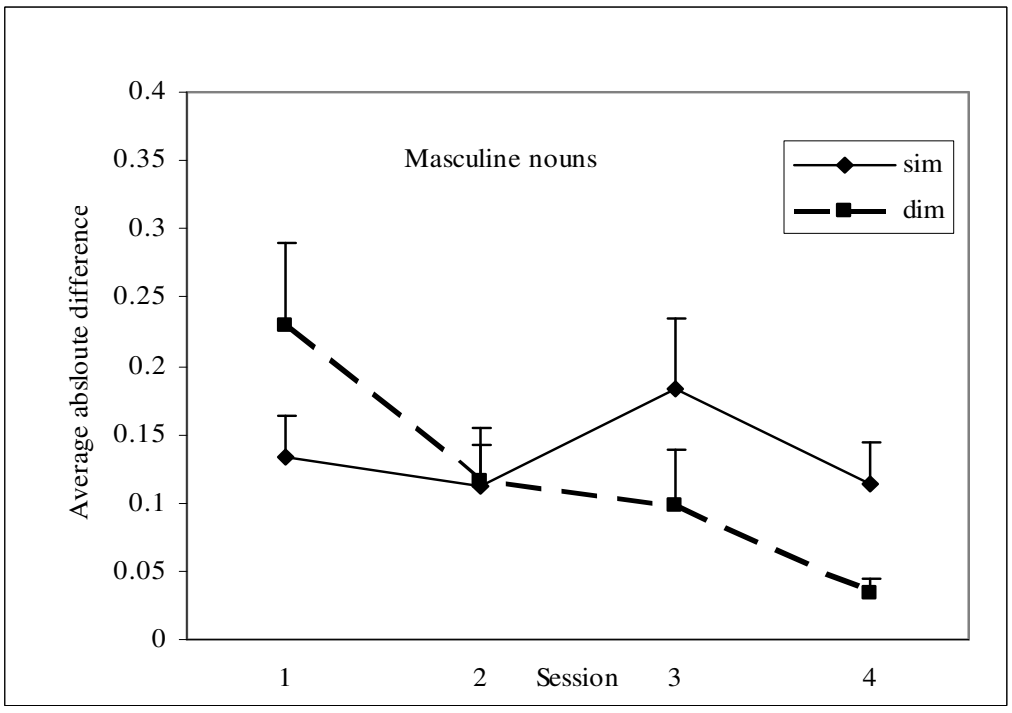


Figure 5.12. Average absolute difference (and 1 S.E.M.) between output and targeted activations over four sessions and two derivations for the masculine nouns separately in the SRN model.

In sum, the analysis of the SRN model performance showed a clear pseudo-diminutive advantage that emerged in Session 3. In addition, the gender effect observed with the children and the whole-word FFN model emerged in the SRN model as well, with only masculine nouns exhibiting a significant cross-over between simplex and pseudo-diminutive forms. Most importantly, the SRN model showed a much better fit to the children's data in comparison to the whole-word FFN model. In both models, the last parts of the presented words were morpho-phonologically the most salient parts which may have provided cues for gender classification. This view goes along with the general notion that children are also highly sensitive to ends of words in the process of grammatical categorisation (Slobin, 1973). In the case of Serbian nouns, the endings of words provide the information that *–a* endings are associated with feminine nouns and consonant endings with masculine nouns. The observed difference between the two types of models might have been due to the different ways of presentation not only of the last syllable but also of the entire word. In order to test to which extent both models rely on the last syllable in categorising nouns as feminine or masculine, I constructed one more FFN model in which I presented only the last syllables of words. If the last-syllable FFN model exhibits patterns similar to the SRN model, this would indicate that both children and the SRN model do rely on the endings of the words in the process of gender categorisation. If on the other hand, the overall pattern for the last-syllable FFN model is similar to the whole-word FFN, this would indicate that the children and the SRN model were using some additional information for classifying nouns as feminine or masculine, for example phonological regularities within the first and the second syllable. The following part

of this chapter will provide a description of the simulation which included only the last syllable of the words, and discuss further the observed results in the context of the previous experimental and simulation results on gender categorisation of simplex and pseudo-diminutive nouns.

5.2.4. Feed-forward network with last syllable only

5.2.4.1. Network architecture

The network architecture was similar to the whole word FFN model, apart from the number of input units (240 for the whole word model vs. 80 for the last syllable model). The 80 input units in this model corresponded to the phonological representation of the last syllable with phonemes represented as a sequence of 16 binary phonetic features (Figure 5.13).

The output layer had 2 units representing the gender of nouns ("1 0" for feminine and "0 1" for masculine). The hidden layer had 20 units.

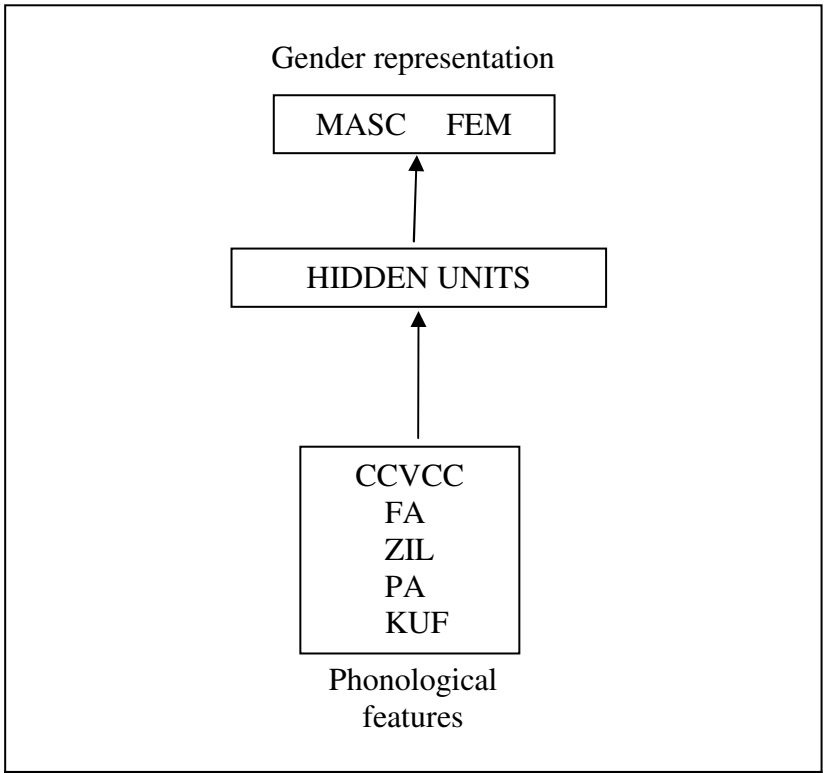


Figure 5.13. The last-syllable FFN model of gender categorisation.

5.2.4.2. Materials

From both sets of words used for training and testing of the previous models I extracted the last syllables (e.g. /fa/ for /kru-fa/, /pa/ for /mom-pu-pa/, /zil/ for /bo-zil/ and /luf/ for /bo-zi-luf/. All other aspects of the materials were identical to the whole-word FFN model.

5.2.4.3. Training and testing procedure

I constructed 24 subject-networks which consisted of a gender-training network and four session-networks per subject. The order of training and testing, the counterbalancing of different groups of novel simplex and pseudo-diminutive nouns over 24 different subject-networks as well as all other parameters including the number of epochs were identical to the whole-word FFN model (see Figure 5.5).

5.2.4.4. Results and discussion

The network performance was again measured as the absolute difference between the targeted values for the first output unit ("1" for feminine and "0" for masculine nouns) and the obtained output activation values. The average absolute differences between target and output values for the first output unit over four sessions are presented in Table 5.4.

Table 5.4. Mean absolute differences between target and output values per last syllable FFN. Standard derivations are given in parentheses.	
Session 1	0.11 (0.007)
Session 2	0.09 (0.006)
Session 3	0.08 (0.007)
Session 4	0.07 (0.007)

A 4 (session) x 2 (derivational status: simplex vs. pseudo-diminutive) x 2 (gender: feminine vs. masculine) within-subjects ANOVA yielded a significant main effect of session, $F(3,69) = 142.9$, $p < 0.001$, $\eta^2=0.51$ with the network performance improving over all four sessions, a main effect of derivation, $F(1,23) = 242.3$, $p < 0.001$, $\eta^2=0.05$ indicating that the pseudo-diminutive advantage was present at all sessions. The analysis also revealed the following two two-way interactions: session x gender, $F(3,69) = 3.41$, $p < 0.05$, $\eta^2=0.003$, mainly due to the feminine nouns showing inferior performance in Session 1 compared to masculine nouns, and derivation x gender, $F(1,23) = 278.2$, $p < 0.001$, $\eta^2=0.15$, indicating that the derivation effect was mainly carried by feminine simplex nouns, as well as a three-way interaction between session, derivation and gender, $F(3,69) = 3.37$, $p < 0.05$, $\eta^2=0.008$, mainly carried by feminine simplex nouns in the first session (see Figures 5.14.-5.16.).

To conclude, the last-syllable FFN model exhibits a pseudo-diminutive advantage starting from the first session. Moreover, feminine pseudo-diminutive nouns were created phonologically more transparent than masculine nouns, probably due to less invariance at the end of words. In addition, the observed pattern

of performance of the last-syllable FFN model was quite different from the children's performance and almost identical to the whole-word FFN model, suggesting that the last syllables were mainly responsible for the pseudo-diminutive advantage observed in both FFN models, but not for the performance of the SRN model and the children.

The last part of this chapter will provide a more detailed analysis on what phonological regularities of the last, but also of the first or second syllable may be responsible for the emergence of the pseudo-diminutive advantage observed both with children and the SRN.

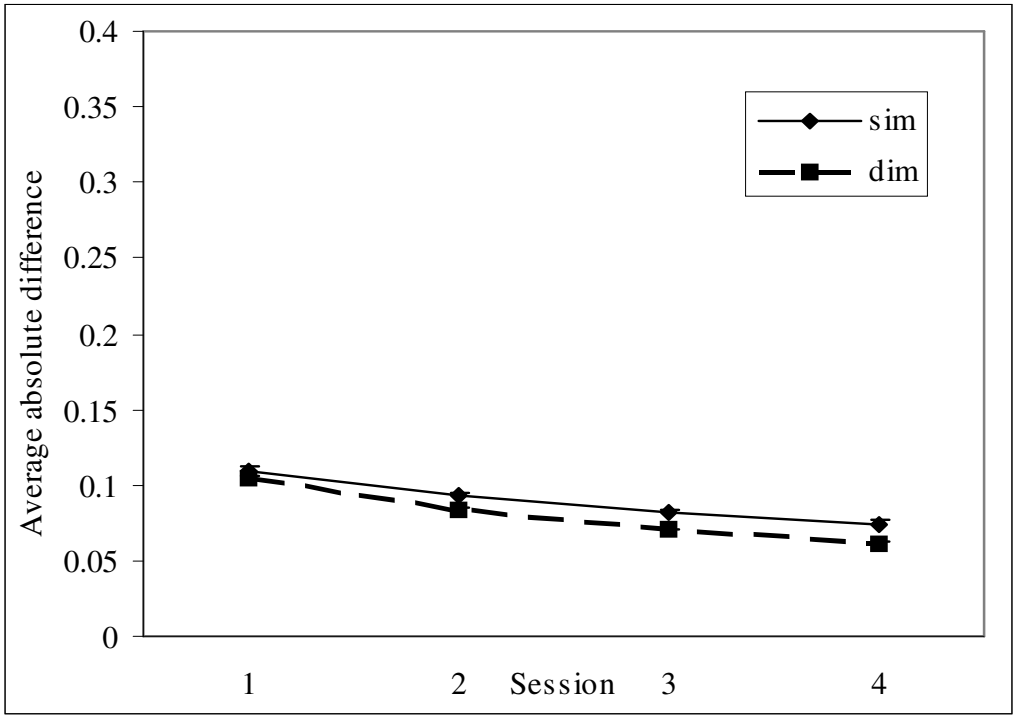


Figure 5.14. Average absolute difference (and 1 S.E.M.) between output and targeted activations over four sessions and two derivations for the last-syllable FFN model.

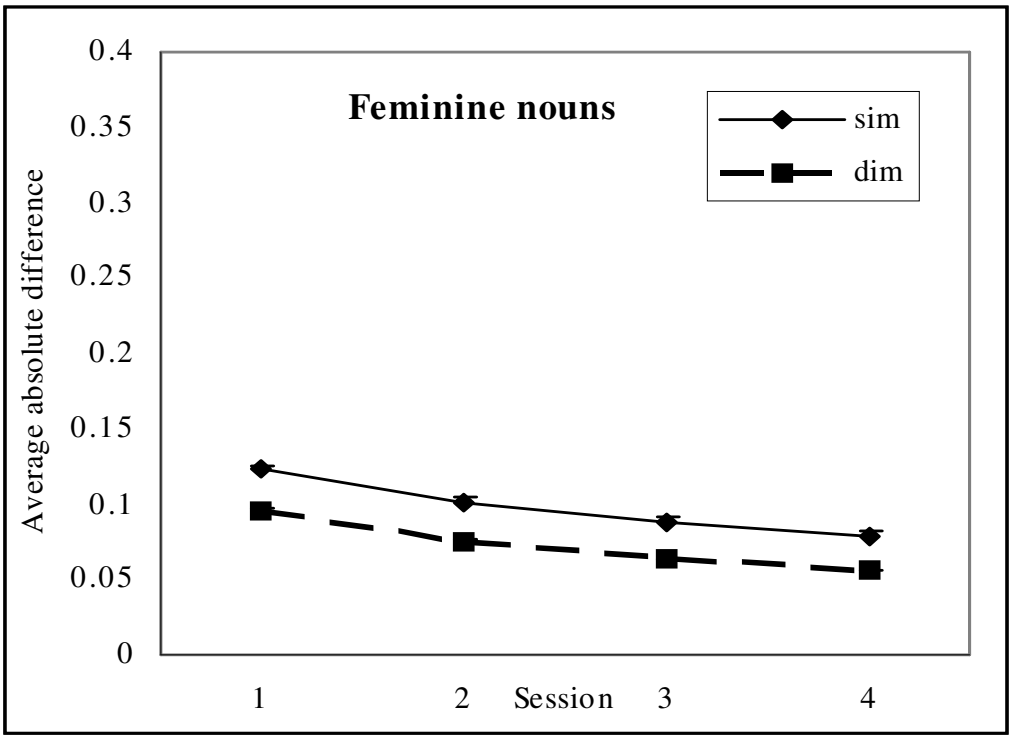


Figure 5.15. Average absolute difference between output and targeted activations over four sessions and two derivations for feminine nouns separately (and 1 S.E.M.) in the last-syllable FFN model.

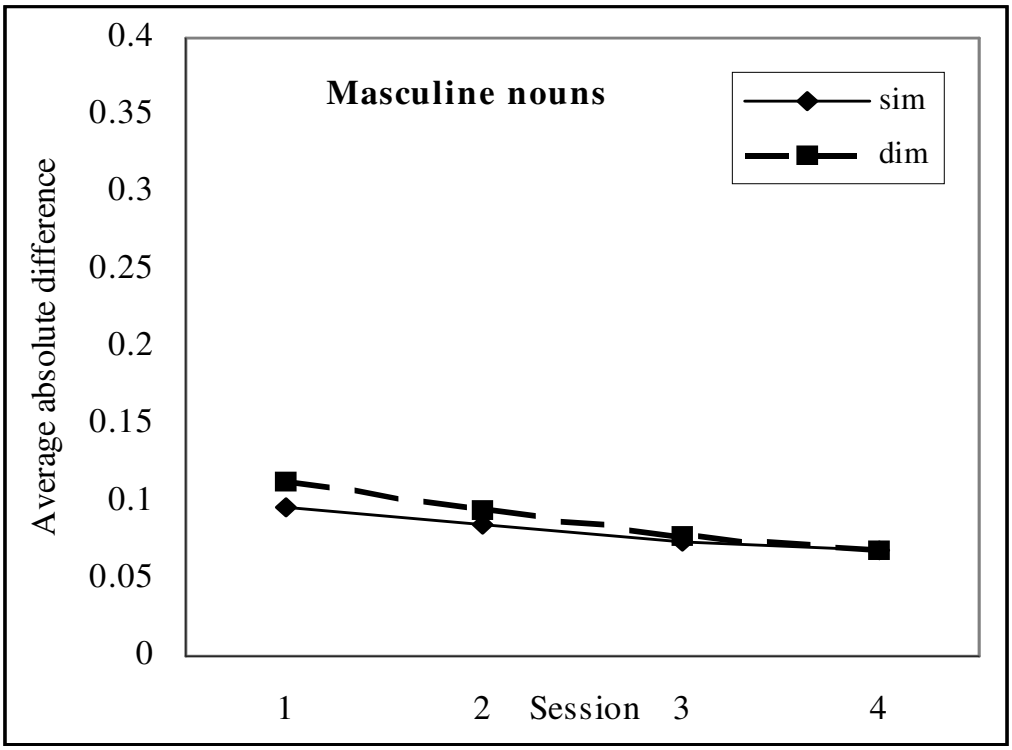


Figure 5.16. Average absolute difference between output and targeted activations over four sessions and two derivations for masculine nouns separately (and 1 S.E.M.) in the last-syllable FFN model.

5.2.5. Discussion of the children's and the networks' performance

The experiment introducing artificial suffixes to Serbian children and the neural network simulations of the experimental results showed that both children as well as networks are sensitive to the morpho-phonological regularities in words. This sensitivity allows for a relatively fast clustering of morpho-phonologically similar words to which inflectional changes are applied. Additionally, the results from the experiment and the neural network simulations (both SRN and FFN) showed that the frequency of diminutives is not as important a factor for the emergence of clusters of words which will become facilitating elements in the process of morpho-syntactic inflectional marking of novel items.

The superior fit of the SRN suggests that this learning process is based on a sequential build-up of representations of the entire word, allowing the system to exploit the predictive power of word stems to anticipate regular endings. In other words, the SRN learns the distributional patterns in the 1st and 2nd syllables in addition to the regularities in the 3rd syllables. Further analyses of the structure of novel words used in the experiment showed that the pseudo-diminutives comprised additional cues for gender categorisation in addition to the ending of words, like the vowel *–u* at the end of the second syllable of feminine pseudo-diminutives and the absence of a consonant in the coda of the second syllable of masculine pseudo-diminutives (see Table 5.5).

