Stem overgeneralizations in the acquisition of Croatian verbal morphology: Evidence from parental questionnaires

Gordana Hržica
University of Zagreb

Tomislava Bošnjak Botica
Institute of Croatian Language and Linguistics

Sara Košutar
UiT The Arctic University of Norway

Abstract

Studies on verbal overgeneralization often focus on languages with low morphological complexity. The Croatian conjugational system exhibits varying degrees of complexity, and this complexity is not primarily based on the number of inflectional morphemes, but on an elaborate system of stem changes. During early language development, children face the difficult task of acquiring this system, using overgeneralized forms to overcome its complexity. To date, studies have used a corpus-based method to retrieve overgeneralizations in child language, which has had limited success in capturing this phenomenon. The aim of this study was to investigate the production of overgeneralized verb forms in Croatian monolingual children aged 2;6 to 5;11 using a questionnaire in which parents report overgeneralizations used by their children. We tested the relationship between the production of overgeneralized forms and features of the input language (token frequency and class size). We hypothesized that the rate of overgeneralizations will depend on input language features, i.e. a higher rate of overgeneralizations for infrequent verbs and for verbs with smaller class size. The items selected for the questionnaire are the verbs with stem change used by parents in the longitudinal *Croatian corpus of child language*. Parents report overgeneralized forms in all verb classes, and verb frequency and class size negatively correlate with the proportion of overgeneralizations. Our results show that children gradually abstract morphological systems in a way that is highly sensitive to the properties of the input.

Keywords: overgeneralization, Croatian verbal morphology, preschool children, token frequency, word class size

1. Introduction

Children begin to acquire inflections as they produce their first words (Clark 2001). A first step in this process is to recognize the morphological structure of the words they hear in the input. To do this, a child must identify the stem and affix and assign the correct meaning to both. Then, a child can use these stems and affixes to form new words. The domain of inflectional morphology also extends beyond the level of a single word, as inflections indicate relationships to other words through grammatical features they encode. During the acquisition of inflectional morphology, children detect patterns based on the words they have already acquired and extend them to form new forms. Gradually, the inflectional paradigms to which certain words belong are established. This is an important step, as generalizations can increase the number of forms acquired (see Ambridge et al. 2012). The most cited example of generalization in the domain of inflectional verbal morphology is the English past simple tense. To form the past tense of regular verbs in English, a child must add the suffix *-ed* to the verb stem (e.g., *jump – jumped*). The general rule of adding the suffix *-ed* can be applied to many verbs. However, English has many irregular verbs that are formed idiosyncratically (e.g., *think-thought*), and this gives rise to unacceptable *overgeneralizations* (e.g., **think – thinked*). Overgeneralizations (also known as *overregularizations*) consist of extending a regular pattern to an irregular form (see Marcus et al. 1992). Children tend to overgeneralize a particular rule in order to cope with the complexity they are confronted with. Overgeneralization of the English past tense is just one example of errors that occur during the acquisition of verbal inflection. In morphologically rich languages, particular forms are not always predictable, also in other tenses, and thus cannot be easily generated by a general rule (e.g., Dąbrowska & Szczerbiński 2006; Savičiūtė et al. 2018; Engelmann et al. 2019; Granlund et al. 2019). In naturalistic studies with longitudinal corpora covering 1–2% of children's speech, such overgeneralizations tend to be low in frequency (up to 5% according to Marcus et al. 1992), but there is convincing evidence that they are more frequent and persistent than reported in these studies (e.g., Maratsos 2000; Maslen et al. 2004). For example, research on Croatian-speaking children has shown that overgeneralized forms persist until at least 7 years of age, although the number of errors decreases with age (Hržica & Lice 2013).

Overgeneralizations reflect the complexity of the morphological system and the strategies children use in mastering it. During language acquisition, children use more than one form per slot in the morphological paradigm (both the correct and overgeneralized forms are attested simultaneously). They also use one of the forms more frequently than the other, which is similar to the phenomenon observed in adult languages called *overabundance* (cf. Thornton 2019: 241–242). The use of two forms in children's language is a temporary side effect of the earlier stages of language acquisition and should disappear as children grow older. How children acquire knowledge of verbal inflection and abandon the use of overgeneralized forms remains an open question, especially for highly inflected languages, which remain under-researched in this regard. There is debate about exactly what factors influence the occurrence of overgeneralization. How does the complexity of a highly inflected language affect the production of incorrect

forms? Does frequency affect the production of incorrect forms in all inflectional paradigms in the same way? Are there other factors that matter? Previous studies have mainly used child language corpora analysis or experimental tasks to answer some of these questions. In this paper, we investigate the factors that influence verbal overgeneralization in Croatian, a morphologically complex language. To do this, we use indirect evidence of overgeneralization, i.e. reports of the use of overgeneralized verbs provided by parents through a questionnaire.

1.1 Theoretical background

Overgeneralizations of inflectional forms in child language are an essential part of language acquisition. Therefore, they have been observed within theoretical approaches that focus on the acquisition of inflectional morphology and, more generally, first language acquisition. Indeed, overgeneralizations and the factors contributing to the appearance and disappearance of overgeneralized forms in child language have been addressed by three different theoretical approaches: the generativist, the constructivist and the pre- and proto-morphological approaches.

Generativist approaches assume that the production of inflected forms in children is governed by innate symbolic rules. Early generativist approaches (e.g., Chomsky & Halle 1968) were very formalistic in their predictions and did not predict overgeneralizations for irregular verbs. Later generativist approaches start from rather modified versions of the nativist assumptions (e.g., Pinker & Prince 1988; Prasada & Pinker 1993; Marcus 1995; Pinker 1998, 1999; Alegre & Gordon 1999; Clahsen 1999). There is a default rule (e.g., addition of the morpheme *-ed*) that can be applied to any regular form when that form is not already available to a child, but some generalizations are also made for irregular verbs. The production of inflected forms is fully productive once a child has learned these forms. Although the main focus in these approaches is on the default rule, many generativist accounts also consider other factors, such as frequency and phonological analogy, which may influence the acquisition of inflections. Stored irregular verbs are produced by phonological analogy, whereas regular verbs are not stored in the lexicon and therefore do not promote generalization by analogy with other similar forms (Prasada & Pinker 1993; cf. Albright & Hayes 2003; Hartshorne & Ullman 2006).

The main assumption of constructivist approaches is that children develop their knowledge of morphology gradually from experience (see Tomasello 2003). Like generativist approaches, constructivist approaches focus on rule-based behaviour but do not state any explicit rule (see McClelland & Rumelhart 1985); rather they claim that rules are abstracted directly from the input (e.g., Bybee & Slobin 1982; Bybee & Moder 1983; Plunkett & Marchman 1991; Bybee 1995, 2001; Bates et al. 1996; Plunkett & Juola 1999; Dąbrowska & Szczerbiński 2006; Aguado-Orea & Pine 2015). During the acquisition of verbal inflection, children abstract over instances (i.e., strings of words and morphemes) of single forms (e.g., *He jumped*), based on which they construct specific morphological schemas (e.g., *He X-ed*). Children can acquire the entire verbal system by forming analogies to all the examples stored in memory (see Ambridge & Lieven 2011).

They can form analogies and correctly use the inflected forms only when the particular schema occurs frequently in the input.