CCVCC	CCVCC	CCVCC	CCVCC	CCVCC	CCVCC
TIM	ZA		BO	ZIL	
TIM	ZU →	PA	BO	ZI →	LUF

Table 5.5. Change of word structure in pseudo-diminutives in comparison to their simplex counter parts.

Moreover, I observed that in the first syllable the usual pattern for masculine nouns is CV and for feminine nouns CVC (see Table 5.6). Detailed analyses of the syllabic structure of the 120 nouns used for pre-training in the simulations revealed that within the group of real Serbian nouns the CVC structure in the first syllable is exclusively associated with feminine nouns, and that masculine nouns predominately start with a CV pattern (see Table 5.7). On the other hand, the sample of novel nouns contained a group of pseudo-diminutive masculine nouns which ended in the vowel /u/ in the second syllable. Thus, nouns which started with a non-typical pattern in the first syllable (like *kru-fu-pa* (CCV-CV-CV) which is non-typical beginnings for feminine nouns) or contained a non-typical vowel in the second syllable (like *pa-gu-luf*, where *-u* is a non-typical vowel for masculine nouns), may present problems both for the children and for the SRN networks in the gender categorisation task. An item-based analysis of the performance for each word showed that both children and SRN models exhibited quite different performance for the different words in the first session. As predicted, words with a non-typical phonological structure for the set of novel nouns were more difficult for the children and the SRN model, e.g. feminine nouns like *kru-fu-pa* (CCV-CV-CV) and masculine nouns like *pa-gu-luf* (CV-C/*u*/-CVC) (see Figure 5.17. for feminine nouns and Figure 5.18. for masculine nouns). In comparison to that, the feed-forward networks displayed homogeneous performance for each item not only in the first, but also all other sessions, probably due to the fact that it relied on a comparison of the last syllables of nouns (see Figure 5.17. for feminine nouns and Figure 5.18. for masculine nouns).

To summarise, both the comparison between the children's and the networks' performance as well as an inspection of performance on individual nouns

showed that a network that is sensitive to word-internal distributional patterns in addition to the ending regularities matches the empirical results better.

The last part of this thesis will provide an overview of all corpus and experimental results from this study, as well as a more general discussion of the observed effects in the general context of acquisition and processing of noun morphology in complex morphological systems.

Pseudo-diminutives: **1st syllable:** Masculine: 1 CCVCC, 14 CCVCC, 1 CCVCC; Feminine: 9 CCVCC, 3 CCVCC, 3 CCVCC	*Pseudo-diminutives:* **2nd syllable:** Masculine: 5 /u/, 11 other vowel Feminine: 16 /u/

Table 5.6. Phonological structure of pseudo-diminutive words for the 1st and 2nd syllables.

Table 5.7. Phonological structure of 120 pre-training real Serbian nouns for the 1st syllable.

1st syllable	cCVcc	cCVCc	ccVcc	ccVCc
Masculine%	94.2	0.0	1.9	3.8
Feminine%	78.0	16.9	5.1	1.7

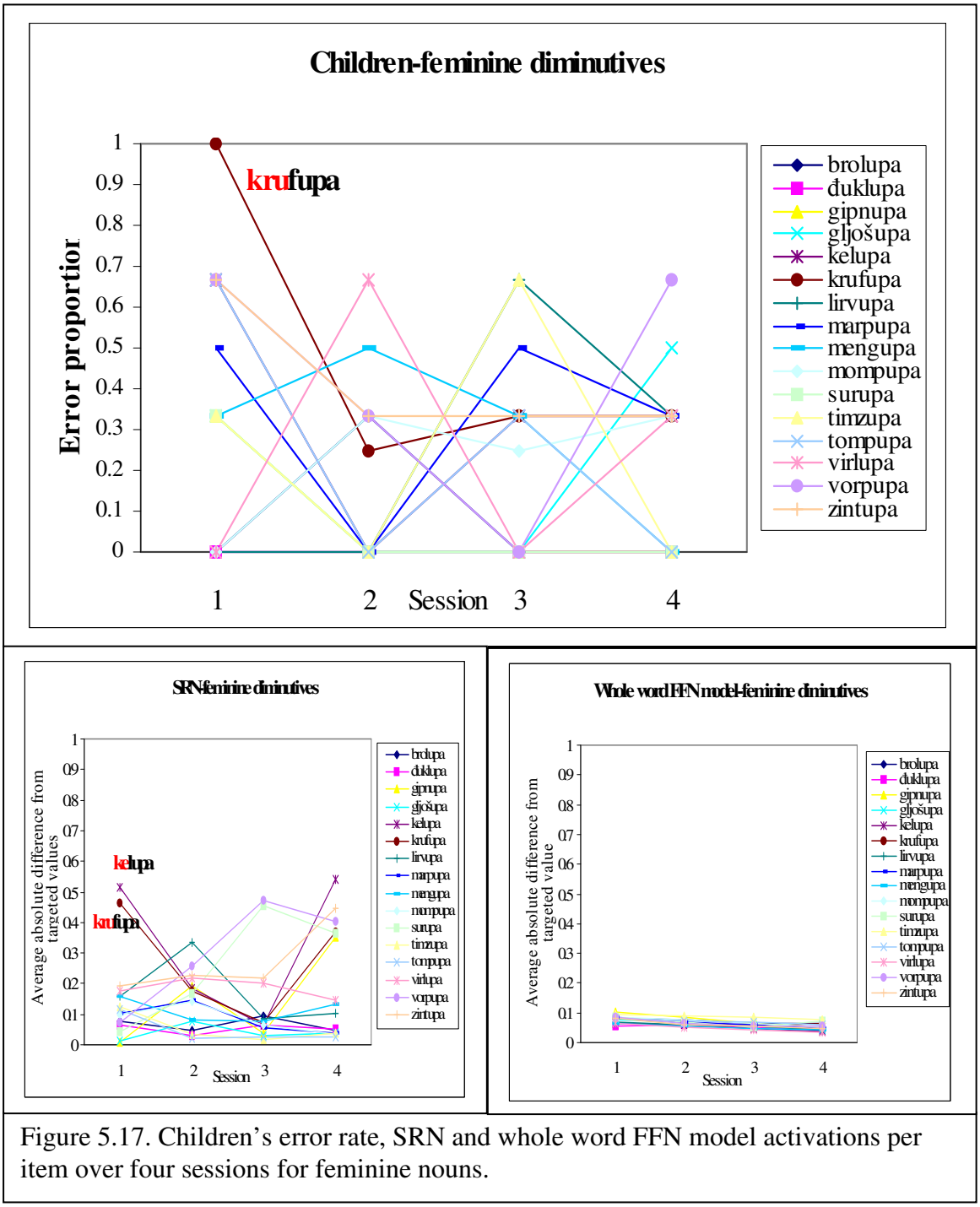


Figure 5.17. Children's error rate, SRN and whole word FFN model activations per item over four sessions for feminine nouns.

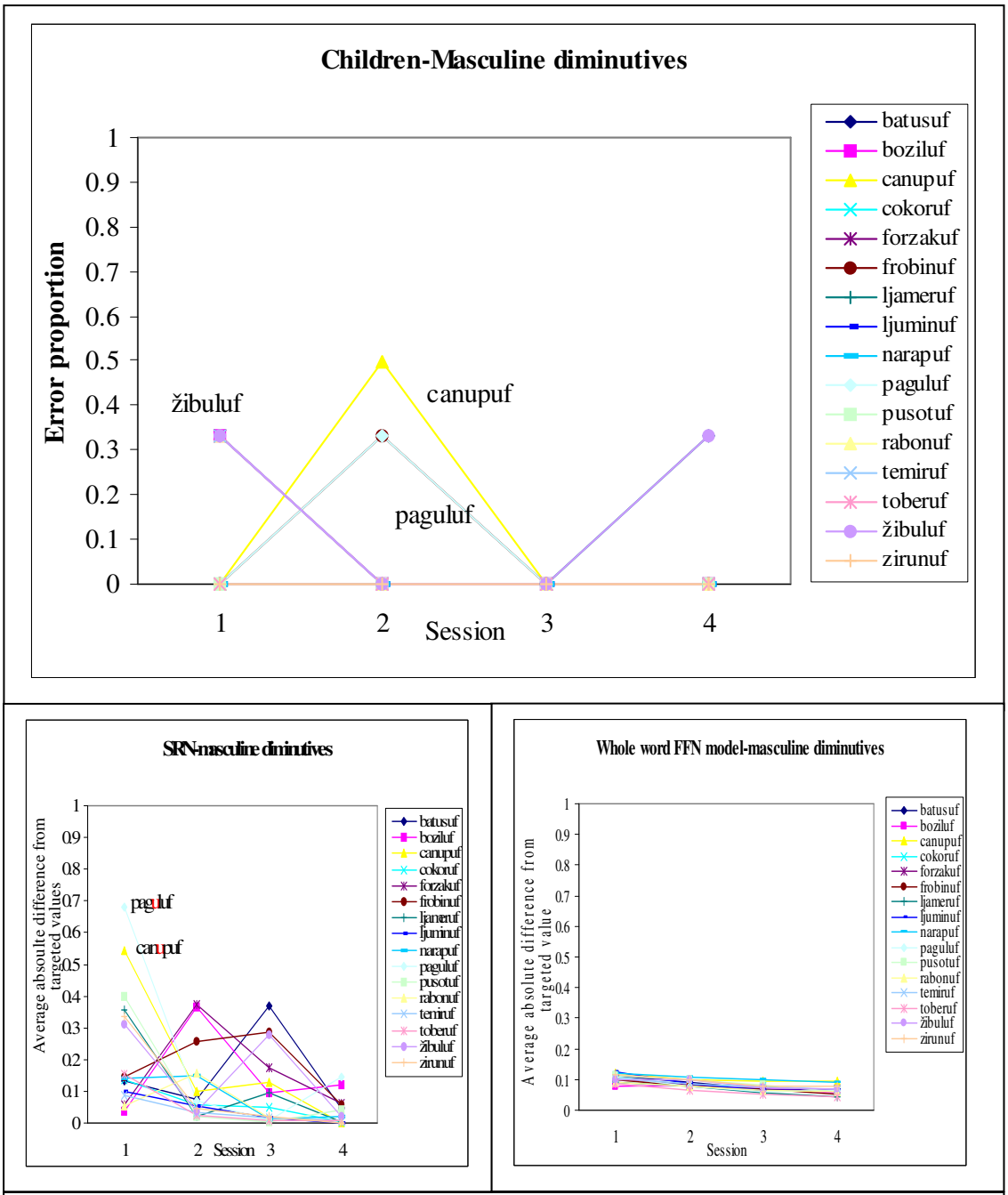


Figure 5.18. Children's error rate, SRN and whole word FFN model activations per item over four sessions for masculine nouns.

Chapter 6

6. Conclusions and general discussion

The main goal of this thesis was to explore in more detail the facilitating effect of diminutives on the acquisition of noun gender and case categories in Serbian, a South-Slavic language with complex inflectional morphology. This study was based on several assumptions on the role of CDS in the acquisition of first language, described in the Introduction. In sum, CDS is a special register evolved primarily for parents (and possibly other adults) to establish an emotional connection with infants and young children, to control children's arousal, as well as to elicit their attention (Fernald, 1992). CDS also tends to provide prosodic, phonological and distributional cues to linguistic structure (Morgan & Demuth, 1996; Monaghan, Christiansen & Charter, submitted). In addition, some of the pragmatic features of this register like the diminutives may lead to a regularisation of morpho-syntactic properties in the input presented to children (Kempe et al, 2001; 2003). Given that children are highly sensitive to regularities, the simplifications observed in CDS can additionally facilitate first language learning.

The present study was divided into two parts. The first part, mainly covered by Chapter 3, presented the first quantitative and qualitative corpus analysis of the distribution of diminutives in Serbian CDS. This corpus analysis was based on one of the biggest corpora so far used in cross-linguistic research on diminutives, allowing for a more detailed description of parental production as a function of age and gender of the children, as well as the gender of the parents. The second part, presented in Chapters 4 and 5, described a set of experimental studies and neural network simulations developed to replicate and further explore the factors underlying the facilitating effect of diminutives on the learning of noun morphology

which had been observed in a set of cross-linguistic studies for Russian, Polish and Lithuanian.

The main findings of the corpus analysis, experimental studies and neural network simulations were:

1) Serbian parents tend to use considerably fewer diminutives when addressing their children (7% of diminutives out of the total number of noun tokens and 11% out total number of lemmata) in comparison to Russian, Polish or Lithuanian parents (20-45% of diminutives), despite the similarities in noun morphology and diminutive productivity in all Balto-Slavic languages.
2) Serbian children exhibited a strong diminutive advantage for both gender agreement and case marking in the same range as Russian children, indicating that the morpho-phonological homogeneity within the cluster of diminutives may play as important a role as their frequency in the grammatical categorisation of novel nouns. Moreover, the experimental study with pseudo-diminutive suffixes showed that the categorisation of nouns into gender categories which capitalises on the morpho-phonological homogeneity of a word cluster, emerges relatively fast, and therefore does not require a high frequency of words belonging to such a cluster in the input. This can explain why the diminutive advantage in Serbian was found despite the much lower frequency of diminutives in CDS compared to Russian, Polish and Lithuanian.
3) All three neural network models showed that the rapidly emerging categorisation of nouns into gender categories can be explained with a simple associative learning mechanism sensitive to the morpho-phonological similarities of nouns. Two network architectures were used to try to model

the experimental data: a feed-forward network which received the phonological structure of the entire noun as input, and a simple recurrent network (SRN) which was designed to capture the sequential nature or language processing by processing the nouns syllable by syllable. The superior fit of the SRN to the experimental data on gender categorisation of pseudo-diminutive nouns suggests that gender learning is based on a very fast sequential build-up of representations of the entire word, allowing the system to exploit the predictive power of word stems to anticipate regularised endings.

These empirical findings and the results of the neural-network simulations for Serbian have extended our understanding of the nature of mechanisms involved in the acquisition and processing of complex morphology systems in two ways:

1. First, the results augmented the growing number of studies demonstrating that noun morphology is learned and processed through a single-route associative learning mechanism based on distributional and morpho-phonological features of nouns. This approach also assumes that both children and adults are sensitive to various levels of generality present in the system. In other words, morphology processing can be seen as a continuum from low-level generalisations of morpho-phonological similar clusters of words to very abstract generalisations which capture general regularities in the language (Bybee, 1995; Albright & Hayes, 2003; Dabrowska, 2004, submitted; Hay & Baayen, 2005 for a general overview). As described in the Introduction, this theoretical model is usually contrasted to the dual-route account (Pinker, 1991; Prasada & Pinker, 1993; Marcus et al., 1995) which presupposes the existence of two separate architectural components: one for

regular forms and the other for exceptions, with the assumption that regular novel items will always be processed with similar accuracy by applying the same default rule. In this context, the experimental studies on grammatical categorisation of novel nouns in Serbian and other Balto-Slavic languages directly question the dual-route account: given that both simplex and diminutive novel nouns were from the same set of regulars, and follow the same rule for gender categorisation, they should have been processed in the same fashion, and a diminutive advantage is difficult to explain. The fact that the diminutive advantage is such a robust phenomenon suggests that languages like Serbian, Russian, Polish and Lithuanian, where there are many complex morphological phenomena such as the existence of various declension and derivational paradigms, do not lend themselves to a straightforward 'rules vs. exceptions' dichotomy (Dabrowska, 2004). The cross-linguistically observed difference in processing between simplex and diminutive nouns can only be explained if gender categorisation and case marking are based on sensitivity to the distribution of morpho-phonological cues in the input. This allows children to operate with low-level generalisations of grammatical categories, before acquiring more abstract ones. In an experimental study on the effect of diminutives on the acquisition of cases in Polish, a diminutive advantage was observed for children up to the age of four, but disappeared in adults which seems to indicate that adult Polish speakers operate on the basis of abstract categorisations of the case marking system (Dabrowska, 2006). On the other hand, this result might have been a consequence of the task which was used in this study (production of novel nouns in a familiar sentence context). Recent cross-linguistic experiments, using different tasks or more complex constructions, showed that even within the class of 'regulars', adult language users can perform more consistently for items that fall

within ‘islands of reliability’ (Albright & Hayes, 2003) or morpho-phonologically densely populated clusters of words (Dabrowska, 2004; submitted). This suggests that categorisations on different levels can persist even in the adult system. In this context, it would be interesting to see whether more sensitive online processing measures can reveal a similar advantage for novel diminutive nouns in adult native speakers of Serbian and other Balto-Slavic languages.

2. In addition to the contribution to the general theoretical framework on processing of inflectional morphology, the experimental studies and neural network simulations in this thesis focused more specifically on the exploration of the acquisition of gender categories in Serbian. The main finding of these studies showed that both children and neural networks used regular endings of words for gender categorisation which is in line with the idea that children are highly sensitive to ends of words during the process of language acquisition (Slobin, 1973). Also, the neural network simulations demonstrated that sensitivity to the endings of words is an emergent property. The system first starts with the exploration of morpho-phonological regularities in the entire word, but then gradually ’zooms’ into the endings of words as the most informative part. In this regard, the most interesting result was the superior fit of the SRN to the experimental data on gender categorisation of pseudo-diminutive nouns which suggests that children are using the regularities present in the stems of the words, in addition to the regularities in the endings. Considering morpho-phonological regularities in the stems of nouns might be especially helpful in systems which exhibit a high level of syncretism, i.e. use identical endings for different functions (see Chapter 2 for more details). To further clarify the use of phonological information in morphology acquisition,

future studies on the nature of the mechanism underlying gender categorisation should include the following points:

a) A more detailed corpus analysis of phonological properties of words than the one provided in Chapter 5 is needed to explore the morpho-phonological cues in the stems (syllabic structure, number of consonants and vowels, etc.) which potentially may help to differentiate nouns into feminine, masculine or neuter;

b) Corpus studies should be accompanied by further experimental studies and neural network simulations to explore which of the potential cues are really used by children in the process of gender categorisation;

c) There is a great need for further exploration of distributional cues used in the process of gender categorisation. The gender effect (masculine nouns were marked with adjectives more accurately) which was observed in both Serbian and Russian gender agreement experiments, indicated that children's performance might be affected by the relatively large number of ambiguous hypocoristic masculine nouns like *meda* 'bearHYP (MASC)', *deka* 'grandfatherHYP (MASC)', etc. It would be interesting to see whether this effect is a consequence of frequent non-matching combinations with pronominal words, e.g. *lep meda* 'beautifulADJ (MASC) bearHYP (MASC)' where a masculine adjective ending is combined with a masculine hypocoristic the ending of which is similar to feminine nouns. Also, the superior fit of the SRN with the experimental data on gender marking of pseudo-diminutives indicates that perhaps some of the novel feminine nouns were closer to the phonological space of masculine nouns, causing both children and neural networks to classify them as a masculine.

In addition to these factors, the gender effect observed in Russian and Serbian can also be a consequence of the difference in length of the masculine and feminine form of adjectives in the nominative case, with masculine adjectives being usually shorter and thus easier to produce, e.g. *lep* 'beautiful (MASC)' vs. *lepa* 'beautiful (FEM)'. In addition to this, Serbian still differentiates adjective aspect (definite (longer) vs. indefinite (shorter)) for masculine adjectives. As a result, there are three groups of adjectives in Serbian: one which only appears in the definite form (e.g. *bratov* 'brother's'), a second group which only appears in indefinite form (e.g. *gornji* 'upper') and a third group which appears in both forms (e.g. *lep/lep-i* 'beautiful'). This feature of Serbian adjectives can be used for the further exploration of adjective length as a factor responsible for the gender effect observed in Serbian and Russian.

Taken together, a further exploration of potential phonological and distributional cues for gender learning in Serbian research will provide the opportunity to determine which of the proposed factors is responsible for the advantage for masculine nouns in gender acquisition. This research will fit well into an already existing body of cross-linguistic research in English, Dutch, French and Japanese which shows that other grammatical categories (e.g. parts-of-speech) can be extracted from the input with high accuracy (over 95%), by relying on the interaction between phonological and distributional cues (Monaghan, Christiansen & Charter, submitted).

Furthermore, it would be interesting to see whether a similar learning mechanism can be used for the explanation of the acquisition of case categories. Recent computational and mathematical modelling work on Serbian case marking

has shown that distributional, phonological and semantic cues are crucial for extracting and producing these categories (Mirković, MacDonald & Seidenberg, 2005; Moscoso del Prado Martin, Kostić, Filipović-Đurđević, submitted). Future studies should explore whether a similar advantage of pseudo-diminutives can also be found with respect to case marking. This, however, will require careful selection of tasks that give children sufficient opportunity to practise the production of cases for novel words so as to be able to differentiate the acquisition phenomena from pronunciation problems.

Finally, this thesis provided some speculative answers to the question as to why diminutives are used less frequently in Serbian CDS. Thus, diminutive production in Serbian CDS can be affected by the following three factors: a) the existence of other derivations for the expression of affection and endearment, like hypocoristics; b) poly-functional usage of diminutive suffixes; c) lack of register specific highly lexicalized diminutive forms associated with CDS, like *mishka* in Russian. There is, of course, also the possibility that socio-cultural differences may be at play, a factor that was beyond the scope of this thesis. To further explore the cross-linguistic differences in the production of diminutives in CDS, it will be necessary to undertake more systematic cross-linguistic comparisons based on comparable corpora across language, as well as experimental work on the production of CDS elicited under controlled laboratory conditions.

In sum, this thesis has strengthened the cross-linguistic evidence for the facilitating effect of diminutives on the acquisition of inflectional morphology and thereby contributed to a better understanding of the mechanisms underlying the role of CDS in language learning in general.

References

Abkarian, G.G., Dworkin, J. P., Abkarian, A.K. (2003). Fathers' speech to their children: perfect pitch or tin ear? *Fathering*, *1*, 27-50.

Akhutina, T., Kurganskij, A., Kurganskaya, M., Polinsky, M., Polonskaya, N., Larina, O., Bates, E. & Appelbaum, M. (2001). Processing of grammatical gender in normal and aphasic speakers of Russian. *Cortex, 37 (3),* 295-326.

Albin, D. D., & Echols, C. H. (1996). Characteristics of stressed and word-final syllables in infant-directed speech: Implications for word-level segmentation. *Infant Behavior and Development*, *19*, 401-418.

Albright, A. & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, *90*, 119-161.

Altmann, G.T. (2002). Statistical learning in infants. *PNAS, 99 (24)*, 15250–15251.

Anđelković, D., Ševa, N. & Moskovljević, J. (2001). *Serbian Corpus of Early Child Language.* Laboratorija za eksperimentalnu psihologiju, Filozofski fakultet i Katedra za opštu lingvistiku, Filološki fakultet, Beograd.

Andrews, E. (1999). Gender roles and preception: Russian diminutives in discourse. In M. Mills (ed.), *Slavic gender linguistics*. Philadelphia: John Benjamins.

Baron, N. (1990). *Pigeon-Birds and Rhyming Words: The Role of Parents in Language Learning.* Englewood Cliffs, New Jersey: Prentice.

Barton, M. W. & Tomasello, M. (1994). The rest of the family: The role of fathers and siblings in early language development. In Gallaway, C. & Richards, B. J. (Eds.), *Input and interaction in language acquisition* (pp. 109-134). New York, NY, US: Cambridge University.

Bauer, L. (1997). Evaluative morphology: In search of universals. *Studies in Language, 21*, 533–575.

Berent, I., Marcus, G., Shimron, J., & Gafos, A. (2002). The scope of linguistic generalizations: Evidence from Hebrew word formation. *Cognition*, *83 (2)*, 113-139.

Bernstein-Ratner, N. & Pye, C. (1984). Higher pitch in babytalk is not universal: Acoustic evidence from Quiche Mayan. *Journal of Child Language*, *11(3)*, 515-522.

Berstein-Ratner, N. (1988). Patterns of parental vocabulary selection in speech to very young children. *Journal of Child Language*, *15*, 481-492.

Berstein-Ratner, N. (1996). From "signal to syntax": But what is the nature of the signal? In Morgan, J.L. & Demuth, K. (Eds), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 135-150). Hillsdale, NJ, England: Lawrence Erlbaum Associates.

Bever, T.G. (Ed.) (1982). *Regressions in mental development: Basic phenomena and theories.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Blake, B. J. (2001). *Case*. Cambridge: Cambridge University Press.

Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn Taking Affects the Quality of Infant Vocalizations. *Journal of Child Language*, *14*, 211-27.

Bowerman, M. (1982). Reorganisational processes in lexical and syntactic development. In Wanner, E. & Gleitman, L. (Eds.), *Language acquisition: The state of the art.* Cambridge, UK: Cambridge University Press.

Brent, M. R. and Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*, 93-125.

Brent, M.R. (1996). Advances in the computational study of language acquisition. *Cognition*, *61*, 1-38.

Breznitz, Z. & Sherman, T. (1987). Speech patterning of natural discourse of well and depressed mothers and their young children. *Child Development*, *58(2)*, 395-400.

Brown, G., Anderson, A., Shillcock, R. & Yule, G. (1984). *Teaching talk: Strategies for production and assessment*. UK: Cambridge University Press.

Burnham, D., Kitamura, C. & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296*, 1435.

Bybee, J. L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*, 425-455.

Cartwright, T. A. & Brent, M. R. (1997). Syntactic categorization in early language acquisition: formalizing the role of distributional analysis. *Cognition*, *63*, 121-70.

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science, 8(6)*, 500-508.

Chater, N. & Vitányi, P. (2002). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*, 19-22.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, *103*, 566-581.

Cherry, L. & Lewis, M. (1976). Mothers and two-year-olds: A study of sex-differentiated aspects of verbal interaction. *Developmental Psychology*, *12(4)*, 278-282.

Christiansen, M. H. & Charter, N. (2001). Connectionist psycholinguistics: capturing the empirical data. *Trends in Cognitive Sciences*, *5(2)*, 82-88.

Christiansen, M., Allen, J. & Seidenberg, M. (1998). Learning to Segment Speech Using Multiple Cues: A Connectionist Model. *Language and Cognitive Processes*, *13*, 221-268.

Clearfield, M. W. & Nelson, N. M. (2006). Sex Differences in Mothers' Speech and Play Behavior with 6-, 9-, and 14-Month-Old Infants. *Sex Roles*, *54(1/2)*, 127-137.

Cooper, R. P., Abraham, J., Berman, S. & Staska, M. (1997). The development of infants'preference for motherese. *Infant Behavior and Development, 20*, 477- 488.

Cooper, R.P. & Aslin, R.N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61(5)*, 1584-95.

Corbett, G. (1991). *Gender*. Cambridge: Cambridge University Press.

Da Fonseca, P. N. & Salomao, N. M. R. (2005). Comparative analysis of maternal and paternal child directed speech / Analise comparativa das falas materna e paterna dirigidas as criancas. *Cadernos de Psicologia e Educacao Paideia*, *15(30)*, 79-91.

Dabrowska (submitted). The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology.

Dabrowska, E. (2004). Rules or schemas? Evidence from Polish. *Language and Cognitive Processes*, *19 (2)*, 225-271.

Dabrowska, E. (2006). Low-level schemas or productive rules: The role of diminutives in the acquisition of Polish case inflections. *Language Sciences*, *28,* 120-135.

Daltas, P. (1985). Some patterns of variability in the use of diminutive and augmentative suffixes in spoken Modern Greek Koine (MGK). *Glossologia*, *4*, 63-88.

Davidson, R. G. & Snow, C. E. (1996). Five-year-olds' interactions with fathers versus mothers. *First Language*, *16*, 223-242.

De Marco, A. (1998). The acquisition of diminutives in Italian. In Gillis, S. (Ed.), *Studies in the acquisition of number and diminutive marking* (pp. 175-192.). Antwerp papers in Linguistics, 95. Antwerp: Universiteit Anwerpen.

DePaulo, M. B. & Coleman, L. (1986). Talking to Children, Foreigners, and Retarded Adults. *Journal of Personality and Social Psychology*, *51(5)*, 945-959.

Dressler, W. (Ed.) (1997). *Studies in pre- and protomorphology*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Durkin, K., Rutter, D. R. & Tucker, H. (1982). Social interaction and language acquisition: Motherese help you. *First Language*, *3(8)*, 107-120.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, *7*, 195-224.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71-99.

Elman, J. L. (2005). Connectionist models of cognitive development: where next? *Trends in Cognitive Sciences*, *9(3)*, 111-117.

Elman, J. L., Bates, E.A., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Englund, K.T. (2005). Voice onset time in infant directed speech over the first six months. *First language*, *25(2)*, 219-234.

Ervin, S. (1964). Imitation and structural change in children's language. In E. Lenneberg (Ed.), *New directions in the study of language.* Cambridge, MA: MIT Press.

Farrar, J. (1990). Discourse and the acquisition of grammatical morphemes. *Journal of child language*, *17*, 607-624.

Farrar, J. (1992). Negative evidence and grammatical morpheme acquisition. *Developmental Psychology*, *28*, 90-98.

Farwell, D., Helmreich, S. & Casper, M. (1995). SPOST: a Spanish Part-of-Speech Tagger. http://crl.nmsu.edu/Publications/farwell/far_etal95.html.

Faszination Tier & Natur. München: Meister Verlag GmbH, IMP B.V.

Fee, E. J. & Shaw, K. (1998). Pitch modifications in Mi'kmaq child-directed speech. In Clark, E. V. (Ed.), *The proceedings of the twenty-ninth annual child language research forum* (pp. 47-54). Chicago, IL, US: Center for the Study of Language and Information.

Ferguson, C.A. & Slobin, D.I. (Eds.) (1973). *Studies of child language development*. New York: Holt, Reinhart & Winston.

Ferguson, C.A. (1978). Talking to children: A search for universals. In Greenberg, J.H. (Ed.), *Universals of human language, Volume I: Method & Theory* (pp. 203-224). Stanford University Press, Stanford, California.

Fernald, A. & McRoberts, G. (1996). Prosodic bootstrapping: A critical analysis of the argument and the evidence. In Morgan, J.L. & Demuth, K. (Eds.), *From Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 365-388). Mahwah, NJ: Erlbaum.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child Development, 60,* 1497-1510.

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In Barkow, J. H., Cosmides, L. & Tooby, J. (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391-428). New York: Oxford.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, *27*, 209-221.

Fernald, A., Taeschner, T., Dunne, J., Papousek, M., de Boysson-Bardies, B. & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*, 477-501.

Fernald. A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behaviour and Development*, *8*, 181-195.

Fisher, C. and Tokura, H. (1996). Prosody in Speech to Infants: Direct and Indirect Acoustic Cues to Syntactic Structure. In Morgan, J. and Demuth, C. (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 343-364). Mahwah, NJ: Lawrence Erlbaum Associates.

Fitneva, S.A., Tobiah, S.J., Christiansen, M.H. & Monaghan, P. (2005). Variation in Children's Use of Phonology to Infer Grammatical Categories. Poster

presented at the *Biennial Meeting of the Society for Research in Child Development*, April, 7th – 10th, Atlanta, GA.

Foulkes, P., Docherty, G.J. & Watt, D.J.L. (2005). Phonological variation in child directed speech. *Language*, *81 (1)*, 177-206.

Freudenthal, D., Pine, J. M. & Gobet, F. (2001). Modelling the optional infinitive stage in MOSAIC: A generalization to Dutch. In Altmann, E.M., Cleeremans, A., Schunn, C.D. & Gray W.D., (Eds.), *Proceedings of the 4th International Conference on Cognitive Modelling* (pp. 79-84). Mahwah, NJ: Erlbaum.

Gerken, L.A. (1996). Prosody's role in language acquisition and adult parsing. *Journal of Psycholinguistic Research*, *25(2)*, 345-356.

Gillis, S. (1997). The acquisition of diminutives in Dutch. In W. Dressler (Ed.), *Studies in pre- and protomorphology* (pp. 165-179). Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Gillis, S. (Ed.) (1998). *Studies in the Acquisition of number and diminutive marking. Antwerp Papers in Linguistics*, 95. Antwerp: Universiteit Anwerpen.

Gleason, B. J., Perlmann, R. U., Ely, R. & Evans, D. W. (1994). The babytalk register: Parents' use of diminutives. In Sokolov, J. L. & Snow, C. A. (Eds.), *Handbook of research in language development using CHILDES* (pp. 50-76). Hillsdale, NJ: Erlbaum.

Gleason, B.J. (1958). The Child's Learning of English Morphology. *Word*, *14*, 150-77.

Gleason, J. B. (1975). Fathers and other strangers: men's speech to young children. In Dato, D.P. (Ed.), *Developmental psycholinguistic theories and applications.* Georgetown University Press, Washington.

Gleitman, L. R., Newport, E.L. & Gleitman, H. (1984). The Current Status of the Motherese Hypothesis. *Journal of Child Language, 11*, 43-79.

Golinkoff, R. M. & Alioto, A. (1995). Infant-directed speech facilitates lexical learning in adults hearing Chinese: Implications for language acquisition. *Journal of Child Language, 22*, 703-726.

Hahn, U. & Nakisa, R.C. (2000). German inflection: Single route or dual route? *Cognitive Psychology*, *41 (4)*, 313–360.

Haman, E. (2003). Early productivity in derivation: A case study of diminutives in the acquisition of Polish. *Psychology of Language and Communication, 7*, 37-56.

Hampson, J. and Nelson, K. (1993). The relation of maternal language to variation in rate and style of language acquisition. *Journal of Child Language*, *20*, 313-342.

Hardy-Brown, K. & & Plomin, R. (1985). Infant Communicative Development: Evidence from Adoptive and Biological Families for Genetic and Evironmental Influeneces on Rate Differences. *Developmental Psychology, 21*, 378-85.

Hay, J. & Baayen, H. (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences*, *9(7)*, 342-348.

Hayashi, A., Tamekawa, Y. & Kiritani, S. (2001). Developmental change in auditory preferences for speech stimuli in Japanese infants. *Journal of speech, language, and hearing research*, *44(6)*, 1189-2000.

Hayes, D. P. & Ahrens, M. G. (1988). Vocabulary simplification for children: A special case of "motherese"? *Journal of Child Language, 15(2),* 395–410.

Herrera, E., Reissland, N. & Shepherd, J. (2004). Maternal touch and maternal child-directed speech: effects of depressed mood in the postnatal period. *Journal Affective Disorder*, *81(1)*, 29-39.

Herrera, M. & Carvallo de, P. Y. (1987). El diminutivo en ninos caraquenos de 7 10 an os de edad: un enfoque psicolingu stico. *Investigacion y Posgrado, 2*, 53–66.

Hoff-Ginsberg, E. & Tardif, T. (1995). Socioeconomic status and parenting. In Bornstein, M. H. (Ed.), *Handbook of parenting, Vol. 2: Biology and ecology of parenting* (pp. 161-188). Hillsdale, NJ, England: Lawrence Erlbaum Associates.

Hoff-Ginsberg, E. (1985). Some contributions of mothers' speech to their children's syntactic growth. *Journal of Child Language*, *12(2)*, 367-385.

Hoff-Ginsberg, E. (1991). Mother-child conversation in different social classes and communicative settings. *Child Development*, *62(4)*, 782-796.

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics*, *19(4)*, 603-629.

Hudson, R.A. (1996). *Sociolinguistics*. Cambridge University Press. Cambridge.

Hymes, D. (1974). *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia.

Ilić, N. & Kostić, A. (2002). Homography-problem in Automatic Lemmatization. Presented at *VIII Empirical Researches in Psychology*, Faculty of Philosophy. University of Belgrade. The Book of Abstracts, 36 (in Serbian).

Jin, J.A. & Naka, M. (2002). Elicitation of children's narratives about past events by Chinese mothers and fathers. *Japanese Journal of Developmental Psychology*, *13(3)*, 274-283.

Joanisse, M. F. & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *PNAS, 96,* 7592-7597.

Jocić, M. (1980). Jedan aspekt usvajanja padeženog sistema na ranom uzrastu: padeži za obeležavanje prostornih odnosa. *Godišnjak SDPLJ*, *4-4, 1980/81*, 87-91.

Johnston, J. R. & Wong, M.-Y. A. (2002). Cultural differences in beliefs and practises concering talk to children. *Journal of Speech, Language, and Hearing Research*, *45*, 916-926.

Johnston, J. R., & Slobin, D. I. (1979). The development of locative expressions in English, Italian, Serbo-Croatian, and Turkish. *Journal of child language, 6(3)*, 529-545.

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

Jusczyk, P., Hirsh-Pasek, K., Kemler-Nelson, D., Kennedy, L., Woodward, A. & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, *24*, 252-293.