The pre- and proto-morphological approach (e.g., Dressler & Karpf 1995; Dressler et al. 2007; Dressler, 2001; Voeĭkova & Dressler 2002; Bittner et al. 2003; Stephany & Voeĭkova 2009; Stephany & Christofidou 2009; Kenstowicz & Kisseberth 2014; cf. Albright & Hayes 2003) shares some aspects with both the generativist and constructivist approaches but overlaps to a greater extent with the constructivist approach. Inflectional morphology does not emerge all at once but develops gradually in three distinct stages. In the pre-morphological stage, no grammatical morphology can be identified, i.e. early word forms are only rote-learned. During the proto-morphological stage, children recognise morphological patterns of analogies or first rules and construct them creatively. In the stage of "morphology proper", children's systems approach adult models (Bittner et al. 2003: 19). Verbal overgeneralizations appear during the proto-morphological stage, when a child tends to generalize over early inflected forms. Generalizations are usually facilitated by two factors: productivity and transparency. Productivity refers to the extent to which a given word class is likely to get new words. Transparency is defined as the absence of morphologically induced changes within a given paradigm. This approach also emphasizes the type frequency of the declension class as a relevant factor (see studies in Stephany & Voeĭkova 2009), i.e. the number of words belonging to a certain class. This concept is closely related to phonological analogy, but includes the role of type frequency. The more productive and/or transparent a given inflectional class is, with higher type frequency, the less overgeneralization should occur (Bittner et al. 2003; Katičić 2003; cf. Hržica 2012). The pre- and proto-morphological approach shares with the constructivist approach the assumption that the acquisition of inflectional morphology begins with the production of simple base forms (Voeĭkova & Gagarina 2012), leading to the construction of the morphological patterns that can then be used to produce new forms. However, according to the pre- and proto-morphological approach, the different stages of morphological acquisition eventually lead to the system of morphological grammar, governed by symbolic rules, while the constructivist approach assumes that morphological acquisition is based on the construction of schemas of different degrees of generality and abstraction (see Stephany & Christofidou 2009: 219). On the other hand, the assumption of the development of symbolic rules is shared by both the generativist and the pre- and proto-morphological approaches (see Savičiūtė et al. 2018), although it is important to note that the pre- and proto-morphological approach is non-nativist, functional by its nature (Stephany & Christofidou 2009: 219). Unlike the generativist and constructivist approaches, the pre- and proto-morphology approach pays special attention to the influence of typological differences between languages and morphological richness on the order and speed of acquisition of inflectional patterns. The three theoretical approaches differ in their assumptions about the process of acquiring inflectional morphology, which includes overgeneralization. However, they also overlap to a certain extent. They all agree that children must somehow learn inflectional paradigms and that they do so on the basis of the language they acquire and which is presented to them as input. Up to a point, they all agree that some features

of the language/input are relevant to the acquisition of inflectional morphology and that there are therefore certain factors that are relevant to the recovery from the use of overgeneralized forms.

1.2 Factors influencing overgeneralizations

There is good evidence that children produce overgeneralizations in the acquisition of verbal morphology, but not all verbs are overgeneralized in the same way and at the same rate. To figure out the patterns of overgeneralizations, one must primarily examine what increases or decreases the likelihood of a verb's overgeneralization. In the area of verbal inflectional morphology acquisition, two main factors can be identified. The first is *token frequency* (or whole-word frequency), which is the number of times the inflected verb appeared in the input (Ambridge et al. 2015). The higher the frequency with which the target inflected form appears in the input, the lower the error rate. There is ample evidence that frequency predicts the occurrence of overgeneralized inflected forms. Both naturalistic (e.g., Maslen et al. 2004; Aguado-Orea & Pine 2015) and experimental studies (e.g., Theakston et al. 2005; Dąbrowska & Szczerbiński 2006; Räsänen et al. 2016) confirmed that children make more errors when they produce the target forms with low token frequency in the input.

The second factor that the most studies focused on is referred to by different names, for example *class size*, *phonological neighbourhood density*, *phonological analogy* (see Luce & Pisoni 1998). It is usually determined as the number of words with phonologically similar stems that carry the same corresponding inflected morpheme (see Kirjavainen et al. 2012; Savičiūtė et al. 2018; Engelmann et al. 2019; Granlund et al. 2019). In this paper, the term *class size* will be used. Although studies differ in the procedure of defining and determining this phenomenon (Savičiūtė et al. 2018), it is predicted that inflectional patterns with many "class neighbours" will have lower error rates. For example, if a child can recall the first-person singular of the Croatian verb *gledati*.INF 'to watch' *gled-a-m* PRS.1SG 'I watch', he or she can, by analogy, generate the same form of the verb *hodati*.INF 'to walk' *hod-a-m*. PRS.1SG 'I walk'. Many studies have shown that children generalise by analogy and that this process is more effective if there are more words that share the same inflectional pattern (e.g., Albright & Hayes 2003; Dąbrowska & Szczerbiński 2006; Kirjavainen et al. 2012; Räsänen et al. 2016).

Most studies investigated the effects of token frequency and class size in the domain of English verbal morphology (e.g., the English past tense). There is some controversy about how exactly these factors affect the acquisition of verbal morphology in languages with a highly complex inflectional system. Studies in these languages have mainly focused on the case marking of nouns (e.g., Polish: Dąbrowska & Szczerbiński 2006; Krajewski et al. 2011; Dąbrowska 2008; Serbian: Mirković et al. 2011; Lithuanian: Savičiūtė et al. 2018). Few studies investigated the effects of token frequency and class size in the domain of verbal inflection, but some contradictory results have been found. Some studies confirmed a facilitative effect of both factors on the acquisition of inflectional morphology (e.g., Polish: Dąbrowska & Szczerbiński 2006;

Räsänen et al. 2016; Finnish & Polish: Engelmann et al. 2019), while others did not find a facilitative effect of token frequency (e.g., Finnish: Kirjavainen et al. 2012) or class size did not have the same effect on word processing across languages (e.g., English and Spanish: Vitevitch & Stamer 2006), implying that there may be cross-linguistic differences with respect to this phenomenon.

Not all the theoretical approaches mentioned earlier take token frequency and class size equally into account when it comes to the acquisition of inflectional morphology. In generativist approaches, token frequency and class size are usually implicit in the models, which includes notions such as storage of word forms and phonological analogy between them. These factors are considered relevant for the acquisition of irregular verbs, but not for regular verbs. For constructivist approaches, token frequency and class size are central mechanisms for the acquisition of inflectional morphology. For the pre- and proto-morphological approaches, token frequency and class size are as important as paradigm transparency.

We argue that there are different patterns of overgeneralization in morphologically complex languages and that there is no class that can simply be defined as a single standard class. Overgeneralizations can occur both within and outside verb classes, which differ in terms of transparency. The effects of token frequency and class size are relevant to the use of overgeneralized forms even in more transparent classes.

1.3 Methodological aspects of overgeneralization studies

Information on early language production is usually retrieved *via* three different sources: child language corpora or speech/language samples, parental questionnaires, and experimental tasks. Previous studies on overgeneralization mostly used child language corpora (Argus 2009; Hržica 2012, Anđelković & Mirić 2017) or experimental tasks (Dąbrowska & Szczerbiński 2006; Prasada & Pinker 1993; Ambridge 2010), and occasionally the parental questionnaire (Marchman et al. 1997; Frank et al. 2021). Child language can be recorded at predetermined intervals, controlling for environmental factors (e.g., situation, other speakers, etc.). This type of language sampling is common when designing longitudinal corpora of child language, a rich source of diverse information about child language development. This methodology has allowed researchers not only to capture the overgeneralized words produced by children but also to determine how frequently overgeneralized forms occur compared to correct ones. It should be noted, however, that this methodology is generally rich in data, but has limitations in capturing low frequency phenomena (Rowland et al. 2008). For most naturalistic corpora, sampling was sparse, between one and two hours of spontaneous speech per month. This means that the corpus captured 1% or 2% of children's speech (Rowland et al. 2008). Less frequent phenomena in child language, such as over-generalization of words with medium or low frequency, may be underestimated or overlooked altogether. Experimental studies of overgeneralization are also quite common. Experiments in general (without disregarding their diversity) provide specific information about a particular aspect of language. They are well structured, controlled and based on stimuli that require a response from a child. Compared to child language

corpora, they are more precise, but by definition less ecologically valid than the study of naturalistic conversation (e.g., Ambridge & Rowland 2013).