Kaplan, P., Goldstein, M., Huckeby, E., & Cooper, R. P. (1995). Habituation, sensitization, and infants' responses to motherese speech. *Developmental Psychobiology*, *28*, 45-57.

Kaplan, P.S., Bachorowski, J. & Zarlengo-Strouse, P. (1999). Directed Speech Produced by Mothers with Symptoms of Depression Fails to Promote

Associative Learning in 4-Month-Old Infants. *Child Development*, *70*, 560 – 570.

Kaplan, P.S., Bachorowski, J.A., Smoski, M.J. & Zinser, M. (2001). Role of diagnosis and medication use in effects of maternal depression on infant-directed speech. *Infancy*, *2(4)*, 537-548.

Kaplan, P.S., Bachorowski, J.A., Smoski, M.J. & Hudenko, W.J. (2002). Infants of depressed mothers, although competent learners, fail to learn in response to their own mothers' infant-directed speech. *Psychological Science*, *13(3)*, 268-71.

Keibel, H. & Elman, J. (2004). Discovering word categories: Distributional regularities in German child-directed speech. Poster presented at the international conference *Architectures and Mechanisms for Language Processing* (AMLaP 2004). Aix-en-Provence, France.

Kelly, M.H. (1996). The role of phonology in grammatical category assignment. In Morgan, J.L. & Demuth, K. (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp.249-262). Mahwah, NJ: Lawrence Erlbaum Associates.

Kemler-Nelson, D.G., Hirsh-Pasek, K., Jusczyk, K. P. W. & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of child Language*, *16*, 55-68.

Kempe V. & Brooks, P. J. (2001). The role of diminutives in Russian gender learning: Can child-directed speech facilitate the acquisition of inflectional morphology? *Language Learning*, *51*, 221-256.

Kempe V., Brooks, P. J. & Pirott, L. (2001). How can child-directed speech facilitate the acquisition of morphology? In Almgren, M., Barrena, A.,

Ezeizabarrena, M.J., Idiazabal, I. & MacWhinney, B. (Eds.), *Research on Child Language Acquisition: Proceedings of the 8th conference of the International Association for the Study of Child Language* (pp. 1237-1247). Medford, MA: Cascadilla Press.

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.

Kempe, V., Brooks, P. J. & Gillis, S. (2005). Diminutives in child-directed speech supplement metric with distributional word segmentation cues. *Psychonomic Bulletin and Review*, *12(1),* 145-151.

Kempe, V., Brooks, P. J., Gillis, S. & Samson, G. (in press). Diminutives facilitate word segmentation in natural speech: Cross-linguistic evidence. *Memory & Cognition.*

Kempe, V., Brooks, P. J., Mironova, N. & Fedorova, O. (2003). Diminutivization supports gender acquisition in Russian children. *Journal of Child Language, 30*, 471-485.

Kempe, V., Brooks, P. J., Mironova, N., Pershukova, A. & Fedorova, O. (in press). Playing with word endings: Morphological variation in the learning of Russian noun inflections. *British Journal of Developmental Psychology.*

Kempe, V., Ševa, N., Brooks, P.J., Mironova, N., Pershukova, A. & Fedorova, O. (in preparation). Diminutives and case marking: Cross-linguistic comparisons between Russian and Serbian.

Klajn, I. (2003). *Tvorba reči u srpskom jeziku: Sufiksacija i konverzija*. Beograd: Zavod za udžbenike i nastavna sredstva, Matica srpska i Institut za srpski jezik.

Korecky-Kroell, K. & Dressler, W. (2004). Austrian German diminutives. Paper presented at the *11th International Morphology Meeting*. February, 13th-17th, Vienna, Austria.

Kostić, Đ. & Vidosavljević, S. (1995). *Govor i jezik deteta u razvoju*. Zavod za udžebnike i nastavna sredstva, Beograd.

Kostić, Đ. (1965). *Sintaktičke funkcije padežnih oblika u srpskohrvatskom jeziku* [*Syntactic Functions and Meanings of Serbo-Croatian cases*]. Beograd: Institut za eksperimentalnu fonetiku i patologiju govora.

Kostić, Đ. (1999). *Frequency Dictionary of Contemporary Serbian Language, Vol. I-VII*. Beograd: Institut za Ekesperimentalnu fonetiku i patologiju govora i Laboratorija za eksperimentalnu psihologiju Filozofskogo fakulteta.

Kostić, Đ. (2001). *Kvantitativni opis strukture srpskog jezika – Korpus srpskog jezika.* [*Quantitative Description of Serbian Language Structure – Corpus of Serbian Language*]. Beograd: Institut za eksperimentalnu fonetiku i patologiju govora, Beograd i Laboratorija za eksperimentalnu psihologiju Filozofskog fakulteta u Beogradu.

Kuhl, P.K., Andruski, J.E., Chistovich, I. A., Chistovich, L.A., Kozhevnikova, E.V., Ryskina,V. L., Stolyarova, E. I., Sundberg, U. & Lacerda, F. (1997). Cross-Language Analysis of Phonetic Units in Language Addressed to Infants. *Science, 277 (5326)*, 684 – 686.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* Chicago: University of Chicago Press.

Lanvers, U. (2004). Gender in discourse behaviour in parent–child dyads: a literature review. *Child: Care, Health & Development, 30 (5)*, 481–493.

Leaper, C., Anderson, J. K. & Sanders, P. (1998). Moderators of Gender Effects on Parents' Talk to Their Children: A Meta-Analysis. *Developmental Psychology*, *34(1)*, 3-27.

Lieven, E. V., Pine, J. M. & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language, 24*, 187-219.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. Volume 1: Transcription format and programs. Volume 2: The Database*. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the Verb Learning Model. *Cognition, 29*, 121-157.

Marcus, G., Brinkmann, U., Clahsen, H., Wiese, R., Woest, A. & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology, 29,* 189–256.

Matthews, J., Ichile, I., Newman, R. & Berstein-Ratner, N. (2004). Mothers' and Fathers' Child-addressed Speech Varies by Race and SES. Presented at *ASHA Convention*, November 18th-20th, Philadelphia,USA.

McLaughlin, B., White, D., McDevitt, T. & Raskin, R. (1983). Mothers' and fathers' speech to their young children: Similar or different? *Journal of Child Language*, *10*, 245-252.

Melzi, G. & King, K.A. (2003). Spanish diminutives in mother-child conversations. *Journal of Child Language, 30*, 281-304.

Messer, D. (1980). The episodic structure of maternal speech to young children. *Journal of Child Language*, *7(1)*, 29-40.

Mimica, I. (1987). The influence of word order and case inflection on sentence comprehension at preschool children. *Radovi-Filozofski Fakultet Zadar-Razdio Filozofije Psihologije, Sociologije i Pedagogije*, *27(4)*, 159-168.

Mintz, T.H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*, 91–117.

Mintz, T.H., Newport, E.L. & Bever, T.G. (2002). The distributional structure of grammatical categories in speech to young children, *Cognitive Science*, *26*, 393–424.

Mirković, J., MacDonald, M. C. & Seidenberg, M. S. (2005). Where does gender come from? Evidence from a complex inflectional system. *Language and Cognitive Processes, 20*, 139-168.

Mirković, J., Seidenberg, M. S. & Joanisse, M. F. (2002). Morphology in an inflectionally rich language: Implications for the rules vs. connections debate. *Poster presented at the 43rd Annual Meeting of the Psychonomic Society, Kansas City, KS.*

Monaghan, P. Chater, N. & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorization. *Cognition*, *96(2)*, 143-182.

Monaghan, P., Christiansen, M.H. & Chater, N. (submitted). The Phonological-Distributional Coherence Hypothesis: Cross-linguistic evidence in language acquisition.

Morgan, J. L. & Demuth, K. (Eds.) (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition,* Mahwah, NJ: Erlbaum.

Morgan, J.L. (1996). Prosody and the Roots of Parsing. *Language and Cognitive Processes*, *11(1-2)*, 69-106.

Moscoso del Prado Martin, F., Kostić, A. & Filipović-Đurđević, D. (submitted). The Missing Link between Morphemic Assemblies and Behavioral Responses:a Bayesian Information-Theoretical model of lexical processing. *http://cogprints.org/perl/user_eprints?userid=6204.*

Munakata, Y. & McClelland, J.L. (2003). Connectionist models of development. *Developmental Science*, *6(4)*, 413-429.

Nicely, P., Tamis-LeMonda, C.S. & Bornstein, M.H. (1999). Mothers' attuned responses to infant affect expressivity promote earlier achievement of language milestones. *Infant Behavior and Development*, *22*, 557-568.

Ogle, S. & Maidment, J. A. (1993). Laryngographic analysis of child-directed speech. *European Journal of Disorders of Communication*, *28*, 289-297.

Pegg, J., Werker, J.F. & McLeod, P.J. (1992). Preference for infant-directed over adult- directed speech: Evidence from 7-week-old infants. *Infant Behavior and development*, *15*, 325-345.

Pine, J. M., Lieven, E. V. M. & Rowland, C. F. (1997). Stylistic variation at the 'single-word' stage: relations between maternal speech characteristics and children's vocabulary composition and usage. *Child Development*, *64*, 807-819.

Pinker, S. (1991). Rules of language. *Science, 253*, 530-535.

Plunkett, K. & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition*, *38*, 43-102.

Plunkett, K. and Elman, J.F. (1997). *Exercises in Rethinking Innateness. A Handbook for Connectionist Simulations*. MIT Press, Cambridge, MA.

Plunkett, K. & Marchman, V. (1991). U-shaped learning and frequency effects in a multilayered perceptron: Implications for child language acquisition. *Cognition, 38,* 43–102.

Plunkett, K. & Marchman, V. (1993). From rote learning to system building. *Cognition, 48,* 21–69.

Plunkett, K. & Nakisa, R. (1997). A connectionist model of the Arabic plural system. *Language and Cognitive Processes, 12,* 807–836.

Prasada, S. & Pinker, S. (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes, 8,* 1–56.

Rabain, J. J. & Sabeau, J. E. (1997). Maternal speech to 4-month-old infants in two cultures: Wolof and French. *International Journal of Behavioral Development*, *20(3),* 425-451.

Ramscar, M. (2002). The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology*, *45(1),* 45-94.

Redington, M. & Chater, N. (1998). Connectionist and Statistical Approaches to Language Acquisition: A Distributional Perspective. *Language and cognitive processes*, *13 (2/3),* 129–191.

Redington, M., Chater, N. and Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories, *Cognitive Science*, *22*, 425–469.

Reissland, N., Shepherd, J. & Herrera, R. (2003). The pitch of maternal voice: a comparison of mothers suffering from depressed mood and non-depressed

mothers reading books to their infants. *The Journal of Child Psychology and Psychiatry, 44(2)*, 255-261.

Rowe, M. L, Coker, D. & Pan, B. A. (2004). A Comparison of Fathers' and Mothers' Talk to Toddlers in Low-income Families. *Social Development*, *13(2)*, 278-291.

Rumelhart, D. E. & McClelland, J. L. (1986). On Learning Past Tenses of English Verbs. In Rumelhart, D. E., & McClelland, J. L. and the PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volumes 1 and 2*. (pp. 216-271). Cambridge, MA: MIT Press.

Savić, S. & Polovina, V. (1989). *Razgovorni srpskohrvatski jezik* (*Conversational Serbo-Croatian*). Institute for South Slavic Languages. Faculty of Philosophy. University of Novi Sad.

Savickienė, I. (1998). The acquisition of diminutives in Lithuanian. In Gillis, S. (Ed.), *Studies in the Acquisition of number and diminutive marking* (pp. 115-135). Antwerp Papers in Linguistics, 95. Antwerp: Universiteit Anwerpen.

Savickienė, I., Kempe, V. & Brooks, P. J. (in preparation). The diminutive advantage in gender agreement of Lithuanian children.

Seidenberg, M. S. (1992). Connectionism without tears. In Davis. S. (Ed.), *Connectionism: Theory and practice* (pp. 84-122). New York, NY: Oxford University Press.

Ševa, N. & Kempe, V. (2005). Serbian diminutives aid case-marking acquisition in Serbian. Poster presented at the *Post-Cognitivist Psychology Conference.* July, 4th–6th, University of Strathclyde, Glasgow, UK.

Ševa, N, Kempe, V. & Brooks, P. J. (2005). Serbian diminutives facilitate gender acquisition despite low frequency in child-directed speech. Poster presented at the *Biennial conference of the Society for Research in Child Development*, April, 7th – 10th, Atlanta, Georgia USA.

Ševa, N., Hadjiconstantinou, G. & Kempe, V. (2005). Estimating distributional characteristics of child-directed speech of different languages under controlled laboratory conditions. Poster presented at the *British Psychological Society Developmental Section Conference, University of Edinburgh,* September, 5th –8th, Edinburgh, UK.

Ševa, N., Kempe, V. & Brooks, P. (2006-a). Inducing low-level schema extraction with artificial suffixes. Published in the Proceedings of the *28th Annual Conference of the Cognitive Science Society,* July 26th -29th, Vancouver, Canada.

Ševa, N., Kempe, V. & Brooks, P. (2006-b). How does the diminutive advantage in learning grammatical gender merge? Comparing the performance of Serbian children and neural networks. Poster presented at the *12th Annual Conference on Architectures and Mechanisms for Language Processing,* 30thAugust –1st September, 2006, Nijmegen, Netherlands.

Ševa, N., Kempe, V., Brooks, P.J, Mironova, N., Pershukova, A. & Fedorova, O. (in press). Cross-linguistic evidence for the diminutive advantage: Gender agreement in Russian and Serbian children. *Journal of Child Language.*

Shute, B. & Wheldall, K. (1989). Pitch alterations in British motherese: Some preliminary acoustic data. *Journal of Child Language*, *16(3)*, 503-512.

Slobin, D. I. & Bever, T. G. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition, 12*, 229-265.

Slobin, D.I. (1973). Cognitive prerequisites for the development of grammar. In Ferguson, C.A. & Slobin, D.I. (Eds.), *Studies of child language development*. (pp. 175-208). New York: Holt, Reinhart & Winston.

Slobin, D.I. (Ed.) (1985). *The Crosslinguistic Study of Language Acquisition*. Volume 1&2. Lawrence Erlbaum, Hillsdale, N.J.

Smith, L.B. & Samuelson, L.K. (2003). Different is good: connectionism and dynamic systems theory are complementary emergentist approaches to development. *Developmental Science*, *6(4)*, 434-439.

Snow, C. E., Smith, N. S. H. & Hoefnagel-Hohle, M. (1980). The acquisition of some Dutch morphological rules. *Journal of Child Language*, *7*, 539–53.

Snow, C.E. (1985). Mothers' speech to children learning language. Citation classic, Current Contents, *Social and Behavioral Sciences*, *17(1)*, 18.

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

Stanojčić, Ž. & Popović, Lj. (2003). *Gramatika srpskog jezika*. Zavod za izdavanje udžbenika i nastavna sredstva. Beograd.

Stephany, U. (1997). Diminutives in early child Greek. In Dressler, W. U. (Ed.), *Studies in pre- and protomorphology* (pp. 147–156). Vienna: Osterreichische Akademie der Wissenschaften.

Stevanović, M. (1964). *Savremeni srpskohravtski jezik I-II*. Beograd: Naučna knjiga.

Stoneman, Z. & Brody, G. H. (1981). Two's company, three makes a difference: an examination of mother's and father's speech to their young children. *Child Development*, *52*, 705–707.

Swanson, L.A. & Leonard, L.B. (1994). Duration of function-word vowels in mothers' speech to young children. *Journal of Speech and Hearing Research*, *37*, 1394-1405.

Swanson, L.A., Leonard, L.B. & Gandour, J. (1992). Vowel duration in mothers' speech to young children. *Journal of Speech and Hearing Research*, *35*, 617-625.

Tamis-LeMonda, C. S. and Bornstein, M. (2002). Maternal responsiveness and early language acquisition. In Kail, R. & Reese, H. W. (Eds.), *Advances in Child Development and Behavior, Vol. 29* (pp. 89-127). San Diego, CA: Academic Press.

Tamis-LeMonda, C.S., Bornstien, M. H. & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, *72*, 748-767.

Tamis-LeMonda, C.S., Shannon, J.D., Cabrera, N.J. & Lamb, M.E. (2004). Fathers and mothers at play with their 2- and 3-year-olds: contributions to language and cognitive development. *Child Development*, *75 (6)*, 1806 – 1820.

Tenenbaum, H. R. & Leaper, C. (1998). Gender effects on Mexican-descent parents' questions and scaffolding during toy play: A sequential analysis. *First Language*, *18*, 129-147.

Thiessen, E.D., Hill, E.A. & Saffran, J.R. (2005). Infant directed speech facilitates word segmentation. *Infancy*, 7, 49-67.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge: Cambridge University Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of child language acquisition.* Harvard University Press, Cambridge, MA.

Tomasello, M., Conti-Ramsden, G. & Ewert, B. (1990). Young children's conversations with their mothers and fathers: Differences in breakdown and repair. *Journal of Child Language*, *17*, 115-130.

Trainor, L.J, Austin, C.M., Desjeradins, R.N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychogical science, 11(3),* 188-195.

Voeykova, M. (1997). Early extragrammatical operations in the speech of a Russian child (from 1;04.11 to 2;00.03). In Dressler. W. (Ed.), *Studies in pre- and protomorphology* (pp. 37-43). Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Walker, K. & Armstrong, L. (1995). Do mothers and fathers interact differently with their child or is it the situation which matters? *Child: care, health and development*, *21(3)*, 161-81.

Warren, L. A. & Bohannon, J. N. (1984). Intonation patterns in child-directed speech: Mother-father differences. *Child Development*, *55(4)*, 1379-1385.

Werker, J.F. & McLeod, P.J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology*, *43*, 230-246.

Werker, J.F., Pegg, J. & McLeod, P.J. (1994). A cross-language investigation of infant preferences for infant-directed communication. *Infant behaviour and development*, *17(3)*, 323-333.

Zipf, G.K. (1949). *Human behavior and the principle of least effort.* Addison-Wesley, Reading, MA.

Appendix 1

Table A1.1. Taxonomy of functions and meanings encompassed by Serbian cases (adopted from Kostić, Đ. 1965). [a]

1- without preposition, 2 - with preposition, 3 - with or without preposition

NOMINATIVE

(always without preposition)

1. Subject
2. Predicate
3. In exclamations

GENITIVE

1. Place (1)
2. Partitive genitive (3)
3. Possessive genitive (3)
4. Temporal genitive (3)
5. Ablative genitive (2)
6. Genitive denoting material (3)
7. Slavic genitive (1)
8. Logical subject (1)
9. Denoting reaching (2)
10. Exclusion (2)
11. Cause (2)
12. Qualitative genitive (3)
13. Distal object with verbs denoting approaching (1)
14. Source (2)
15. Comparison (2)
16. Object with verbs denoting moving away, drawing back etc. (1)
17. In oaths (1)
18. Denoting measure, quantity (3)
19. Substitution (2)
20. Genitive of origin (3)
21. Modal genitive (2)
22. Goal (2)
23. Genitive instead of accusative in rhyme completion (1)
24. With demonstrative adverbs (1)
25. In curses (1)
26. Denoting usage (2)
27. Adding (with the preposition *osim* (except)) (2)
28. Noun complement (2)
29. Opposition (2)
30. Denoting surprise (1)
31. Preposition *od* (of) with genitive instead of *o* (about) with locative (2)
32. In some expressions with the preposition *do* (up to)
33. Distinction (2)
34. Verb complement (2)
35. Purpose of action (2)
36. In exclamatory sentences (1)
37. Explicative genitive (1)
38. Agent in passive sentences (2)

39. Intransitive verb complement (1)
40. Indicating origin or source (2)
41. Destination (with the preposition *kod* (at) with genitive, instead of preposition *k* (to) with dative (2)
42. Rank of order, succession (2)
43. Relation (2)
44. Adjectival complement (2)
45. Instrument (1)
46. Exclusion, separation (2)
47. Permission (2)
48. Relation between objects (2)
49. Conditional with the preposition *kod* (with) (2)
50. With expressions of regret (1)
51. Purpose (with the preposition *za* (for))

DATIVE

1. Direction (1)
2. Distal object (1)
3. Object with intransitive and reflexive verbs (1)
4. Purpose (1)
5. Dative commodi and incommodi (1)
6. Logical subject (1)
7. With verbs denoting possibility and obligation (1)
8. With verbs whose object is in dative or accusative (1)
9. Noun complement (1)
10. Possessive dative (1)
11. Ethic dative (1)
12. In oaths (1)
13. Adjectival complement (1)
14. Adverbial complement (1)
15. In exclamations (1)
16. With preposition *k* (*to*) instead of dative alone (2)
17. With preposition *k* meaning *towards* (2)
18. Predicate with the infinitive (1)
19. Time (2)
20. Opposition (2)
21. With the preposition *blizu* (*close to*) denoting place
22. With the preposition *bliže* (*closer to*) denoting comparison

ACCUSATIVE

1. Object with transitive verbs (1)
2. Place (2)
3. Time (2)
4. Verb complement (2)
5. Modality (2)
6. Purpose (2)
7. Preposition *na* (*on*), with verbs denoting movement, goes with the accusative for nouns denoting action or state (2)
8. Contact, touching (2)
9. Top down movement (2)
10. Noun complement (2)

11. Cause (2)

12. Object with some intransitive verbs[11] (1)

13. Adjectival complement (2)

14. Bottom up movement (2)

15. Sign, specification (2)

16. Direction (2)

17. Preposition *na* (*on*) with the accusative instead of *u* (*in*) after the question *where to?* (2)

18. Covering in a sense of protection (2)

19. Intention (2)

20. Division (2)

21. Measure (3)

22. Instrument (2)

23. Goal, purpose (2)

24. Change (2)

25. Price (3)

26. With verbs and expressions denoting mood (in psychological sense) (1)

27. Substitution (2)

28. In oaths (2)

29. Proximal object used brachilogically (1)

30. In some expression accusative with the preposition *od* (*from*), *do* (*up to*), oko (*around*), *s* (*with*), *preko* (*over*) (2)

31. Usefulness or damage (2)

32. Putting together (2)

33. With the expressions denoting some offer (1)

34. Verb complement with the preposition *za* (*for*) (2)

35. (a) accusative instead of instrumental in expressions like *face to face*, (b) with the preposition *iz* (*from*) in expressions like *from day to day*.

36. Logical subject

37. Two accusatives (real object and predicative accusative) (1)

38. With adjectives *dužan* (*obliged to*) and *voljan* (*willing to*) (1)

39. Two accusatives with transitive verbs to learn*, to ask* etc. (1)

40. Accusative as proximal object instead of dative (1)

41. Accusative instead of nominative with passive verbs (1)

42. Accusative in subjectless sentences (1)

43. Accusative with assumed verb, but not expressed (1)

44. Actor is in the accusative with verbs like *to start to*, followed by nouns like *laughter*, *shouting* etc. (1)

45. taking (with the preposition *po* (for)) (2)

46. Side effects (2)

47. Denoting superiority (with the preposition *nad* [*over*]) (2)

48. Comparison (2)

49. Inequality (2)

50. Adding (2)

51. Company (2)

52. Simultaneity (2)

53. Quantity (2)

54. Equality (2)

55. Prepositional object (2)

56. Condition (2)

57. Allowance (2)

58. Accusative with the preposition *za* (*for*) as the subject of the main clause

INSTRUMENTAL

1. Instrument or tools (1)

2. Accompaniment (2)

3. Place (3)

4. With verbs *to make someone... e.g. happy*, *to call someone... e.g. friend*, *to appoint someone...* etc. (1)

5. Adjectival complement (3)

6. Manner (3)

7. Attribute (2)

8. In oaths (1)

9. Unspecified[12]

10. Rank of order, ordering (2)

11. Object (3)

12. Verb complement (3)

13. Place used figuratively, denoting subordination (with the preposition *pod* [under])

14. Separation (2)

15. Sign, specification (3)

16. Distal object (1)

17. Time (3)

18. Noun complement (3)

19. Nominal part of predicate (1)

20. In expressions like *as if...* (1)

21. Side effects (2)

22. Noun in instrumental denoting superiority (1)

23. State (1)

24. Goal (1)

25. Cause (1)

26. With verbs denoting feelings (2)

27. With passive verb forms (1)

28. With gerunds as complement of verbs that go with instrumental (1)

29. Comparison, figuratively (2)

30. Place used figuratively, with the preposition *među* (*between, among*)

31. Covering, protecting (2)

32. With the adjective *satisfied with* (1)

LOCATIVE (always with preposition)

1. Place

3. Time

4. Manner

5. Verb complement

6. Bivalent

7. Noun complement

8. Denoting that something is appropriate

9. Specification

10. Topic of conversation

11. Denoting contact
12. Side effects
13. Adjectival completion
14. Denoting dependence (figuratively)
15. Opposition
16. With count nouns and collective nouns
17. Origin, source
18. *By means of...*
19. Proportion
20. Comparison
21. Unspecified (Instances of locative that could not be classified in the above categories).

a. Taxonomy of functions and meanings of Serbian cases was part of the project of the Institute for Experimental Phonetics and Speech Pathology in Belgrade, aimed to specify the probability of occurrence of Serbian grammatical forms and phonological structure of Serbian language. The project was conceived, guided and supervised by Prof. Đorđe Kostić from 1957 to 1965.

Appendix 2

Lemmatization of Serbian corpus

The semi-automatic lemmatization of the parental utterances from the *Serbian Corpus of Early Child Language* (Anđelković et al., 2001), was carried out with the adapted MOR program from the CHILDES package (MacWhinney, 2000), developed for morpho-syntactic rule based tagging of English.

1. General structure of MOR program

The MOR program consists the four parts:

a) a language free parser;

b) a lexicon which is adaptable to different languages and which can contain both a list of lemmata and a list of inflectional and derivational forms of words.

Example (1) represents an output of tagged lemmata for English lexicon:

can {[scat v:aux]}

a {[scat det]}

an {[scat det]} "a"

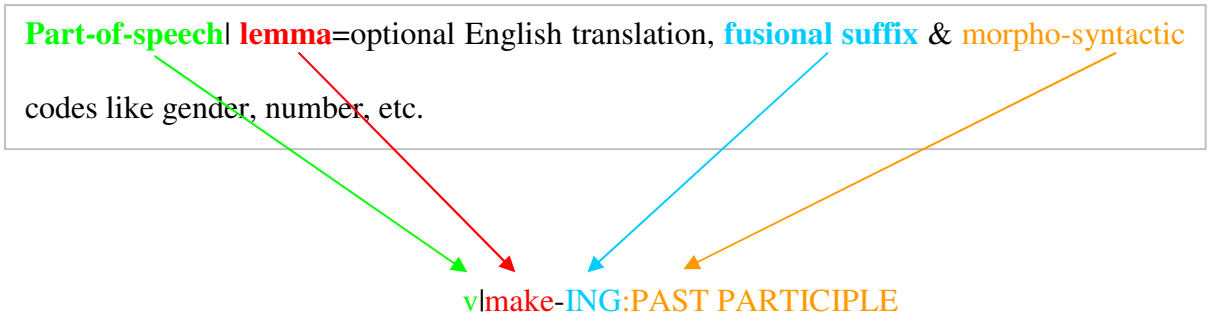
go {[scat v] [ir +]}

went {[scat v] [tense past]} "go&PAST";

c) a set of rules which are used for deriving regular forms of words if the overloading of the lexicon has to be avoided (e.g., –ing and –ed forms of verbs) and

d) an additional ruled-based program POST for resolving morpho-syntactic ambiguities.

The first action of the parser is the loading of the run-time lexicon made consisting a list of lemmata and derived morphological forms for regular words, based on a set of existing rules. After that, the parser matches the items from the text with the words from the virtual run-time lexicon. The output of lemmatization is printed in the %mor line in the following form:



The MOR program frequently gives more than one solution for a particular item due to the high level of homography. For example, in English, where conversion is one of the most productive derivational processes, homography usually occurs in different parts of speech. Example (3) illustrates an utterance in which two words have more than one interpretation[20].

(3) *CHI: I want to go back.

%mor: pro|I v|want **inf|to^prep|to** v|go **adv|back^n|back^v|back** .

Ambiguities presented in Example (3) can be resolved in principle in two ways: a) **manually** which is very time consuming and labour intensive, but yields the highest degree of accuracy and b) **(semi)automatically**, by using probability information or complex linguistic rules which saves time and labour, but increases the number of errors.

The MOR package originally allowed only for ruled base automatic disambiguation, with pre-specified set of rules for English. For any other language, researchers would have to develop language specific rules for disambiguation. Below I will describe how I adopted MOR for lemmatizaton of Serbian CDS.

[20] Ambiguous codes are separated by '^' sign.

2. Lemmatization of Serbian language

As described in Chapter 2, Serbian has a complex inflectional morphology. Until now, a rule based morpho-syntactic parser for Serbian has not been developed, and given the complexity of the system, it is a question whether any ruled based account would achieve an acceptably low error rate (less than 5%). On the other hand, previous lemmatization of Serbian written language (Ilić & Kostić, 2002) showed that if we use simple frequency criteria, where the most frequent lemmata are the correct ones, the error rate did not exceed 5%. In order to use the same probability based procedure for the lemmatization of the spoken language (in this case CDS), information on the frequency of lemmata would have to be introduced indirectly into the MOR parser. This can be achieved by utilizing the stack-memory within the parser which operates on the simple principle: Last In First Out (LIFO). This means that a simple sorting of lemmata in descending order would allow the parser to put the last word on the list (the most frequent one) as the first one in the %mor line. In this way the program can use information on the frequency of lemmata without actually comparing the frequency rate for each lemma.

Materials:

Estimates on the accuracy of the adapted MOR program for Serbian were based on 4 samples of the utterances of two of the most productive parents when the children were 1;8 and 3;8 years old. Overall, the sample contained 4000 words.

Procedure:

The lexicon of the *Frequency Dictionary of Serbian Contemporary Language* (Kostić, 1999) was adapted for the of the MOR program. Example (4) presents outlook for the lemma *kuća* 'house' with all inflectional forms:

(4)

kuća {[scat n]} "kuća"

kuća {[scat n]} "kuća"

kuće {[scat n]} "kuća"

kućo {[scat n]} "kuća"

kući {[scat n]} "kuća"

kuće {[scat n]} "kuća"

Lemmata from the dictionary and their inflectional forms were sorted in the descending order by the frequency of lemmata. Table A2.1. presents the system of part-of-speech tags used for lemmatization of Serbian CDS.

Table A2.1. List of codes used in lemmatization of Serbian CDS

Code	*Part of speech*	*Code*	*Part of speech*
adj	adjective	**n**	noun
adv	adverb	**neo**	neologism
bab	babbling	**num**	number
chi	child word	**onoma**	onomatopoeic word
co	communicator	**prep**	preposition
conj	conjunction	**pro**	pronoun
fam	Family word	**ptl**	particle
int	filled pause	**unk**	unknown
l	Letter	**v**	verb

The utterances were lemmatised with the command:

mor +lsrb +gsrb *.cut

The +lsrb and +gsrb commands are used for activating the Serbian lexicon and the Serbian set of rules which was empty in this case, because program indirectly used frequency information.

Results:

Similar to English, the Serbian version of the MOR program often produced the coding in which the lemmata had more than one interpretation.

Example (5) presents one of the mother's utterances, where three out of five words in the sentence had more than one interpretation:

(5) *MAJ: ona se može sama igrati .

%mor: **pro|on^pro|onaj prt|se^pro|sebe** v|moći **pro|sam^adj|sam^n|sama** v|igrati .

%eng: she can play on her own.

The Figure A2.1. illustrates the distribution of unindentified words, words with only one interpretation, and words with more than one lemma in two registers: CDS and written language. Overall, the correlations of the homography distribution between the four samples ranged between 0.897 and 0.995, indicating that the number of lemmata with more than one interpretation does not vary a lot across different speakers. The averaged homography distribution for CDS also correlated very highly with the homography distribution for the written language, with $r=0.934$, $p<0.01$. The overall percentage of words with more than one interpretation was higher in CDS with approximately 55% of words in comparison to written language with approximately 47% of ambiguous interpretations. This is mainly due to the register differences manifested in the increase of functional words in spoken language which are the main carriers of homography in Serbian.

Small differences like an increase from 5% to 8% of unindentified words in the CDS register were mainly due to an increase of verbs in second person which are typical for the spoken register, but not frequent in the *Frequency Dictionary of Contemporary Serbian Language* (Kostić, 1999) which was based on the written language.

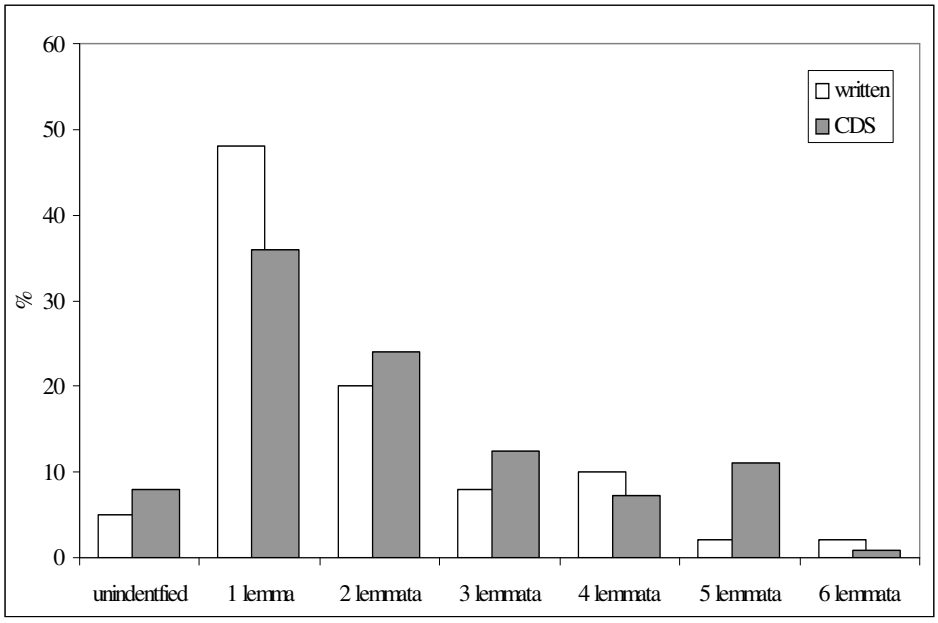


Figure A2.1. Homography distribution in Serbian, across CDS and written language.

Lemmatization of 4000 words from the CDS sample was checked manually in order to determine the level of accuracy, if we assume that the most frequent lemmata is the correct one.

Figure A2.2. illustrates the percentage of correct lemma coding in the first, second or third frequency rank for CDS and written language.

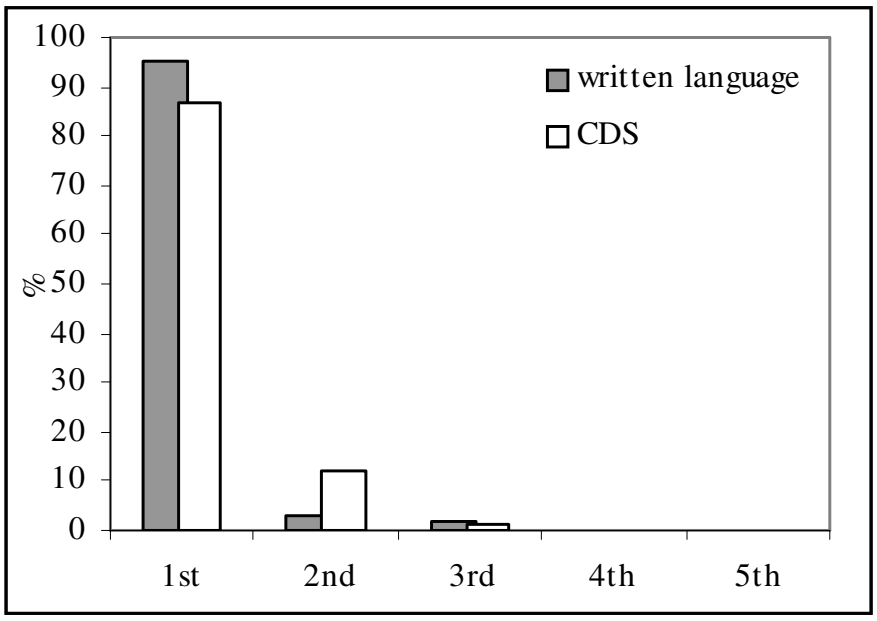


Figure A2.2. The percentage of correct lemma coding on the first, second or third place of frequency in CDS and written language.