Parental questionnaires are designed and (often) standardized to obtain a general measure of child language development. The most common one, used for many languages, is *MacArthur–Bates communicative development inventories* (CDI; Fenson et al. 2007). Parents are asked whether their child produces certain gestures, words, and morphological forms, and they mark what they find appropriate. The CDI has been used in many studies and has been shown to correlate with laboratory-based measures of children's performance (see Dale et al. 1989; Dale 1991). However, parental questionnaires are not often used in studies focusing on more specific questions of language development. To our knowledge, only recently parental reports in which parents report on the child's language have been used to investigate overgeneralization in language development (Frank et al. 2021). There might be a great potential in using parental questionnaires in such studies. This method can accommodate a large number of items, including those with different frequencies. If distributed online, it can reach a diverse population (e.g., enabling geographical and socioeconomic diversity). The CDI has been shown to have some limitations (reviewed in Ambridge & Rowland 2013) that have to be taken into consideration when extending the methodology. For example, parents tend to underestimate children's knowledge of less prominent aspects of vocabulary (such as verbs compared to nouns, e.g., Tardif et al. 1999). However, 'child-like' forms such as overgeneralizations can be easily identified by parents. It can be concluded that this method could be relevant (or at least comple-mentary) to overgeneralization research.

1.4 Croatian verbal morphology

Most languages within the Indo-European family have two or more conjugation types, depending on the differences between the verb base and/or endings in two or more inflectional forms, typically the infinitive and the present tense. Slavic morphological features are among the most complex with their 4 to 10 conjugation types, and numerous subtypes. Croatian verbs change their conjugation forms depending on the tense, mood number, person, gender or other language-specific factors. They change their infinitive or canonical form in eight inflectional categories: present tense, two simple past tenses (aorist and imperfect), imperative, two gerunds (past and non-past), two verbal adjectives (active and passive). Present, aorist and imperfect tenses, along with verbal adjectives, each have six different endings depending on number (singular, plural) and person (1st, 2nd, 3rd). Person and number are fused and expressed by verbal suffixes. Based on their phonological features, verbs are categorized into groups, types, and classes. The phono-logical forms of the infinitive and the present tense predict the phonological changes of verbs in inflection. There are different approaches to verb categorization and hence different models of conjugational description (e.g., Barić et al. 1997; Silić & Pranjković 2005). According to the latest categorization (Jelaska & Bošnjak Botica 2019), Croatian verbs are divided into 3 groups, containing 10 different classes or types. Groups are formed based on the thematic vowel in present tense, while types are categorized based on

the relation between two stems: the infinitive and the present one. This categorization is partially based on the prototype principle, with the 1st class figuring as the most transparent and productive class, while the last one, the 10th, is the least transparent and completely unproductive. Due to its strong heterogeneity, the 10th class has been additionally subcategorized. The other verb classes are found at different levels of transparency and productivity. Previous studies in Croatian have shown that most of the verbs in child language are overgeneralized following the pattern of the first conjugation class, which is considered to be the most frequent and transparent one (e.g., Hržica 2011, 2012).

In addition to all the inflectional morphemes, further complexity arises from the fact that different verbs in the same (synthetic) inflectional categories behave differently and exhibit an elaborate system of stem alternations. These features make Croatian very suitable for producing a large number of deviations from canonicity within a conjugation paradigm made by children during the early stages of language acquisition. Certain deviations from canonicity – like overabundant forms (two forms realizing the same cell within a paradigm, see Thornton 2019), e.g. NOM.PL. *prsteni* and *prstenovi* 'rings'; PRS.1SG. *šetam* and *šećem* 'I walk') – can be found in adult speakers as well (Lečić 2015; Bošnjak Botica & Hržica 2016; Bošnjak Botica et al. 2022). Croatian verbs have a structure that includes several slots – for more than 94% of verbs there are three obligatory slots in the canonical form: root, thematic suffix (usually a vowel or vowel + consonant) and inflectional morphemes. The root and the thematic suffix form the stem of a verb (1):

(1) *gledati*   *gled-a-ti*
     watch.INF
     gledam   *gled-a-m*
     watch.PRS.1SG

The remaining verbs have only two obligatory slots, root and inflectional morphemes (e.g., *jesti*: *jes-ø-ti* 'to eat'), lacking an overt thematic suffix in the infinitive form. These verbs undergo specific phonological and morphological changes in inflection that originate from the earliest language state and cannot be governed by any contemporary rule. Therefore, this verb class is completely unproductive and is sometimes, for pedagogical purposes, considered irregular (Jelaska & Bošnjak Botica 2019). However, other stem alternations are also evidenced among the majority of other verb classes, as in (2) and (3), and in those classes overgeneralized forms have been attested (Hržica 2012), used by children to overcome the complexity of the system.

(2) *skakati*   *skak-a-ti*
     jump.INF
     skačem   *skač-e-m*   *skakam*   *skak-a-m*
     jump.PRS.1SG   jump.PRS.1SG.ERR

(3) *putovati*   *put-ov-a-ti*
     travel.INF
     putujem   *put-uj-e-m*   *putovam*   *put-ov-a-m*
     travel.PRS.1SG   travel.PRS.1SG.ERR

Some previous studies on the acquisition of morphology in Croatian (e.g., Anđel et al. 2000; Xanthos et al. 2011) are in line with the Natural Morphology approach (Dressler 1985; Dressler & Karpf 1995; Dressler 2005), which implies differences in children's morphological behaviour as a function of language typology and system complexity, claiming that children who speak morphologically rich languages respond earlier to specific structural features of their language. According to Dressler (2005: 8), morphological richness is related to the amount of productive morphology, and special attention according to the Natural Morphology approach is given to the productivity of morphological patterns. A study conducted on verb forms in more inflected languages showed that overgeneralization can be found in both stem and verb suffixes (e.g., Clahsen et al. 2002 for Spanish; Gagarina 2008 for Russian; Hržica 2012 for Croatian; Anđelković & Mirić 2017 for Serbian). Therefore, it is worth investigating how the complexity of the verbal system (understood, following Dressler (2005), as a morphological richness accompanied by unproductive morphology) in highly inflected languages affects the production of overgeneralized forms.

When searching for the phenomenon of overgeneralization in verb paradigms in the literature, the most represented language one comes across mainly is English and its past tense (e.g., *goed* instead of *went*; *bringed* instead of *brought*). Despite their morphological richness, Slavic languages seem to be less represented, although some research is conducted either directly, observing overgeneralizations (Hržica 2012; Anđelković & Mirić 2017) or indirectly, within a broader topic of verbal acquisition (Anđel et al. 2000; Katičić 2003; Gagarina 2008; Hržica & Lice 2013; Chejnová 2018).

As far as Croatian is concerned, the first findings on the acquisition of verbal morphology and morphology in general, with emphasis on inflectional errors that can be considered as overgeneralizations, were obtained within the Vienna research group on pre- and proto-morphology (Anđel et al. 2000; Katičić 2003). Although lacking sufficient data that can provide us with more reliable conclusions, those studies indicated children's preferences for more transparent suffixes, as well as for fully transparent and productive verb classes. The first study that focused on overgeneralizations in verb paradigms in Croatian confirmed the crucial role of transparency of a given paradigm revealing that the concept has more levels (Hržica 2012). The main goal was to determine the ratio of overgeneralizations among different conjugation types, which allows deeper insights into the phonological and morphological properties of some forms that encourage over-generalizations. The same study also confirmed that Croatian children acquire the rich inflectional system of verb endings quite early, while more difficulties are encountered in the formation of the stem. Several patterns of overgeneralization were found. Most often the overgeneralized form was formed following the pattern of the verb classes without a stem change. For example, the verb *plakati* 'cry.INF' has the regular form *plačem* 'cry.PRS.1.SG', but is overgeneralized as **plakam* 'cry.PRS.1.SG' by analogy with *gledati* 'watch.INF' from *gledam* 'watch.PRS.1.SG'. There is also a significant number of verbs in which one morphological form (e.g., participle) is formed by using a more frequent stem (e.g., present tense stem). For example, the verb *donijeti* 'bring.INF' has the regular form *donio* 'bring.PTCP.M.SG' and the irregular form **doneseo* 'bring.PTCP.M.SG'. The overgeneralized form was formed after the present tense stem of the same verb

(*donesem* 'bring.PRS.1.SG'). The study conducted on Serbian suggests that typologically similar languages show similar patterns of overgeneralizations (Anđelković & Mirić 2017).

In the course of the acquisition of morphology, children tend to generalize forms using analogy patterns in both the stem and/or the final part of the verb and to follow the most transparent model. Although important as indicators of tendencies in the acquisition of verbs in Croatian, all three above-mentioned studies are based on corpus data up to the age of 3 and with a very small number of children investigated (i.e., one to three), which prevents us from drawing more reliable conclusions. When it comes to possible factors that might play a role in this process, such as frequency or analogy, there is no unanimous conclusion about their role.

1.5 Aim and hypotheses

The aim of this study is to test the production of overgeneralized verb forms in Croatian monolingual children ages 2;6 to 5;11, using the questionnaire in which parents report overgeneralizations of their children. Previous research has confirmed the impact of frequency and class size on the production of overgeneralized forms in children (e.g., Kirjavainen et al. 2012; Engelmann et al. 2019). Not many studies have investigated the influence of these factors in morphologically rich languages, especially on the acquisition of verbal morphology (see Granlund et al. 2019). We aimed to investigate the relationship between the production of overgeneralized verb forms and features of the input language (token frequency and class size). We hypothesize that parents will report overgeneralized forms in all verb classes in which the stem changes, but that the frequency of overgeneralizations will depend on the features of the input, i.e. it will be predicted by the frequency of verbs (higher rate of overgeneralizations for infrequent verbs) and class size in lemmas, tokens and tokens of selected types (higher rate of overgeneralizations for verbs with smaller class size).

2. Method

2.1 Participants

Ethical approval for the research was obtained from the Ethical committee of the Faculty of Education and Rehabilitation Sciences University of Zagreb. Participants were recruited via social media. Only parents of children aged 2;6 to 5;11 years were invited to participate in the study. The invitation included a link to the SurveyMonkey online platform where participants could access a consent form. The consent form included a clear and detailed description of the purpose of the study, specific procedures related to the study, information about minimal risks, benefits, confidentiality, and the possibility to withdraw from the study. Altogether, 87 parents finished the questionnaire, therefore, we obtained data for 87 children: five 2-year-olds, thirty-three 3-year-olds, twenty-five 4-year-olds and twenty-four 5-year-olds. The data about participants is presented in

Table 1: Number and age of participants (N = 87)

	Age of participants				
N of participants	Age group	Min	Max	M	SD in months
5	Two–year–olds	2;8	2;11	2;10	1.4
33	Three–year–olds	3;0	3;11	3;6	3.5
25	Four–year–olds	4;1	4;11	4;5	3.4
24	Five–year–olds	5;0	5;11	5;5	3.5

Note: N = number; Min = minimum; Max = Maximum; M = mean; SD = standard deviation

Table 1. This age range was chosen on the basis of previous studies in Croatian which have shown that children do not fully master verbal morphology before the age of 3, i.e. overgeneralized forms were attested after this age (e.g., Anđel et al. 2000; Hržica 2012). However, parents of younger children were the least interested in participating in the study, so there is not much information about the youngest participants. All parents of preschool children were approached via a social media campaign. Thus, it is not clear whether parents of younger children are simply not interested in such studies or whether there is some other reason that led to a low number of participants in this group.

2.2 Materials

For the purpose of this study, an online parental questionnaire was developed. Parents were asked to indicate how often their child produces a particular verb form (correct and overgeneralized) using a 5-point Likert scale. Selected verbs belonged to several classes with the stem change. Overgeneralized forms were designed either by:

a. Analogy to the most transparent verb class (1st class), e.g. for the verb *plakati* 'cry.INF' the regular form *plačem* 'cry.PRS.1.SG' and the overgeneralized form **plakam* 'cry.PRS.1.SG' were offered. The overgeneralized form was formed following the pattern of the verb classes without a stem change (e.g., *gledati* 'watch.INF' from *gledam* 'watch.PRS.1.SG').

b. Use of the more frequent verb stem of a particular verb, e.g. for the verb *donijeti* 'bring.INF' regular form *donio* 'bring.PTCP.M.SG' and irregular form **doneseo* 'bring.PTCP. M.SG' were offered. The overgeneralized form was formed after the present tense stem of the same verb (*donijeti* 'bring.INF' regular form *donesem* 'bring.PRS.1.SG' ∼ irregular form **doneseo* 'bring. PTCP.M.SG').

2.2.1 *Selection of verbs*

In selecting items for the questionnaire, we first had to determine which verbs of certain conjugation classes might be overgeneralized by children. We selected six conjugation classes with stem changes (which makes about 25% of all verbs). The corresponding verb examples are given in (4). Five of the verb classes are thematic (4th, 5th, 6th, 7th and 8th), while the 10th consists of athematic verbs. Other verb classes with stem change (1st, 2nd,

3rd and 9th) were not considered for the questionnaire. Previous corpus-based studies including Croatian children up to the age of three have not provided evidence for stem overgeneralizations in these classes in the present tense (see Anđel et al. 2000; Hržica 2012).

Verbs belonging to 4th and 6th classes change only their thematic vowel ($a > i$; $a > e$), while verbs belonging to 5th, 7th and 8th classes also change the consonant. In addition to preserving a thematic vowel in the present tense form, the 10th class also involves adding or deleting a particular consonant within the stem. Therefore, to produce a certain morphological form in all selected verb classes, a child must perform two tasks: attach the appropriate inflectional morpheme to the verb stem and change the stem.

(4) Stem alternations in six conjugation classes selected for the questionnaire

4th	*trčati*	*trčim*
	run.INF	run.PRS.1.SG
5th	*plakati*	*plačem*
	cry.INF	cry.PRS.1.SG
6th	*penjati*	*penjem*
	climb.INF	climb.PRS.1.SG
7th	*putovati*	*putujem*
	travel.INF	travel.PRS.1.SG
8th	*davati*	*dajem*
	give.INF	give.PRS.1.SG
10th	*donijeti*	*donesem*
	bring.INF	bring.PRS.1.SG
	doći	*dođem*
	come.INF	come.PRS.1.SG

In the next step, we had to determine which verbs of a given conjugation classes children regularly encounter in their language environment, so we focused on language sources that provide at least partial information about spoken language.

The verbs for the questionnaire were selected from the *Croatian corpus of child language* (Kovačević 2002) to ensure that they occur in the child's language environment, i.e. child-directed speech (CDS). The corpus includes language samples of spontaneous interaction of three children and adult speakers in their family environment. It was collected during the period of early language acquisition from the onset of speech to approximately 3 years of age. 1269 verbs were found, 397 of which belonged to classes with stem changes. Most verbs were given in their basic, non-prefixed form and we considered only verbs that occurred at least 3 times. We used two lexical databases to refine the selection: *Croatian lexical database* (Kuvač Kraljević & Olujić 2018) and *Croatian psycholinguistic database* (Peti-Stantić et al. 2018). The databases contain information on estimated age of acquisition and subjective word frequency retrieved from native speakers. Only verbs with the age of acquisition up to 5 and a subjective frequency of at least 3 ("very often") were selected. Based on these criteria, a total of 36 verbs were included in the study. The verbs belong to the 4th class ($N = 6$), 5th class ($N = 16$), 6th class ($N = 2$), 7th class ($N = 1$), 8th class ($N = 1$), and 10th class ($N = 10$).