The results of the lemmatization of CDS, showing that 87% of the correctly coded words were the most frequent ones, indicate that the error rate for spoken language register is higher in comparison to written language with 95% of correct answers. A closer analysis of the correctly coded words on the second frequency rank revealed that the increase of errors in CDS was mainly due to homography between some of the inflectional forms of two pronouns: *taj* 'that' and *ti* 'you'. In written language (based on the *Frequency Dictionary*), the demonstrative pronoun is more frequent than the second person pronoun. Since the frequency of the second person pronoun increases in the spoken language, due to the dyadic nature of conversational discourse, the difference in frequency between its homographic counterparts is probably more levelled in comparison to the written language.

A similar problem was observed for the *conjunction vs. particle* interpretation, where there was an increase in the usage of the particle *i* 'and' in comparison to the conjunction *i* 'and'. Figure A2.3. presents the distribution of the most problematic pair of first and second ranked interpretations by frequency, where the second rank was the correct one.

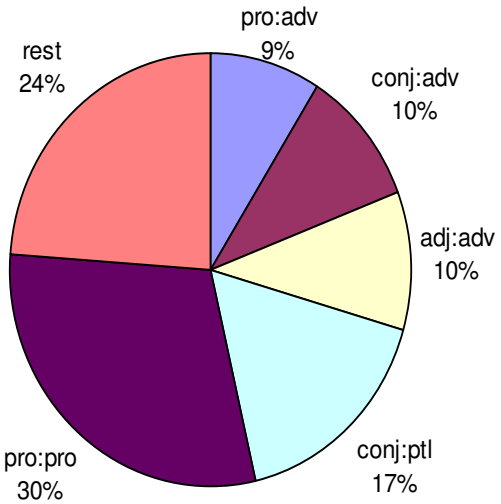


Figure A2.3. The distribution of correctly coded words on the second rank frequency across different parts of speech.

An analysis of the confusion matrix analysis of errors showed that within each type of words, adverbs were the most error prone, with only 60% of adverbs labelled accurately. Figure A2.4 depicts the percentage of errors within each part of speech.

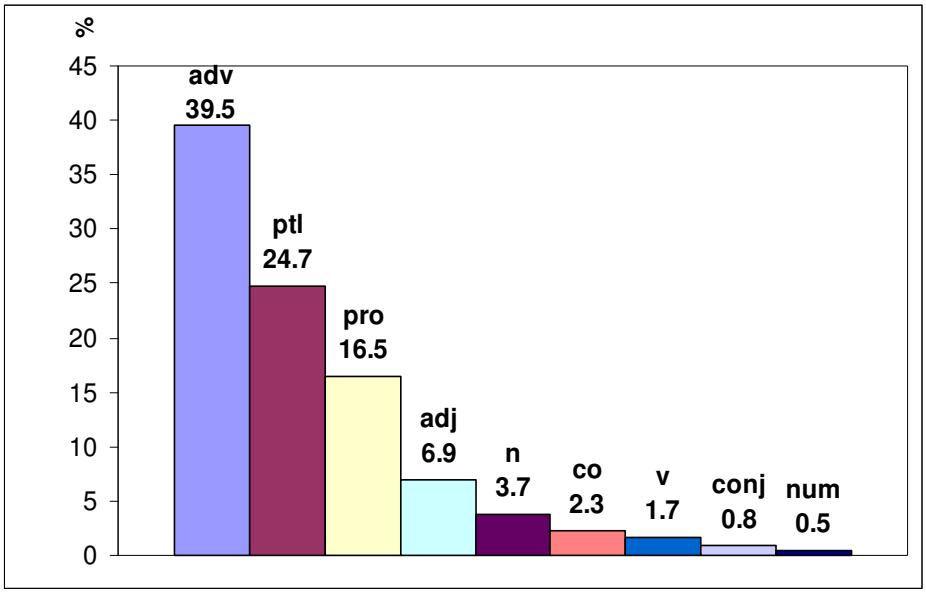


Figure A2.4. Percentage of errors across different parts of speech.

Both the analyses of the correctly coded words on the second frequency rank and the confusion matrix for different parts of speech showed that errors are not distributed homogeneously. This outcome of lemmatization of Serbian CDS is probably due to slight register differences expressed in an incompatibility between distributions in the *Frequency Dictionary* and the spoken utterances. This indicates that in order to improve automatic lemmatization of Serbian CDS, the MOR parser should be fed with additional information on conditional probabilities only for adverbs, particles and pronouns, or to manually check each instance of these parts of speech.

Appendix 3

Table A3.1. Percentage of diminutive tokens out of all common nouns in a sample of Serbian CDS, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivisible nouns	Simplex	Grand Total
Feminine	Feminine	5.1	1.4	11.2	3.6	31.3	52.5
	Opaque	0.0	0.0	0.0	0.9	0.2	1.2
	Unknown	0.00	0.00	0.00	0.02	0.00	0.02
Feminine Total		5.1	1.4	11.2	4.5	31.5	53.7
Masculine	Feminine	0.0	0.0	5.4	0.1	0.7	6.2
	Masculine	1.7	0.4	0.6	4.5	22.6	29.8
Masculine Total		1.7	0.4	6.0	4.6	23.4	36.0
Neuter	Neuter	0.2	0.0	0.0	2.3	7.4	9.8
Neuter Total		0.2	0.0	0.0	2.3	7.4	9.8
Unknown	Feminine	0.0	0.0	0.0	0.0	0.0	0.0
	Unknown	0.0	0.0	0.0	0.4	0.0	0.4
Unknown Total		0.0	0.0	0.0	0.4	0.0	0.4
Grand Total		**7.0**	**1.8**	**17.2**	**11.7**	**62.3**	**100.0**

Table A3.2. Percentage of diminutive lemmata out of all common nouns in a sample of Serbian CDS, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivisible nouns	Simplex	Grand Total
Feminine	Feminine	6.8	1.3	1.0	6.6	27.8	43.4
	Opaque	0.0	0.0	0.0	1.2	0.3	1.5
	Unknown	0.0	0.0	0.0	0.1	0.0	0.1
Feminine Total		6.8	1.3	1.0	7.8	28.1	45.0
Masculine	Feminine	0.2	0.0	0.5	0.3	0.8	1.9
	Masculine	4.1	0.3	0.7	11.9	25.2	42.2
Masculine Total		4.3	0.3	1.2	12.2	26.1	44.0
Neuter	Neuter	0.7	0.1	0.0	5.2	4.0	9.9
Neuter Total		0.7	0.1	0.0	5.2	4.0	9.9
Unknown	Feminine	0.0	0.0	0.0	0.0	0.1	0.1
	Unknown	0.0	0.0	0.0	0.8	0.1	0.9
Unknown Total		0.0	0.0	0.0	0.8	0.2	1.0
Grand Total		**11.7**	**1.6**	**2.2**	**26.1**	**58.3**	**100.0**

Table A3.3. Percentage of diminutive tokens out of all common nouns in Serbian CDS when children were at the age of 1;8 years, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivisible nouns	Simplex	Grand Total
Feminine	Feminine	5.5	1.2	17.0	2.3	28.2	54.2
	Opaque	0.0	0.0	0.0	0.6	0.1	0.7
Feminine Total		5.5	1.2	17.0	2.9	28.3	54.9
Masculine	Feminine	0.0	0.0	7.1	0.1	1.0	8.2
	Masculine	1.8	0.0	1.0	4.4	20.4	27.6
Masculine Total		1.8	0.0	8.1	4.5	21.4	35.8
Neuter	Neuter	0.2	0.0	0.0	1.7	6.7	8.6
Neuter Total		0.2	0.0	0.0	1.7	6.7	8.6
Unknown	Unknown	0.0	0.0	0.0	0.7	0.0	0.7
Unknown Total		0.0	0.0	0.0	0.7	0.0	0.7
Grand Total		**7.5**	**1.2**	**25.0**	**9.9**	**56.5**	**100.0**

Table A3.4. Percentage of diminutive lemmata out of all common nouns in Serbian CDS when children were at the age of 1;8 years, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivisible nouns	Simplex	Grand Total
Feminine	Feminine	8.0	1.2	2.1	4.4	29.5	45.3
	Opaque	0.0	0.0	0.0	0.9	0.5	1.4
Feminine Total		8.0	1.2	2.1	5.3	30.0	46.7
Masculine	Feminine	0.2	0.0	1.4	0.4	1.1	3.0
	Masculine	3.9	0.0	1.1	9.4	25.2	39.6
Masculine Total		4.1	0.0	2.5	9.8	26.3	42.6
Neuter	Neuter	0.7	0.0	0.0	4.3	4.6	9.6
Neuter Total		0.7	0.0	0.0	4.3	4.6	9.6
Unknown	Unknown	0.0	0.0	0.0	1.1	0.0	1.1
Unknown Total		0.0	0.0	0.0	1.1	0.0	1.1
Grand Total		**12.8**	**1.2**	**4.6**	**20.4**	**60.9**	**100.0**

Table A3.5. Percentage of diminutive tokens out of all common nouns in a sample of Serbian CDS when children were at the age of 2;2 years, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivisible nouns	Simplex	Grand Total
Feminine	Feminine	5.8	1.6	12.1	3.3	30.1	52.8
	Opaque	0.0	0.0	0.0	1.1	0.2	1.3
Feminine Total		5.8	1.6	12.1	4.3	30.3	54.1
Masculine	Feminine	0.1	0.0	6.6	0.1	0.9	7.7
	Masculine	1.8	0.5	0.9	4.5	21.0	28.7
Masculine Total		1.9	0.5	7.5	4.6	21.9	36.4
Neuter	Neuter	0.0	0.0	0.0	1.9	6.9	8.9
Neuter Total		0.0	0.0	0.0	1.9	6.9	8.9
Unknown	Feminine	0.0	0.0	0.0	0.0	0.1	0.1
	Unknown	0.0	0.0	0.0	0.5	0.0	0.6
Unknown Total		0.0	0.0	0.0	0.5	0.1	0.6
Grand Total		7.6	2.0	19.6	11.4	59.3	100.0

Table A3.6. Percentage of diminutive lemmata out of all common nouns in a sample of Serbian CDS when children were at the age of 2;2 years, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivisible nouns	Simplex	Grand Total
Feminine	Feminine	7.3	1.4	1.7	4.7	30.1	45.2
	Opaque	0.0	0.0	0.0	1.1	0.3	1.4
Feminine Total		7.3	1.4	1.7	5.8	30.4	46.6
Masculine	Feminine	0.3	0.0	1.1	0.3	0.9	2.6
	Masculine	4.0	0.5	1.2	10.1	24.9	40.8
Masculine Total		4.4	0.5	2.3	10.4	25.9	43.5
Neuter	Neuter	0.0	0.0	0.0	4.0	5.0	9.0
Neuter Total		0.0	0.0	0.0	4.0	5.0	9.0
Unknown	Feminine	0.0	0.0	0.0	0.0	0.2	0.2
	Unknown	0.0	0.0	0.0	0.6	0.2	0.8
Unknown Total		0.0	0.0	0.0	0.6	0.3	0.9
Grand Total		11.7	1.9	4.0	20.9	61.5	100.0

Table A3.7. Percentage of diminutive tokens out of all common nouns in a sample of Serbian CDS when children were at the age of 2;8 years, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivizible nouns	Simplex	Grand Total
Feminine	Feminine	6.4	1.9	7.0	3.8	34.9	53.9
	Opaque	0.0	0.0	0.0	1.1	0.5	1.6
	Unknown	0.0	0.0	0.0	0.1	0.0	0.1
Feminine Total		6.4	1.9	7.0	5.0	35.4	55.7
Masculine	Feminine	0.0	0.0	4.3	0.1	0.2	4.6
	Masculine	2.2	0.3	0.5	2.8	21.7	27.6
Masculine Total		2.2	0.3	4.8	2.9	21.9	32.3
Neuter	Neuter	0.1	0.0	0.0	2.7	9.2	11.9
Neuter Total		0.1	0.0	0.0	2.7	9.2	11.9
Unknown	Feminine	0.0	0.0	0.0	0.0	0.1	0.1
	Unknown	0.0	0.0	0.0	0.1	0.0	0.1
Unknown Total		0.0	0.0	0.0	0.1	0.1	0.2
Grand Total		**8.7**	**2.2**	**11.8**	**10.7**	**66.6**	**100.0**

Table A3.8. Percentage of diminutive lemma out of all common nouns in a sample of Serbian CDS when children were at the age of 2;8 years, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivizible nouns	Simplex	Grand Total
Feminine	Feminine	7.9	1.4	1.9	5.2	30.8	47.2
	Opaque	0.0	0.0	0.0	1.0	0.6	1.7
	Unknown	0.0	0.0	0.0	0.2	0.0	0.2
Feminine Total		7.9	1.4	1.9	6.4	31.5	49.1
Masculine	Feminine	0.0	0.0	0.8	0.2	0.2	1.2
	Masculine	3.9	0.6	0.8	6.4	25.3	37.1
Masculine Total		3.9	0.6	1.7	6.6	25.5	38.3
Neuter	Neuter	0.2	0.0	0.0	4.3	7.7	12.2
Neuter Total		0.2	0.0	0.0	4.3	7.7	12.2
Unknown	Feminine	0.0	0.0	0.0	0.0	0.2	0.2
	Unknown	0.0	0.0	0.0	0.2	0.0	0.2
Unknown Total		0.0	0.0	0.0	0.2	0.2	0.4
Grand Total		**12.0**	**2.1**	**3.5**	**17.6**	**64.8**	**100.0**

Table A3.9. Percentage of diminutive tokens out of all common nouns in a sample of Serbian CDS when children were at the age of 3;2 years, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivizible nouns	Simplex	Grand Total
Feminine	Feminine	4.7	1.5	12.0	3.7	31.0	53.0
	Opaque	0.0	0.0	0.0	1.2	0.2	1.4
Feminine Total		4.7	1.5	12.0	4.9	31.2	54.4
Masculine	Feminine	0.0	0.0	4.1	0.1	0.8	4.9
	Masculine	1.9	0.8	0.1	4.7	23.0	30.5
Masculine Total		1.9	0.8	4.2	4.7	23.8	35.4
Neuter	Neuter	0.3	0.0	0.0	1.9	7.9	10.0
Neuter Total		0.3	0.0	0.0	1.9	7.9	10.0
Unknown	Unknown	0.0	0.0	0.0	0.1	0.0	0.1
Unknown Total		0.0	0.0	0.0	0.1	0.0	0.1
Grand Total		**6.9**	**2.3**	**16.2**	**11.6**	**62.9**	**100.0**

Table A3.10. Percentage of diminutive lemmata out of all common nouns in a sample of Serbian CDS when children were at the age of 3;2 years, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivizible nouns	Simplex	Grand Total
Feminine	Feminine	8.1	1.3	1.4	4.7	28.3	43.9
	Opaque	0.0	0.0	0.0	1.8	0.5	2.3
Feminine Total		8.1	1.3	1.4	6.5	28.9	46.2
Masculine	Feminine	0.0	0.0	0.9	0.2	0.7	1.8
	Masculine	3.6	0.4	0.2	9.0	27.1	40.3
Masculine Total		3.6	0.4	1.1	9.2	27.8	42.1
Neuter	Neuter	0.7	0.0	0.0	4.5	6.1	11.4
Neuter Total		0.7	0.0	0.0	4.5	6.1	11.4
Unknown	Unknown	0.0	0.0	0.0	0.4	0.0	0.4
Unknown Total		0.0	0.0	0.0	0.4	0.0	0.4
Grand Total		**12.5**	**1.6**	**2.5**	**20.6**	**62.8**	**100.0**

Table A3.11. Percentage of diminutive tokens out of all common nouns in a sample of Serbian CDS when children were at the age of 3;8 years, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivisible nouns	Simplex	Grand Total
Feminine	Feminine	3.1	1.1	6.7	4.8	33.4	49.1
	Opaque	0.0	0.0	0.0	0.9	0.2	1.1
Feminine Total		3.1	1.1	6.7	5.6	33.6	50.2
Masculine	Feminine	0.0	0.0	4.0	0.0	0.6	4.7
	Masculine	1.1	0.5	0.2	5.8	26.9	34.5
Masculine Total		1.1	0.5	4.3	5.9	27.5	39.2
Neuter	Neuter	0.3	0.0	0.0	3.2	7.0	10.6
Neuter Total		0.3	0.0	0.0	3.2	7.0	10.6
Unknown	Unknown	0.0	0.0	0.0	0.0	0.0	0.0
Unknown Total		0.0	0.0	0.0	0.0	0.0	0.0
Grand Total		**4.5**	**1.6**	**11.0**	**14.8**	**68.1**	**100.0**

Table A3.12. Percentage of diminutive lemmata out of all common nouns in a sample of Serbian CDS when children were at the age of 3;8 years, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivisible nouns	Simplex	Grand Total
Feminine	Feminine	4.7	1.2	1.2	6.0	28.5	41.6
	Opaque	0.0	0.0	0.0	1.3	0.4	1.7
Feminine Total		4.7	1.2	1.2	7.3	28.9	43.3
Masculine	Feminine	0.0	0.0	0.7	0.1	0.6	1.4
	Masculine	2.3	0.4	0.4	10.6	30.2	44.0
Masculine Total		2.3	0.4	1.2	10.8	30.8	45.5
Neuter	Neuter	0.7	0.1	0.0	5.0	5.2	11.1
Neuter Total		0.7	0.1	0.0	5.0	5.2	11.1
Unknown	Unknown	0.0	0.0	0.0	0.1	0.0	0.1
Unknown Total		0.0	0.0	0.0	0.1	0.0	0.1
Grand Total		**7.8**	**1.7**	**2.3**	**23.3**	**64.9**	**100.0**

Table A3.13. Distribution of derivational classes out of all common noun tokens in a sample of utterances of mother B.G., across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivisible nouns	Simplex	Grand Total
Feminine	Feminine	5.0	0.7	5.0	4.2	32.6	47.4
	Opaque	0.0	0.0	0.0	0.5	0.4	0.9
Feminine Total		5.0	0.7	5.0	4.7	32.9	48.3
Masculine	Feminine	0.0	0.0	6.2	0.2	1.2	7.6
	Masculine	2.8	0.6	0.5	6.2	23.6	33.7
Masculine Total		2.8	0.6	6.6	6.5	24.8	41.3
Neuter	Neuter	0.0	0.0	0.0	2.2	7.7	9.8
Neuter Total		0.0	0.0	0.0	2.2	7.7	9.8
Unknown	Unknown	0.0	0.0	0.0	0.6	0.0	0.6
Unknown Total		0.0	0.0	0.0	0.6	0.0	0.6
Grand Total		**7.8**	**1.3**	**11.6**	**14.0**	**65.4**	**100.0**