Table 2: Overgeneralized forms designed for the questionnaire

Class	Infinitive	Correct form	Overgeneralized form
4th	*trčati* run.INF	*trčim* run.PRS.1.SG	**trčam* run.PRS.1.SG
5th	*plakati* cry.INF	*plačem* cry.PRS.1.SG	**plakam* cry.PRS.1.SG
6th	*penjati* climb.INF	*penjem* climb.PRS.1.SG	**penjam* climb.PRS.1.SG
7th	*putovati* travel.INF	*putujem* travel.PRS.1.SG	**putovam* travel.PRS.1.SG
8th	*davati* give.INF	*dajem* give.PRS.1.SG	**davam* give.PRS.1.SG
10th	*donijeti* bring.INF	*donio* bring.PTCP.M.SG	**donesao* **doneseo* **donesio* bring.PTCP.M.SG

2.2.2 *Designing the questionnaire*

In the questionnaire, parents had to estimate the frequency with which their child uses a particular verb form. For all verbs, both the correct and the assumed overgeneralized form were presented (see Table 2). For the 4th, 5th, 6th, 7th, and 8th class overgeneralized forms that a child might produce were created based on our predictions that children retain the infinitive stem when producing present tense forms if the verb ends in *-ati*, analogous to the most transparent 1st class of verbs. For the 10th class, we assume that children overgeneralize either by:

(a) using the present tense pattern and the thematic vowel of the 1st type (*-a-*)
    donijeti   *donio*   **donesao*
    bring.INF   bring.PTCP.M.SG   bring.PTCP.M.SG
(b) using the present tense stem
    donijeti   *donio*   **doneseo*
    bring.INF   bring.PTCP.M.SG   bring.PTCP.M.SG
(c) using the imperative stem
    donijeti   *donio*   **donesio*
    bring.INF   bring.PTCP.M.SG   bring.PTCP.M.SG

For each verb, the parents had to estimate how often their child produces certain form, ranging from 1 (never) to 5 (very often) on a 5-point Likert scale, as shown in (5).

(5) How often does your child use *trčim* 'I run' (including forms *trčiš* 'you run', *trči* 'she runs', *trčimo* 'we run'…)?
    1 – never   2 – rarely   3 – sometimes   4 – often   5 – very often

It is important to note that items were displayed randomly, so participants estimated each form (correct and overgeneralized) independently.

2.3 Variables for the analysis

2.3.1 *Frequency of verbs*

Verb frequency was calculated from the two sources, one being a longitudinal child language corpus and a second corpus of written adult language.

In previous studies on overgeneralizations in language acquisition, both corpus-based and experimental (e.g., Argus 2009; Hržica 2012; Engelmann et al. 2019; Granlund et al. 2019), objective frequency was calculated from available corpora of children's language, using only utterances produced by adults, as these were assumed to be representative of the input children receive. For this study, child-directed speech (CDS) from the *Croatian corpus of child language* (Kovačević 2002) was used.

Based on previous research on Croatian child language, we assumed that overgeneralizations are also present in the language of children aged 3 to 6 (e.g., Hržica 2012), so we included parents of children older than those who participated in the corpus of child language. It would have been preferable to use corpora that provide more information about the input that three-, four-, and five-year-old children receive – namely, corpora of spontaneous speech of children of this age – but such corpora are not available in Croatian. Therefore, we also included the largest available adult language corpus, as it was done in previous studies (Ambridge & Pine 2016; Kirjavainen et al. 2009). The *Croatian web corpus* (*hrWaC*; Ljubešić & Klubička 2014) consists of texts collected from the Internet, that range from books and newspapers portals to forms such as blogs and forums. The current version of the corpus (v2.0) contains 1.9 billion tokens and is annotated with lemma, morphosyntax and dependency syntax layers. Objective frequency was calculated for each verb lemma using the *NoSketch Engine* (Rychlý 2007; Kilgarriff et al. 2014).

From the two corpora, we obtained two types of frequency information for each of the preselected verbs. The first was the overall frequency of all morphological types of a verb. The second was the frequency of the specific morphological verb types that were central to our study. For example, for verbs where overgeneralization occurs only in the present tense, we calculated the frequency of forms in the present tense. Information on the frequency of verbs from the questionnaire can be found in Table 3. Each of the verbs selected for the questionnaire occurred at least once in selected morphological types in the corpus of child language (in child directed speech), and at least 50 times in selected morphological types in the *hrWaC*.

2.3.2 *Calculating class size of verbs*

The class size was calculated from the two sources, the longitudinal child language corpus and the adult written language corpus.

Table 3: Frequency of all/specific morphological types of verbs from the questionnaire based on child-directed speech (CDS) in *Croatian corpus of child language* (Kovačević 2002) and *hrWaC* (Ljubešić & Klubička 2014)

	Min	Max	*M*	*SD*
Croatian corpus of child language (CDS) – overall frequency	1	567	64	118
Croatian corpus of child language (CDS) – frequency of selected types	1	116	18	26
hrWaC – overall frequency	7,500	5,670,049	389,280	979162
hrWaC – frequency of selected types	56	1,080,732	107,409	202,935

Note: Min – lowest frequency of the individual verb selected for the questionnaire; Max – highest frequency of the individual verb selected for the questionnaire; *M* – mean frequency of verbs selected for the questionnaire; *SD* – standard deviation

In previous studies on overgeneralizations in language acquisition (e.g., Argus 2009; Hržica 2012; Engelmann et al. 2019; Granlund et al. 2019), the class size (referred to also as *phonological neighbourhood density*) was calculated from available corpora of children's language, taking into account only utterances produced by adults. We used the *Croatian child language corpus* (Kovačević 2002) to extract all verbs from child directed speech ($N = 1352$) and we assigned them to morphological classes. The procedure was repeated for the *hrWaC* corpus (Ljubešić & Klubička 2014). All verbs (lemmas) that appeared in the corpus with a frequency of at least 5 ($N = 11.114$) were assigned to a corresponding morphological verb class (according to Jelaska & Bošnjak Botica 2019), and we added subclasses in classes 5 and 10 based on phonological criteria (see Table 4). Class 5 consists of verbs that are palatalised (iotated) in the formation of the present tense stem. Several consonants are palatalised in different verbs. However, since the frequency of occurrence of these patterns can change, we have divided class 5 into 6 subclasses, each consisting of verbs in which the same consonant is palatalised. Class 10 is traditionally considered the least regular class of Croatian verbs. It consists of different patterns of stem change. Therefore, we have treated each type of stem change as a separate subclass. The class size of each verb included in the questionnaire was calculated in lemmas and in tokens.

2.4 Procedure

Prior to commencing the research, a short pilot study was conducted to assess the validity of the questionnaire and to estimate the time required for participants to complete it. After parents had provided informed consent, they were given a link to the questionnaire. Their task was to rate the frequency with which their child uses a particular verb form on a 5-point Likert scale. The questionnaire took about 10 to 15 minutes. The questions were randomly selected, and respondents could leave the platform and edit their answers afterwards. Each respondent was presented with the

Table 4: Class size of each verb included in the questionnaire

	Croatian child language corpus – CDS			hrWaC	
	Lemmas	Tokens	Tokens of selected types	Lemmas	Tokens
Class 4 *trč-a-ti trč-i-m* run.INF run.PRS.1SG	23	463	259	93	120873
Class 5					

Subclass 1
  bris-a-ti briš-e-m
  wipe.INF wipe.PRS.1SG
	17	353	153	61	62173

Subclass 2
  šet-a-ti šeć-e-m
  walk.INF walk.PRS.1SG
	14	45	37	85	13340

Subclass 3
  rez-a-ti rež-e-m
  cut.INF cut.PRS.1SG
	34	607	279	94	114830

Subclass 4
  pomag-a-ti pomaž-e-m
  help.INF help. PRS.1SG
	29	142	51	94	114830

Subclass 5
  plak-a-ti plač-e-m
  cry.INF cry.PRS.1SG
	13	234	118	56	9318

Subclass 6
  jah-a-ti jaš-e-m
  ride.INF ride.PRS.1SG
	7	127	79	20	2698

Class 6 *penj-a-ti penj-e-m* climb.INF climb.PRS.1SG	21	212	127	146	99148
Class 7 *put-ov-a-ti put-uj-e-m* travel.INF travel.PRS.1SG	30	97	54	770	1725344
Class 8 *da-va-ti da-je-m* give.INF give.PRS.1SG	4	45	34	31	41782
Class 10					

Subclass 3
  uze-∅-ti uzm-e-m
  take.INF take.PRS.1SG
	7	314	85	13	27332

Continued

Table 4: *Continued*

	Croatian child language corpus – CDS			hrWaC	
	Lemmas	Tokens	Tokens of selected types	Lemmas	Tokens
Subclass 5 *donije-∅-ti dones-e-m* bring.INF bring.PRS.1SG	6	350	159	16	35410
Subclass 7 *jes-∅-ti jed-e-m* eat.INF eat.PRS.1SG	48	9331	368	39	53427
Subclass 9 *obu-∅-ći obuč-e-m* dress.INF dress PRS.1SG	51	5238	1094	48	91429
Subclass 10 *mo-∅-ći mog-u* can.INF can.PRS.1SG	3	3303	323	3	658091

Table 5: Reported overgeneralized usage per verb

	Percentage of parents who reported overgeneralized usage of a verb			
	Min	Max	*M*	*SD*
Per verb	16%	64%	34%	13.001

Note: Min = minimal percentage of parents per verb who reported that their children use overgeneralized form of a verb; Max = maximal percentage of parents per verb who reported that their children use overgeneralized form of a verb; *M* = mean percentage of parents per verb who reported that their children use overgeneralized form of a verb; *SD* = standard deviation

same set of questions. Parents received a language profile of their child as a gesture of thanks for participating in the study.