Table A3.14. Distribution of derivational classes out of all common noun lemmata in a sample of utterances of mother B.G., across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivisible nouns	Simplex	Grand Total
Feminine	Feminine	8.3	0.9	1.6	3.9	27.3	42.0
	Opaque	0.0	0.0	0.0	0.9	0.5	1.4
Feminine Total		8.3	0.9	1.6	4.8	27.7	43.4
Masculine	Feminine	0.0	0.0	0.9	0.5	0.9	2.3
	Masculine	4.4	0.5	0.2	10.9	26.6	42.5
Masculine Total		4.4	0.5	1.2	11.3	27.5	44.8
Neuter	Neuter	0.0	0.0	0.0	4.4	6.7	11.1
Neuter Total		0.0	0.0	0.0	4.4	6.7	11.1
Unknown	Unknown	0.0	0.0	0.0	0.7	0.0	0.7
Unknown Total		0.0	0.0	0.0	0.7	0.0	0.7
Grand Total		**12.7**	**1.4**	**2.8**	**21.2**	**61.9**	**100.0**

Table A3.15. Distribution of derivational classes out of all common noun tokens in a sample of utterances of father A.G., across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivisible nouns	Simplex	Grand Total
Feminine	Feminine	4.3	0.6	6.8	5.6	36.1	53.4
	Opaque	0.0	0.0	0.0	0.2	0.0	0.2
Feminine Total		4.3	0.6	6.8	5.8	36.1	53.5
Masculine	Feminine	0.0	0.0	6.5	0.1	1.6	8.2
	Masculine	1.1	0.0	0.3	5.1	24.3	30.9
Masculine Total		1.1	0.0	6.9	5.2	25.9	39.1
Neuter	Neuter	0.0	0.0	0.0	1.4	5.5	7.0
Neuter Total		0.0	0.0	0.0	1.4	5.5	7.0
Unknown	Unknown	0.0	0.0	0.0	0.4	0.0	0.4
Unknown Total		0.0	0.0	0.0	0.4	0.0	0.4
Grand Total		**5.4**	**0.6**	**13.7**	**12.8**	**67.5**	**100.0**

Table A3.16. Distribution of derivational classes out of all common noun lemmata in a sample of utterances of father A.G., across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivisible nouns	Simplex	Grand Total
Feminine	Feminine	5.4	0.8	1.9	5.4	33.2	46.9
	Opaque	0.0	0.0	0.0	0.3	0.0	0.3
Feminine Total		5.4	0.8	1.9	5.7	33.2	47.1
Masculine	Feminine	0.0	0.0	1.4	0.3	1.1	2.7
	Masculine	2.5	0.0	0.5	10.1	27.0	40.1
Masculine Total		2.5	0.0	1.9	10.4	28.1	42.8
Neuter	Neuter	0.0	0.0	0.0	3.8	5.4	9.3
Neuter Total		0.0	0.0	0.0	3.8	5.4	9.3
Unknown	Unknown	0.0	0.0	0.0	0.8	0.0	0.8
Unknown Total		0.0	0.0	0.0	0.8	0.0	0.8
Grand Total		**7.9**	**0.8**	**3.8**	**20.7**	**66.8**	**100.0**

Table A3.17. Percentage of diminutive tokens out of all common nouns in a sample of Serbian ADS, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivisible nouns	Simplex	Grand Total
Feminine	Feminine	0.5	0.1	0.7	5.9	32.7	39.9
	Opaque	0.0	0.0	0.0	1.2	1.6	2.7
	Unknown	0.0	0.0	0.0	0.1	0.0	0.1
Feminine Total		0.5	0.1	0.7	7.1	34.3	42.8
Masculine	Feminine	0.0	0.0	0.4	0.2	0.2	0.8
	Masculine	0.2	0.0	0.2	19.8	23.7	43.8
Masculine Total		0.2	0.0	0.6	20.0	23.9	44.6
Neuter	Neuter	0.0	0.0	0.0	8.4	4.1	12.5
Neuter Total		0.0	0.0	0.0	8.4	4.1	12.5
Unknown	Feminine	0.0	0.0	0.0	0.0	0.0	0.0
	Unknown	0.0	0.0	0.0	0.1	0.0	0.1
Unknown Total		0.0	0.0	0.0	0.1	0.0	0.1
Grand Total		**0.7**	**0.2**	**1.3**	**35.6**	**62.2**	**100.0**

Table A3.18. Percentage of diminutive lemmata out of all common nouns in a sample of Serbian ADS, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Diminutives	Hypocoristics	Nondiminutivisible nouns	Simplex	Grand Total
Feminine	Feminine	0.9	0.2	0.4	7.9	31.0	40.3
	Opaque	0.0	0.0	0.0	2.1	0.5	2.6
	Unknown	0.0	0.0	0.0	0.1	0.0	0.1
Feminine Total		0.9	0.2	0.4	10.0	31.5	43.0
Masculine	Feminine	0.0	0.0	0.2	0.4	0.2	0.9
	Masculine	0.5	0.1	0.2	23.3	18.2	42.3
Masculine Total		0.5	0.1	0.4	23.7	18.4	43.1
Neuter	Neuter	0.1	0.0	0.0	10.2	3.3	13.6
Neuter Total		0.1	0.0	0.0	10.2	3.3	13.6
Unknown	Feminine	0.0	0.0	0.0	0.1	0.0	0.1
	Unknown	0.0	0.0	0.0	0.2	0.0	0.2
Unknown Total		0.0	0.0	0.0	0.3	0.0	0.3
Grand Total		**1.4**	**0.2**	**0.8**	**44.3**	**53.2**	**100.0**

Table A3.19. Percentage of diminutive tokens in a sample of written language, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Dim.	Non-diminutivisible	Simplex	Grand Total
Feminine	Feminine	0.3	0.0	6.7	37.4	44.4
	Opaque	0.0	0.0	2.4	0.4	2.8
Feminine Total		0.3	0.0	9.1	37.8	47.2
Masculine	Feminine	0.0	0.0	0.1	0.0	0.1
	Masculine	0.2	0.0	19.0	14.3	33.5
Masculine Total		0.2	0.0	19.1	14.3	33.6
Neuter	Neuter	0.0	0.0	9.5	9.6	19.1
Neuter Total		0.0	0.0	9.5	9.6	19.1
Grand Total		**0.6**	**0.01**	**37.7**	**61.7**	**100.0**

Table A3.20. Percentage of diminutive lemmata in a sample of written language, across three genders and four declensions.

Gender	Declension	Diminutives	Frozen Dim.	Non-diminutivisible	Simplex	Grand Total
Feminine	Feminine	1	0.05	16.8	15.9	33.7
	Opaque	0	0	5.45	0.05	5.5
Feminine Total		1	0.05	22.25	15.95	39.2
Masculine	Feminine	0	0	0.5	0.1	0.6
	Masculine	0.6	0	35.55	6.5	42.65
Masculine Total		0.6	0	36.05	6.6	43.25
Neuter	Neuter	0.4	0	16.85	0.3	17.55
Neuter Total		0.4	0	16.85	0.3	17.55
Grand Total		**2.0**	**0.1**	**75.2**	**22.9**	**100.0**

Table A3.21. Distribution of all diminutive like endings of words across derivational classes and endings of words for noun tokens and lemmata in Serbian CDS.

Tokens	Diminutives	Frozen Diminutives	Hypocoristic	Non-diminutivisible	Simplex	Total
-ac	0.07	0.00	0.00	0.00	0.00	0.07
-ak	0.01	0.00	0.00	0.47	1.41	1.89
-ce	0.12	0.00	0.00	0.07	0.42	0.62
-če	0.05	0.01	0.00	0.06	0.31	0.42
-ić	1.63	0.39	0.00	0.17	0.01	2.20
-ica	5.08	1.43	0.22	2.09	1.02	9.84
Total	**6.96**	**1.83**	**0.22**	**2.86**	**3.18**	**15.04**

Lemmata	Diminutives	Frozen Diminutives	Hypocoristic	Non-diminutivisible	Simplex	Total
-ac	0.12	0.00	0.00	0.00	0.00	0.12
-ak	0.06	0.00	0.00	0.96	1.81	2.83
-ce	0.36	0.00	0.00	0.18	0.18	0.72
-če	0.30	0.06	0.00	0.06	0.24	0.66
-ić	3.92	0.30	0.00	0.36	0.06	4.64
-ica	6.81	1.20	0.12	4.10	1.02	13.25
Total	**11.57**	**1.57**	**0.12**	**5.66**	**3.31**	**22.23**

Table A3.22. Distribution of all diminutive like endings of words across derivational classes and endings of words for noun tokens and lemmata in Serbian ADS.

Tokens	Diminutives	Frozen Diminutives	Hypocoristic	Non-diminutivisible	Simplex	Total
-ac	0.0	0.0	0.0	1.3	0.2	1.5
-ak	0.0	0.0	0.0	2.2	0.5	2.6
-ce	0.0	0.0	0.0	0.0	0.1	0.1
-če	0.0	0.0	0.0	0.5	0.0	0.5
-ić	0.2	0.0	0.0	0.1	0.0	0.3
-ica	0.5	0.1	0.0	3.1	0.3	4.0
Total	**0.7**	**0.2**	**0.0**	**7.2**	**1.1**	**9.2**

Lemmata	Diminutives	Frozen Diminutives	Hypocoristic	Non-diminutivisible	Simplex	Total
-ac	0.0	0.0	0.0	1.4	0.2	1.6
-ak	0.0	0.0	0.0	2.6	0.9	3.4
-ce	0.1	0.0	0.0	0.1	0.2	0.3
-če	0.0	0.0	0.0	0.1	0.1	0.2
-ić	0.5	0.1	0.0	0.3	0.0	0.9
-ica	0.9	0.2	0.0	4.3	0.4	5.8
Total	**1.4**	**0.2**	**0.0**	**8.7**	**1.8**	**12.2**

Table A3.23. Distribution of all diminutive like endings of words across derivational classes and endings of words for noun tokens and lemmata for written Serbian.

Tokens	Diminutives	Frozen Diminutives	Hypocoristic	Non-diminutivisible	Simplex	Total
-ac	0.0	0.0	0.0	1.1	0.8	1.9
-ak	0.1	0.0	0.0	1.8	0.8	2.7
-ce	0.0	0.0	0.0	0.0	8.4	8.5
-če	0.0	0.0	0.0	0.0	0.0	0.0
-ić	0.1	0.0	0.0	0.0	0.0	0.1
-ica	0.3	0.008	0.0	2.6	0.0	2.9
Total	**0.6**	**0.008**	**0.0**	**5.5**	**10.0**	**16.1**

Lemmata	Diminutives	Frozen Diminutives	Hypocoristic	Non-diminutivisible	Simplex	Total
-ac	0	0	0.0	2.15	0.1	2.25
-ak	0.1	0	0.0	3	0.25	3.35
-ce	0.2	0	0.0	0.25	0.05	0.5
-če	0.2	0	0.0	0.05	0	0.25
-ić	0.5	0	0.0	0	0	0.5
-ica	0.95	0.05	0.0	4.45	0	5.45
Total	**1.95**	**0.05**	**0.0**	**9.9**	**0.4**	**12.3**

Appendix 4

Estimating distributional characteristics of child-directed speech of different languages under controlled laboratory conditions

Nada Ševa , George Hadjiconstantinou, Vera Kempe
University of Stirling

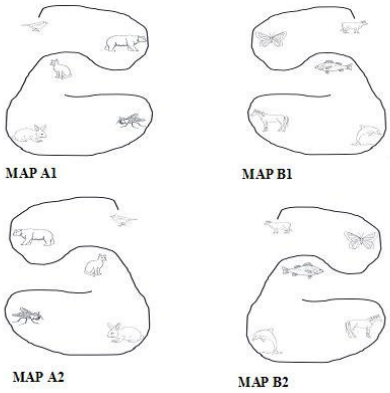
Most studies on morphological characteristics of child-directed speech (CDS) are based on analyses of speech corpora from a limited number of speakers. Such analyses have shown that diminutives are one of the most pervasive features of CDS. This has been attributed to their almost universal pragmatic and semantic characteristics [4], and to their beneficial effects for word segmentation [6] and the acquisition of noun morphology [7]. Still, the frequency of diminutives varies widely across languages. For example, a high frequency of diminutives has been observed in Russian, Spanish, Italian, Dutch, Greek, and Lithuanian CDS [2, 3, 5, 9, 10]. On the other hand, languages like German or Serbian [5, 8, 11] exhibit a low frequency of diminutives in CDS, despite a similar degree of diminutive productivity. However, we cannot exclude the possibility that the cross-linguistic estimates may be flawed due to discrepancies in the representativeness and reliability of the corpora. Moreover, most studies failed to provide an ADS baseline for the same adult speakers, which makes it difficult to assess the relative difference in diminutive usage between ADS and CDS. Considering that thorough quantitative research on representativeness and reliability of spoken corpora is still missing, and that collecting and transcribing spontaneous speech is extremely time consuming and labour intensive, it is unlikely that we will obtain systematically collected samples for more languages in the near future. This study examines whether it is possible to find an alternative, controlled way of collecting CDS and ADS samples, in order to obtain a wider range of cross-linguistic data.

Experimental study

We selected two typologically distinct languages differing in diminutive frequency in CDS: Greek has a high frequency of 32% [10] and Serbian has a low frequency of 8% [11].

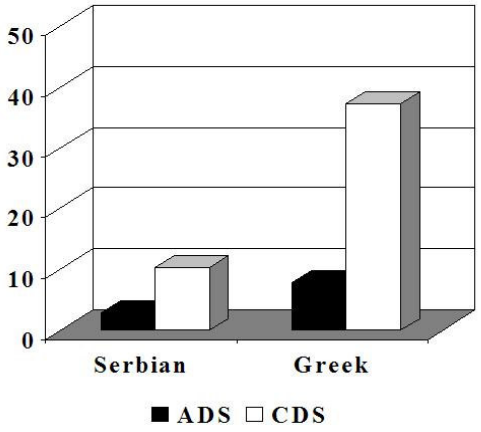
Method

We designed a simplified Map Task [1] that could be used to instruct both an adult and a small child. All maps consisted of five animal pictures connected by a route. For each language, two map versions containing different landmarks were constructed. Within each map version, orientation of the route line was mirror imaged, which gave us 4 different maps in total. Noun gender and length (in syllables) were matched across maps. All nouns were readily diminutivisable. The participants' task was to describe the route so that a prospective listener could re-draw it. All participants completed the task twice within a span of at least 7 days, once describing the route to an imaginary adult, and the other time to an imaginary 2-year old child. Order of speech register (ADS vs. CDS) and map version was counterbalanced across the two sessions. All descriptions were audio-recorded, and coded for number of landmark noun mention and number of diminutives.



	Serbian animal names	Greek animal names
Map A	*ptica*FEM [bird]	*kounelí*NEUT [rabbit]
	*medved*MAS [bear]	*elafí*NEUT [deer]
	*macka*FEM [cat]	*psarí*NEUT [fish]
	*zec*MAS [rabbit]	*petaloudá*FEM [butterfly]
	*muva*FEM [fly]	*alogo*NEUT [horse]
Map B	*krava*FEM [cow]	*poulí*NEUT [bird]
	*leptir*MAS [butterfly]	*delfíni*NEUT [dolphin]
	*riba*FEM [fish]	*pontíki*NEUT [mouse]
	*konj*MAS [horse]	*agelada*FEM [cow]
	*delfin*MAS [dolphin]	*skílos*NEUT [dog]

Participants
Serbian: Twenty-four native Serbian speakers (17 female) mean age, 22.1 (S.D.=4.3, range 18-33 years). were tested at the University of Belgrade.
Greek: Twenty-two native Greek speakers (10 female), mean age 23.4 (S.D =1.8, range 20-27 years), were tested at the University of Stirling.



Results

2 (register: ADS vs. CDS) x 2 (presentation sequence) ANOVAs revealed a main effect of register in Serbian, $F(1,22)=5.345$, $p<0.05$, and in Greek, $F(1,20)=29.35$, $p<0.001$. The effects of presentation sequence as well as the interactions did not reach significance. The results confirm that in both languages, diminutives are used significantly more often in CDS than in ADS. Moreover, using controlled elicitation, we were able to replicate the cross-linguistic differences: The frequency of diminutives in Greek CDS was higher than in Serbian CDS as confirmed by a significant interaction between register and language in a joint analysis ($F(1,44)=13.23$, $p<0.001$).

General Discussion

In sum, controlled elicitation of CDS and ADS can provide a robust estimate of diminutive frequency, and replicate the cross-linguistic differences observed in corpus-analyses. Given the important role of distributional characteristics of CDS in language development, our methodology can be used as a convenient and reliable method for cross-linguistic comparisons of CDS registers, to determine the role of input in language acquisition.

References
[1] Brown, G., Anderson, A., Shillcock, R., & Yule, G. (1984). *Teaching talk: Strategies for production and assessment*. Cambridge, England: Cambridge University Press.
[2] De Marco, A. (1993) "The acquisition of diminutives in Italian" in S. Gillis (Ed.) *Studies in the acquisition of number and diminutive markings*, Papers in Linguistics 9/2, Universiteit Antwerpen, Antwerp, 115-132.
[3] Gillis, S. (1997) The acquisition of diminutives in Dutch. In W. Dressler (ed.) *Studies in pre- and protomorphology*, Vienna. Verlag der Österreichischen Akademie der Wissenschaften, 165-179.
[4] Jurafsky, D. (1996) Universal tendencies in the semantics of the diminutive, *Language*,*72*, 533-578.
[5] Kempe V., Brooks, P. J. & Boom, L. (2001). Slavic child-directed speech facilitates the acquisition of morphology? In M. Almgren, A. Barrena, M-J. Ezeizabarrena, I. Idiazabal, & B. MacWhinney (Eds.), *Research on Child Language Acquisition. Proceedings of the 8th conference of the International Association for the Study of Child Language* (pp. 1237-1247). Medford, MA: Cascadilla Press.
[6] Kempe, V., Brooks, P. J. & Gillis, S. (in press). Diminutives in child-directed speech supplement metric with distributional word segmentation cues. *Psychonomic Bulletin and Review*.
[7] Kempe, V., Brooks, P.J., Mironova, N. and Fedorova, O. (2003). Diminutive-mediated noun-gender acquisition in Russian children. *Journal of child language* **30**, 471-485.
[8] Kernodle-Korell, K. & Dressler, W. (2004). Austrian German diminutives. Paper presented at the *11th International Morphology Meeting*, Vienna, Austria.
[9] Savickiene, I. (1998) "The acquisition of diminutives in Lithuanian" in S. Gillis (Ed.) *Studies in the acquisition of number and diminutive markings*, Papers in Linguistics 9/2, Universiteit Antwerpen, Antwerp, 173-192.
[10] Stephany, U. (1997) Diminutives in early child Greek. In W. U. Dressler (Ed.), Studies in pre- and protomorphology. Vienna.Österreichische Akademie der Wissenschaften.
[11] Seva, N, Kempe, V., Brooks, P.J., Mironova, N. and Fedorova, O. (submitted). Cross-linguistic evidence for the diminutive advantage. Gender agreement in Russian and Serbian children.