2.5 Data analysis

We observed the percentage of parents who reported overgeneralized forms for each verb (regardless of the point on the Likert scale, as long as it is not 'never') to examine which verbs are used in both correct and overgeneralized forms. The frequency of both the correct and overgeneralized forms was included in the analysis. The results were also presented based on the morphological verb classes. Spearman's correlation was used to examine the relationship between the class size and frequency of a given verb with: (a) the estimated subjective frequency of using the correct form of the verb; (b) the estimated subjective frequency of using the overgeneralized form of the verb (both measured using the Likert scale); (c) the proportion of using the overgeneralized form (ratio of the frequency of the overgeneralized form in the sum of the frequency of both the correct

and overgeneralized forms of a verb). The proportion of the overgeneralized form was calculated as a separate measure because the overall frequency of the overgeneralized form might depend on the overall frequency of the verb. Frequencies were taken from the longitudinal child language corpus and the written adult language corpus, including total frequency per verb and frequency of selected (targeted) morphological forms per verb. Class size was taken from the longitudinal child language corpus and the written adult language corpus, including class size per lemma, per token, and per selected (targeted) morphological type. Due to the large differences in relative numbers, both frequency and class size were converted to ranks.

3. Results

Descriptive statistics for the overall report of overgeneralizations is presented in Table 5. As can be seen from the table, parents reported on overgeneralized forms in child language for all the verbs included in the questionnaire. Each verb included in the questionnaire was reported as overgeneralized by some parents. The percentage of parents who reported overgeneralized usage is minimally 16% per verb, and the maximum reaches 64% per verb. On average, each verb was reported as overgeneralized by 34% of parents.

Parents reported overgeneralized forms in all verb classes. Overgeneralized forms of verbs were most frequently reported for classes 8, 6, 5 and 10 (over 30%), although it needs to be taken into account that some classes or subclasses consist of only one verb. Table 6 shows percentages of parents who reported overgeneralized forms of a verb per verb class.

Although all verbs included in the questionnaire were reported as overgeneralized, the frequency with which they were reported by parents varied. An overview of the responses at the different levels of the Likert scale can be found in Table 7. When reporting on overgeneralized usage of a verb, the parents most frequently chose the option 'Never'. Other options were chosen less frequently, with the option 'Very often' being chosen least frequently.

The occurrence of overgeneralized forms can vary greatly with age. We tested overgeneralizations in children of a wide age range (Table 8). The average results show that parents recognised overgeneralizations, although this is a rare phenomenon in children's speech. Overgeneralized forms were reported for children of all ages and for every child who participated in the study, at least some forms were reported as overgeneralized. In this study, overgeneralized forms were not shown to be prevalent in younger age groups. Rather, it appears that they are more common in the middle age group, which is consistent with the hypothesised U-shaped curve of language development. The frequency of overgeneralization obviously decreases with age, but what we do not yet know for morphology-rich languages is when exactly children withdraw from overgeneralization.

We tested the relationship between the objective frequency of verbs (frequency was calculated as a sum of frequencies of selected (targeted) morphological types) retrieved from child-directed speech (CDS) in the *Croatian child language corpus* (Kovačević 2002)

Table 6: Reported overgeneralized usage per verb class

		Percentage of parents who reported overgeneralized usage of a verb			
	N	*Min*	*Max*	*M*	*SD*
Class 4 *trč-a-ti trč-i-m* run.INF run.PRS.1SG	6	19%	32%	26%	4.918
Class 5	16	25%	58%	38%	9.511

Subclass 1
    bris-a-ti briš-e-m
    wipe.INF wipe.PRS.1SG
		28%	48%	36%	10.111

Subclass 2
    šet-a-ti šeć-e-m
    walk.INF stroll.PRS.1SG
		44%	44%	44%	–

Subclass 3
    rez-a-ti rež-e-m
    cut.INF cut.PRS.1SG
		25%	52%	38%	9.365

Subclass 4
    plak-a-ti plač-e-m
    cry.INF cry.PRS.1SG
		29%	29%	29%	–

Subclass 5
    jah-a-ti jaš-e-m
    ride.INF ride.PRS.1SG
		34%	58%	43%	12.701

Class 6 *penj-a-ti penj-e-m* climb.INF climb.PRS.1SG	2	22%	51%	37%	20.223
Class 7 *put-ov-a-ti put-uj-e-m* travel.INF travel.PRS.1SG	1	20%	20%	20%	–
Class 8 *da-va-ti da-je-m* give.INF give.PRS.1SG	1	52%	52%	52%	–
Class 10 *mo-0-ći, mog-u* can.INF can.PRS.1SG	10	16%	64%	32%	17.128

Note: N = number of verbs per class; Min = minimal percentage of parents per verb class who reported that their children use overgeneralized form of a verb; Max = maximal percentage of parents per verb class who reported that their children use overgeneralized form of a verb; *M* = mean percentage of parents per verb class who reported that their children use overgeneralized form of the verb; *SD* = standard deviation

Table 7: Parent's reports of children's usage of overgeneralized forms of verbs at different levels of the Likert scale

Levels of the Likert scale	
Never	73%
Rarely	12%
Sometimes	8%
Often	5%
Very often	2%

Table 8: Descriptive statistics for overgeneralized forms in three age groups

Age group	Min	Max	M	SD
2;8–3;11	1.1	2.8	1.8	0.378
4;0–4;11	1.5	3.6	2.2	0.557
5;0–5;11	1.2	2.7	2	0.411

Note: Min = minimum; Max = maximum; M = mean; SD = standard deviation

Table 9: Correlations of objective frequency of verbs, only for targeted morphological forms, based on CDS (*Croatian corpus of child language*, Kovačević 2002) and *hrWaC* (Ljubešić & Klubička 2014), and subjective frequency of corrects forms, overgeneralized forms and the proportion of overgeneralized forms (estimated by parents)

Variables	Frequency in *Croatian corpus of child language* (CDS)	Frequency in *hrWaC*
Frequency of correct forms	0.635**	0.482**
Frequency of overgeneralized forms	−0.035	−0.076
Proportion of overgeneralized forms	−0.459**	−0.345*

Note: 95% confidence intervals are shown in brackets.
** $p < 0.01$, * $p < 0.05$

and *hrWaC* (Ljubešić & Klubička 2014), and three measures in the questionnaire based on parents' estimations: frequency of correct forms, frequency of overgeneralized forms and the proportion of overgeneralized forms. The results are presented in Table 9.

The results showed a significant moderate positive correlation between the frequency of verbs in CDS and the frequency of correct forms estimated by parents ($r_s(36) = 0.635**$, $p < 0.01$). There was a significant moderate positive correlation between the frequency of verbs in *hrWaC* and the frequency of correct forms estimated by parents ($r_s(36) = 0.482**$, $p < 0.01$). We also found a significant negative moderate correlation between the proportion of overgeneralized forms and the objective verb frequency in CDS ($r_s(36) = -0.459**$, $p < 0.01$) and in *hrWaC* ($r_s(36) = -0.345*$, $p < 0.05$). Other correlations were not significant.

Table 10: Correlations of class size in lemmas, tokens and tokens of selected types, based on CDS (*Croatian corpus of child language*, Kovačević 2002) and *hrWaC* (Ljubešić & Klubička 2014), and subjective frequency of corrects forms, overgeneralized forms and the proportion of overgeneralized forms (estimated by parents)

Variables	Frequency of correct forms	Frequency of overgeneralized forms	Proportion of overgeneralized forms
Class size in lemmas (CDS – *Croatian child language corpus*)	−0.127	−0.480[**]	−0.179
Class size in lemmas (*hrWaC*)	−0.210	−0.099	0.020
Class size in tokens (CDS – *Croatian child language corpus*)	0.306	−0.416[*]	−0.498**
Class size in tokens (*hrWaC*)	−0.304	−0.328	−0.414*
Class size in tokens of selected types (CDS – *Croatian child language corpus*)	0.370*	−0.376[*]	−0.507**

Note: 95% confidence intervals are shown in brackets.

** $p < 0.01$, * $p < 0.05$

Finally, we tested the relationship between the class size in lemmas, tokens and tokens of selected types, based on CDS (*Croatian corpus of child language*, Kovačević 2002) and *hrWaC* (Ljubešić & Klubička 2014), and three measures from the questionnaire based on parents' estimations: frequency of correct forms, frequency of overgeneralized forms and the proportion of overgeneralized forms.

As can be seen from Table 10, there was a significant negative moderate correlation between the class size in lemmas (CDS) and the frequency of overgeneralized forms ($r_s(36) = −0.480$, $p < 0.01$). We found a significant negative moderate correlation between class size in tokens (CDS) and the frequency of overgeneralized forms ($r_s(36) = −0.416*$, $p < 0.05$), as well as between class size in tokens (CDS) and the proportion of overgeneralized forms ($r_s(36) = −0.498**$, $p < 0.01$). Class size in tokens in *hrWaC* also positively moderately correlated with the proportion of overgeneralized forms ($r_s(36) = −0.414*$, $p < 0.05$). Class size determined for selected (targeted) morphological types positively moderately correlated with the frequency of correct forms ($r_s(36) = 0.370$, $p < 0.05$), negatively correlated with the frequency of overgeneralized forms ($r_s(36) = −0.376$, $p < 0.05$), as well as with the proportion of overgeneralized forms ($r_s(36) = −0.507**$, $p < 0.01$). Other correlations did not reach the significance.

4. Discussion

The aim of this study was to investigate the production of overgeneralized verb forms in Croatian monolingual children aged 2;6 to 5;11, using a questionnaire in which parents report overgeneralizations of their children. We aimed to find out whether children of

this age produce overgeneralized forms of verbs belonging to stem-changing verb classes and what the relationship between the production of overgeneralized verb forms and features of the input language (token frequency and class size) is. More specifically, we expected higher rates of overgeneralization for rare verbs and higher rates of overgeneralization for verbs with smaller class size. Previous research on overgeneralizations was mainly based on corpus or experimental methods and only occasionally the parental questionnaire.

Parents reported varying frequencies of use of overgeneralized forms for verbs, but each verb was reported as overgeneralized by at least 16% of parents. These results are consistent with our first hypothesis, which states that parents report overgeneralized forms for all verbs. Looking at these results within the stem-changing classes, the average percentage of reported overgeneralized forms varies. For most classes, it ranges from 25% to 40%. Future studies should investigate whether these differences are significant when other factors (such as frequency) are controlled for, and whether the difference in class transparency matters.

We examined the relationship between the verb frequency and the use of overgeneralized forms. Parents reported that children more often overgeneralize low-frequency verbs, and less often high frequency verbs. This is true for both frequency counts: one was obtained from the child language corpus (input data) and the other from the adult written language corpus. Our results are consistent with previous studies that have confirmed a facilitating effect of frequency on the acquisition of inflectional morphology. Part of these studies focused mainly on the acquisition of noun inflection (e.g., Dąbrowska & Szczerbiński 2006; Krajewski et al. 2011; Granlund, 2019). In the studies that investigated the role of frequency in the acquisition of verbal morphology, token frequency appeared to be an important factor influencing overgeneralization. Engelmann et al. (2019) found a robust effect of token frequency on the production of person/number verb forms in Finnish and Polish children aged 3 to 5 years. Räsänen et al. (2016) tested Finnish-speaking children aged 2 to 4 years. In that study, error rates in marking person/number were higher for low-frequency verb forms. Kirjavainen et al. (2012) found a significant facilitatory effect of lemma frequency on the production of (semi-)productive verb forms in Finnish children aged 4 to 6 years. However, the effect of token frequency was inhibitory, i.e. children performed better when presented with words of low frequency. A similar effect has been observed in other studies. Clahsen et al. (2004), for example, found that German children produced participle forms with lower frequency faster than those with higher frequency.

Next, we examined the relationship between the class size of verbs and the use of overgeneralized forms. The results showed a negative relationship between class size and both the frequency and the proportion of overgeneralizations. Parents reported that children more often overgeneralize verbs that belong to smaller classes, and the less often overgeneralize verbs that belong to larger classes, i.e. less overgeneralization is reported for patterns that are more frequent in the language surrounding children. This is true for class size in tokens, and class size in tokens of selected (targeted) types, both derived from the corpus of child language and the adult language corpus. Similar observations

were made in previous studies, although the measure was defined differently. Kirjavainen et al. (2012) found that children made fewer errors when they produced verbs with a higher number of phonological friends (defined as clusters of verbs that differed in only one phoneme). Engelmann et al. (2019) observed a robust effect of class size (members of the same class similar in both stem and inflected form). Finally, Räsänen et al. (2016) confirm that children rely on phonological analogy, i.e. lower error rates were observed for verbs with higher phonological neighbourhood density (defined as morphological class size). In contrast, Vitevitch and Stamer (2006) did not find the same effect of PND (clusters of words with similar sounds) on the processing of spoken nouns in Spanish and English and concluded that differences in morphology between two languages affect processing.

It is important to note that we considered different measures of frequency and class size, some based on the corpus of child language, others based on the written corpus. The data from the parental questionnaire showed a higher correlation with frequency and class size from the child language corpus, suggesting that although the children in this study are older than the children from the Croatian child language corpus, this source seems to be more relevant in this and similar studies. Frequency and class size also showed better correlations when calculated on tokens of specific morphological forms, i.e. those where overgeneralization is expected. Parents reported that children produce a higher proportion of overgeneralization in verbs with lower frequency and a higher proportion of overgeneralization in verbs with lower class size. In more precise measures of frequency and class size, targeting selected morphological forms and dividing verbs into a larger number of classes, the correlations were higher, contributing to the general idea that the language surrounding children is important for the acquisition of inflectional morphology.

The results of this study, which highlight both token frequency and class size as relevant factors in the production of overgeneralization, contribute to theoretical debates on the acquisition of inflectional morphology in general. Generativist approaches take frequency and phonological analogy into account when dealing with the acquisition of irregular verbs (verbs not inflected according to the default rule). However, they consider that these factors are generally not that important for the acquisition of inflectional morphology in general. Constructivist approaches, on the other hand, consider both the token frequency and the frequency of occurrence of a given schema (i.e., class size) in the input to be relevant for acquisition. The pre- and proto-morphological approach considers type frequency, i.e. the size of the specific inflectional paradigm, to be one of the most important factors in the acquisition of inflectional morphology, together with token frequency.