Figure A4.1. Detailed presentation of experimental study on elicitation of diminutive usage in controlled experimental situations (poster presentation, Ševa, Hadjiconstantinou, Kempe, 2005).

Appendix 5

Table A5.1. Materials for Gender agreement experiment for the Serbian gender marking task.	
Masculine	**Feminine**
Training nouns	
medveḍ [bear]	mačkạ [cat]
zečić [rabbitdim]	zebrica [zebradim]
pauk [spider]	žaba [frog]
golubičić [pigeondim]	sovicạ [owldim]
Testing nouns	
pingvin / pingvinčić [penguin]	ptica / ptičica [bird]
slon / slončić [elephant]	zmija / zmijica [snake]
krokodil / krokodilčić [crocodile]	pčela / pčelica [bee]
papagaj / papagajčić [parrot]	kornjača / kornjačica [turtle]
lav / lavić [lion]	muva / muvica [fly]
majmun / majmunčić [monkey]	riba / ribica [fish]
leptir / leptirić [butterfly]	buva / buvica [beetle]
konj / konjić [horse]	žirafa / žirafica [giraffe]
žabul / žabulić	lirva / lirvica
pusot / pusotić	sura / surica
rabon / rabonić	brula / brulica
forzak / forzačić	timza / timzica
zirun / zirunić	vigla/viglica
narap / narapić	gljoša / gljošica
cokor / cokorić	krufa / krufica
batus / batusić	mompa / mompica

Appendix 6

Table A6.1. Materials for case-marking experiment with Serbian children.

	Name	**Gloss**	**Type**	**Gender**	**Genitive**	**Dative**	**dim-genitive**	**dim-dative**
1	Kappa	hat	familiar	fem	od kape	prema kapi	od kapice	prema kapici
2	knjiga	book	familiar	fem	od knjige	prema knjizi	od knjižice	prema knjižici
3	Vaza	vase	familiar	fem	od vaze	prema vazi	od vazice	prema vazici
4	Točak	wheel	familiar	mas	od točka	prema točku	od točkića	prema točkiću
5	prsten	ring	familiar	mas	od prstena	prema prstenu	od prstenčića	prema prstenčiću
6	Tanjir	bike	familiar	mas	od tanjira	prema tanjiru	od tanjirića	prema tanjiriću
7	Krpa	cloth	familiar	fem	od krpe	prema krpi	od krpice	prema krpici
8	Čaša	glass	familiar	fem	od čaše	prema čaši	od čašice	prema čašici
9	metla	broom	familiar	fem	od metle	prema metli	od metlice	prema metlici
10	Papir	paper	familiar	mas	od papira	prema papiru	od papirića	prema papiriću
11	Kaiš	belt	familiar	mas	od kaiša	prema kaišu	od kaišića	prema kaišiću
12	krompir	potato	familiar	mas	od krompira	prema krompiru	od krompirića	prema krompiriću
13	marpa		novel	fem	od marpe	prema marpi	od marpice	prema marpici
14	menga		novel	fem	od menge	prema mengi	od mengice	prema mengici
15	tompa		novel	fem	od tompe	prema tompi	od tompice	prema tompici
16	Tober		novel	mas	od tobera	prema toberu	od toberića	prema toberiću
17	Bozil		novel	mas	od bozila	prema bozilu	od bozilića	prema bozilićу
18	ljumin		novel	mas	od ljumina	prema ljuminu	od ljuminića	prema ljuminiću
19	đukla		novel	fem	od đukle	prema đukli	od đuklice	prema đuklici
20	vorpa		novel	fem	od vorpe	prema vorpi	od vorpice	prema vorpici
21	Zinta		novel	fem	od zinte	prema zinti	od zintice	prema zintici
22	frobin		novel	mas	od frobina	prema forbinu	od forbinića	prema forbiniću
23	pagul		novel	mas	od pagula	prema paguli	od pagulića	prema pagulićу
24	ljamer		novel	mas	od ljamera	prema ljameru	od ljamerića	prema ljameriću

Table A6.2. Presentation sequences for the case-marking experiment with Serbian children.

	Presentation Sequence 1		**Presentation Sequence 2**		**Presentation Sequence 3**		**Presentation Sequence 4**	
1	čaša	'od'	krpa	'prema'	točak	'prema'	vazica	'prema'
2	tanjirić	'prema'	ljamerić	'od'	tanjirić	'od'	vazica	'od'
3	pagulić	'prema'	krompirić	'od'	zintica	'od'	krpa	'prema'
4	marpa	'prema'	vazica	'od'	bozil	'prema'	tompica	'od'
5	čaša	'prema'	ljuminić	'od'	knjižica	'prema'	papirić	'prema'
6	zintica	'od'	vorpica	'prema'	točak	'od'	knjižica	'od'
7	bozil	'prema'	kaiš	'od'	tompica	'prema'	menga	'od'
8	metlica	'od'	menga	'prema'	kapa	'prema'	menga	'prema'
9	kaiš	'od'	točak	'od'	metlica	'od'	bozil	'prema'
10	ljuminić	'od'	bozil	'prema'	vazica	'od'	kapa	'prema'
11	prsten	'prema'	pagulić	'prema'	vorpica	'od'	frobin	'prema'
12	đukla	'od'	ljuminić	'prema'	vazica	'prema'	tanjirić	'od'
13	bozil	'od'	kapa	'od'	papirić	'od'	tanjirić	'prema'
14	knjižica	'prema'	zintica	'od'	tober	'prema'	bozil	'od'
15	kaiš	'prema'	menga	'od'	kaiš	'od'	prsten	'prema'
16	prsten	'od'	metlica	'prema'	vorpica	'prema'	vorpica	'prema'
17	menga	'od'	knjižica	'prema'	tober	'od'	krpa	'od'
18	frobin	'prema'	prsten	'od'	ljuminić	'prema'	točak	'prema'
19	marpa	'od'	točak	'prema'	ljamerić	'od'	tompica	'prema'
20	kapa	'prema'	đukla	'prema'	zinitica	'prema'	ljamerić	'od'
21	frobin	'od'	metlica	'od'	marpa	'od'	đukla	'prema'
22	pagulić	'od'	krompirić	'prema'	pagulić	'prema'	frobin	'od'
23	kapa	'od'	vorpica	'od'	krpa	'od'	metlica	'od'
24	metlica	'prema'	ljamerić	'prema'	krompirić	'od'	zinitica	'prema'
25	papirić	'prema'	tanjirić	'od'	kaiš	'prema'	marpa	'od'
26	vorpica	'od'	tober	'prema'	frobin	'od'	čaša	'prema'
27	tanjirić	'od'	papirić	'od'	tanjirić	'prema'	đukla	'od'
28	menga	'prema'	marpa	'prema'	tompica	'od'	krompirić	'od'
29	krpa	'od'	knjižica	'od'	đukla	'prema'	ljuminić	'prema'
30	ljamerić	'od'	marpa	'od'	prsten	'od'	ljuminić	'od'
31	tober	'od'	tanjirić	'prema'	menga	'prema'	papirić	'od'
32	točak	'prema'	vazica	'prema'	kapa	'od'	krompirić	'prema'
33	vazica	'od'	krpa	'od'	marpa	'prema'	točak	'od'
34	krompirić	'od'	papirić	'prema'	krompirić	prema'	pagulić	'od'
35	ljuminić	'prema'	tompica	'od'	ljuminić	'od'	prsten	'od'
36	vazica	'prema'	pagulić	'od'	krpa	'prema'	ljamerić	'prema'
37	đukla	'prema'	prsten	'prema'	frobin	'prema'	marpa	'prema'
38	točak	'od'	čaša	'prema'	bozil	'od'	pagulić	'prema'
39	krpa	'prema'	frobin	'prema'	čaša	'od'	kaiš	'prema'
40	papirić	'od'	zinitica	'prema'	knjižica	'od'	zintica	'od'
41	zinitica	'prema'	bozil	'od'	papirić	'prema'	kaiš	'od'
42	tompica	'od'	čaša	'od'	đukla	'od'	kapa	'od'
43	tompica	'prema'	kapa	'prema'	ljamerić	'prema'	vorpica	'od'
44	tober	'prema'	tober	'od'	metlica	'prema'	tober	'prema'
45	krompirić	'prema'	kaiš	'prema'	menga	'od'	knjižica	'prema'
46	ljamerić	'prema'	đukla	'od'	pagulić	'od'	tober	'od'
47	knjižica	'od'	frobin	'od'	čaša	'prema'	čaša	'od'
48	vorpica	'prema'	tompica	'prema'	prsten	'prema'	vazica	'prema'

	Presentation Sequence 5		**Presentation Sequence 6**		**Presentation Sequence 7**		**Presentation Sequence 8**	
1	prstenčić	‘prema'	krompir	‘prema’	kaišić	‘od'	točkić	‘prema’
2	krompir	od'	zinta	‘prema’	pagul	‘od’	toberić	‘od’
3	kapica	‘prema’	tanjir	‘od’	čašica	‘od’	vorpa	‘prema’
4	kaišić	‘prema’	krpica	‘od’	metla	‘od’	točkić	‘od’
5	vaza	‘prema’	kaišić	‘prema’	vaza	‘od’	ljamer	‘prema’
6	bozilić	‘prema’	vaza	‘od’	kaišić	‘prema’	tanjir	‘od’
7	đuklica	‘prema’	knjiga	‘od’	knjiga	‘od’	marpica	‘od’
8	tompa	‘od’	frobinić	‘prema’	marpica	‘od’	tompa	‘od’
9	zinta	‘prema’	ljumin	‘od’	čašica	‘prema’	tompa	‘prema’
10	tanjir	‘prema’	mengica	‘od’	točkić	‘prema’	kaišić	‘prema’
11	marpica	‘prema’	vorpa	‘prema’	tanjir	‘prema’	toberić	‘prema’
12	ljumin	‘od’	zinta	‘od’	ljumin	‘prema’	frobinić	‘prema’
13	toberić	‘od’	bozilić	‘prema’	ljumin	‘od’	kaišić	‘od’
14	frobinić	‘prema’	vaza	‘prema’	bozilić	‘od’	krompir	‘prema’
15	pagul	‘od’	knjiga	‘prema’	tompa	‘prema’	papir	‘prema’
16	metla	‘od’	marpica	‘prema’	zinta	‘prema’	đuklica	‘prema’
17	knjiga	‘od’	tompa	‘od’	prstenčić	‘prema’	mengica	‘od’
18	krpica	‘od’	kaišić	‘od’	frobinić	‘prema’	bozilić	‘od’
19	tompa	‘prema’	kapica	‘od’	papir	‘prema’	pagul	‘od’
20	mengica	‘prema’	čašica	‘od’	ljamer	‘od’	metla	‘prema’
21	vorpa	‘od’	mengica	‘prema’	vorpa	‘prema’	bozilić	‘prema’
22	točkić	‘prema’	metla	‘prema’	ljamer	‘prema’	kapica	‘od’
23	frobinić	‘od’	metla	‘od’	mengica	‘od’	ljamer	‘od’
24	kaišić	‘od’	prstenčić	‘od’	kapica	‘od’	čašica	‘od’
25	đuklica	‘od’	ljamer	‘prema'	krompir	‘prema’	mengica	‘prema’
26	točkić	‘od'	papir	‘od'	toberić	‘prema’	vaza	od'
27	tanjir	‘od'	pagul	‘prema’	bozilić	‘prema’	đuklica	od'
28	čašica	‘prema’	marpica	‘od'	vorpa	‘od'	metla	od'
29	zinta	od'	toberić	‘od’	prstenčić	‘od’	zinta	‘od’
30	toberić	‘prema’	čašica	prema'	frobinić	‘od’	krompir	‘od’
31	krpica	‘prema’	točkić	prema'	toberić	‘od’	krpica	‘od’
32	mengica	‘od’	točkić	‘od’	đuklica	‘prema’	krpica	‘prema’
33	ljamer	‘od’	bozilić	‘od’	kapica	‘prema’	vorpa	‘od’
34	knjiga	‘prema’	frobinić	‘od’	tanjir	‘od’	ljumin	‘prema’
35	kapica	‘od’	krpica	‘prema’	knjiga	‘prema’	marpica	‘prema’
36	papir	‘od’	toberić	‘prema’	metla	‘prema’	tanjir	‘prema’
37	čašica	‘od’	đklica	‘od’	krompir	‘od’	knjiga	‘prema’
38	pagul	‘prema’	đuklica	‘prema’	zinta	‘od’	papir	‘od’
39	ljamer	‘prema’	pagul	‘od’	marpica	‘prema’	vaza	‘prema’
40	marpica	‘od’	ljumin	‘prema’	tompa	‘od’	kapica	‘prema’
41	krompir	‘prema’	krompir	‘od’	đuklica	‘od’	ljumin	‘od’
42	vorpa	‘prema’	ljamer	‘od’	pagul	‘prema’	knjiga	‘od’
43	vaza	‘od’	vorpa	‘od’	papir	‘od’	frobinić	‘od’
44	ljumin	‘prema’	tompa	‘prema’	vaza	‘prema’	čašica	‘prema’
45	metla	‘prema’	papir	‘prema’	krpica	‘od’	pagul	‘prema’
46	bozilić	‘od’	kapica	‘prema’	krpica	‘prema’	zinta	‘prema’
47	papir	‘prema’	prstenčić	‘prema’	točkić	‘od’	prstenčić	‘prema’
48	prstenčić	‘od’	tanjir	‘prema’	mengica	‘prema’	prstenčić	‘od’

Appendix 7

Table A7.1. Four groups of novel words used in the experiment with artificial suffixes and neural network simulations of the experiment.

	Blue list	**Red list**	**Gender**	**Picture**
Group A	đukla	đukl**upa**	feminine	object
	marp**upa**	marpa	feminine	object
	bozil	bozil**uf**	masculine	object
	frobin**uf**	fronbin	masculine	object
	brola	brol**upa**	feminine	animal
	momp**upa**	mompa	feminine	animal
	batus**uf**	batus	masculine	animal
	forzak	forzak**uf**	masculine	animal
Group B	menga	meng**upa**	feminine	object
	tomp**upa**	tompa	feminine	object
	ljamer	ljamer**uf**	masculine	object
	temir**uf**	temir	masculine	object
	lirva	lirv**upa**	feminine	animal
	timz**upa**	timza	feminine	animal
	žibul	žibul**uf**	masculine	animal
	rabon**uf**	rabon	masculine	animal
Group C	vorpa	vorp**upa**	feminine	object
	zint**upa**	zinta	feminine	object
	pagul	pagul**uf**	masculine	object
	tober**uf**	tober	masculine	object
	virl**upa**	virla	feminine	animal
	sura	sur**upa**	feminine	animal
	pusot	pusot**uf**	masculine	animal
	narap**uf**	narap	masculine	animal
Group D	kela	kel**upa**	feminine	object
	gipn**upa**	gipna	feminine	object
	ljumin	ljumin**uf**	masculine	object
	canup**uf**	canup	masculine	object
	kruf**a**	kruf**upa**	feminine	animal
	gljoš**upa**	gljoš**a**	feminine	animal
	zirun	zirun**uf**	masculine	animal
	cokor**uf**	cokor	masculine	animal

Template nouns	zec (rabbit)	masculine
	mačka (cat)	feminine
	tanjir (plate)	masculine
	viljuška (fork)	feminine

Appendix 8

phoneme	consonant	vocalic	obsturent	sonorant	lateral	continuant	noncontinuant	voiced	voiceless	nasal	labial	conoral	palatal	high	distributed	dorsal
p	1	0	1	0	0	0	1	0	1	0	1	0	0	0	0	0
b	1	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0
t	1	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0
d	1	0	1	0	0	0	1	1	0	0	0	1	0	0	0	0
k	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	1
g	1	0	1	0	0	0	1	1	0	0	0	0	0	0	0	1
f	1	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0
v	1	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0
c	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0
s	1	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0
z	1	0	1	0	0	1	0	1	0	0	0	1	0	0	0	0
ž	1	0	1	0	0	1	0	1	0	0	0	0	0	0	1	0
š	1	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0
dž	1	0	1	0	0	1	1	1	0	0	0	1	0	0	1	0
dj	1	0	1	0	0	1	1	1	0	0	0	1	1	0	0	0
č	1	0	1	0	0	1	1	0	1	0	0	1	0	0	1	0
ć	1	0	1	0	0	1	1	0	1	0	0	1	1	0	0	0
h	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1
m	1	0	0	1	0	0	1	1	0	1	1	0	0	0	0	0
n	1	0	0	1	0	0	1	1	0	1	0	1	0	0	0	0
nj	1	0	0	1	0	1	1	1	0	1	0	0	1	0	0	0
l	1	0	0	1	1	1	0	1	0	0	0	1	0	0	1	0
lj	1	0	0	1	1	1	0	1	0	0	0	0	1	0	1	0
r	1	0	0	1	0	1	0	1	0	0	0	1	0	0	0	0
j	1	0	0	1	0	1	0	1	0	0	0	0	1	0	0	0
i	0	1	0	1	0	1	0	1	0	0	1	1	0	1	0	0
e	0	1	0	1	0	1	0	1	0	0	1	1	0	0	0	0
a	0	1	0	1	0	1	0	1	0	0	0	0	0	0	0	1
o	0	1	0	1	0	1	0	1	0	0	0	0	0	0	0	1
u	0	1	0	1	0	1	0	1	0	0	0	0	0	1	0	1

Table A8.1. Features of Serbian phonemes, used also in Mirković et al,. 2005.