Most work on overgeneralization of verbs has focused on the regular vs. irregular past tense in English (but see Räsänen et al. 2016; Kirjavainen et al. 2012; Engelmann at al. 2019). In this study, parents reported that their children produced overgeneralized forms according to two different patterns. While some overgeneralized forms were formed by adhering to the most transparent and largest class size, others were formed by using the more frequent verb stem of the specific verb. The latter cannot be explained by simply conforming to the rule of a default class, which is a pattern of overgeneralization

suggested by dual route generativist approaches. Another pattern of overgeneralization of verbs can be explained by the frequency of the form: the more frequently a child hears a form with a particular stem, the more likely he or she is to overgeneralize that form into another morphological form that would require a different stem. Our results suggest that the frequency of a particular verb and the class size of the verb play a role in both overgeneralization patterns. The more frequently the child hears the verb and the more frequently the child hears similar verbs, the lower the proportion of overgeneralized forms, and this is true for different patterns of overgeneralization. This suggests that these factors are relevant to the acquisition of morphology, even if we leave behind the debate about regular vs. irregular verbs. In this respect, the results of this study support theoretical approaches that assume that the acquisition of morphology is generally regulated by input, namely frequency and class size, such as connectionist models of morphology (e.g., Marchman et al. 1997; Meunier & Marslen-Wilson 2004; Mirković et al. 2011).

The results of this study could also support approaches that predict multiple rules for the acquisition of morphology. Albright and Hayes (2003) find that verbs, whether regular or irregular, form islands of different rules and that these islands respond to type frequency (class size) but not to token frequency. On the other hand, the pre- and proto-morphological approach (e.g., Bittner et al. 2003; Stephany & Voeĭkova 2009) also assumes the existence of multiple rules, but these rules are generated by analogy and respond to both class size and token frequency. Our results, where both patterns of overgeneralization correlated with both class size and token frequency, would contribute to the pre- and proto-morphological approach. The methodology of this study differs from previous research. While previous studies have relied on corpus-based research, experiments or computational models, here a parental questionnaire was used. As far as we know, parental questionnaires in which parents report on the child's language have not been used before to study overgeneralization in language development. This method allowed us to accommodate a large number of items, including those with different frequencies and class sizes. It reached a large number of participants from different parts of Croatia and allowed for their geographical and socio-economic diversity. Child overgeneralizations were easily detected by parents who had been trained through a one-week observation period. The results obtained are broadly comparable with previous findings and can be interpreted in the light of earlier studies that differed in methodology. It can be concluded that this methodology could be relevant (or at least complementary) to the study of overgeneralization and could also be useful for other studies dealing with the acquisition of inflectional morphology.

This work has certain limitations. There are other factors that could influence the frequency of overgeneralized forms and that need to be considered, such as morphosyntactic features, the semantics of the verb, and the context in which the verb is generally used. In the present study, verbs were selected according to the morphological classes in which overgeneralized forms are likely to occur, but this factor could play a crucial role. Future work should consider differences in the frequency of over-generalization in different morphological classes, especially in highly inflected languages where the phenomenon has not yet been studied closely in this respect.

5. Conclusions

During the acquisition of verbal inflectional morphology, children are confronted with the complexity of the language to which they are exposed. The result of such complexity is the occurrence of overgeneralized forms, that is, a child extends the general pattern for producing the particular verb form to the verb forms that should not be produced according to that pattern. Two factors are known to influence the occurrence of over-generalized forms, namely token frequency (how often a child is exposed to a particular verb form) and class size (the number of verbs with phonologically similar word stems bearing the same corresponding inflectional morpheme). Previous studies on over-generalization in verbal morphology acquisition have generally focused on languages with lower morphological complexity (mainly English). The phenomenon has not been studied in depth in highly inflected languages and some contradictory results have also been found. Morphological features of Croatian, due to the richness of the system which exhibits varying degrees of complexity, are a fertile ground to study the production of overgeneralized forms.

In the present study, verbal overgeneralization in Croatian was investigated using a novel method. The results show that preschool children of different ages still resort to this mechanism to overcome the complexity of verbal morphology. A relationship was also found between the frequency of overgeneralization and factors considered to be relevant to it in previous studies based on corpus analysis or experimental work. We can conclude, then, that the data collected from parents show similar trends to studies suggesting that children gradually abstract morphological systems in a way that is highly sensitive to the properties of the input.

References

Aguado-Orea, Javier & Pine, Julian M. 2015. Comparing different models of the development of verb inflection in early child Spanish. *PloS one* 10(3). 1–21.

Albright, Adam & Hayes, Bruce. 2003. Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* 90(2). 119–161.

Alegre, Maria & Gordon, Peter. 1999. Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40(1). 41–61.

Ambridge, Ben. 2010. Children's judgments of regular and irregular novel past-tense forms: new data on the English past-tense debate. *Developmental Psychology* 46(6). 1497–1504.

Ambridge, Ben & Kidd, Evan & Rowland, Caroline F. & Theakston, Anna L. 2015. The ubiquity of frequency effects in first language acquisition. *Journal of Child Language* 42(2). 239–273.

Ambridge, Ben & Lieven, Elena. 2011. *Child language acquisition: Contrasting theoretical approaches*. Cambridge: Cambridge University Press.

Ambridge, Ben & Pine, Julian M. & Rowland, Caroline F. & Chang, Franklin & Bidgood, Amy. 2012. The retreat from overgeneralization in child language acquisition: word learning, morphology, and verb argument structure. *Wiley Interdisciplinary Reviews: Cognitive Science* 4(1). 47–62.

Ambridge, Ben, & Rowland, Caroline F. 2013. Experimental methods in studying child language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science* 4(2). 149–168.

Anđel, Maja & Klampfer, Sabine & Kilani–Scoch, Marianne & Dressler, Wolfang U. & Kovačević, Melita. 2000. Acquisition of verbs in Croatian, French and Austrian German – an outline of a comparative analysis. *Suvremena lingvistika* 49–50(1). 5–25.

Anđelković, Darinka & Mirić, Mirjana. 2017. Distribution of verbal overgeneralizations in the Serbian Corpus of Early Child Language. *Psihološka istraživanja* 20(2). 291–310.

Argus, Reili. 2009. The early development of case and number in Estonian. In Stephany, Ursula & Voeĭkova, Maria D. (eds.), *Development of nominal inflection in first language acquisition*, 111–152. Berlin, New York: De Gruyter Mouton.

Barić, Eugenija & Lončarić, Mijo & Malić, Dragica & Pavešić, Slavko & Peti, Mirko & Zečević, Vesna & Znika, Marija. 1997. *Hrvatska gramatika*. Zagreb: Školska knjiga.

Bates, Elizabeth & Elman, Jeffrey & Johnson, Mark M. & Karmiloff-Smith, Annette & Parisi, Domenico & Plunkettt, Kim. 1996. *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Bittner, Dagmar & Dressler, Wolfang U. & Kilani-Schoch, Marianne. 2003. Introduction. In Dagmar Bittner & Wolfang U. Dressler & Marianne Kilani-Schoch (eds.), *Development of verb inflection in first language acquisition: A cross-linguistic perspective*, vii–xxxviii. Berlin, New York: De Gruyter Mouton.

Bošnjak Botica, Tomislava & Hržica, Gordana. 2016. Overabundance in Croatian dual-class verbs. *Fluminensia* 28(1). 83–106.

Bošnjak Botica, Tomislava & Polančec, Jurica & Sviben, Robert. 2022. Korpusno istraživanje hrvatskih imenica s dugom i kratkom množinom. *Jezikoslovlje* 23(1). 35–74.

Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5). 425–455.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan & Slobin, Dan. 1982. Rules and schemas in the development and use of the English past tense. *Language* 58(2). 265–289.

Bybee, Joan & Moder, Carol L. 1983. Morphological classes as natural categories. *Language* 59(2). 251–270.

Chejnová, Pavla. 2018. Child acquisition of Czech modal verbs. *Ad alta, journal of interdisciplinary research*. 7–9.

Chomsky, Noam & Halle, Morris. 1968. *The sound pattern of English*. New York, Evanston and London: Harper & Row.

Clahsen, Harald. 1999. Lexical entries and rules of language: a multi-disciplinary study of German inflection. *Behavioral and Brain Sciences* 22. 991–1060.

Clahsen, Harald & Aveledo, Fraibet & Roca, I. 2002. The development of regular and irregular verb inflection in Spanish child language. *Journal of Child Language* 29(3). 591–622.

Clahsen, Harald & Hadler, Meike & Weyerts, Helga. 2004. Speed production of inflected words in children and adults. *Journal of Child Language* 31(3). 683–712.

Clark, Eve V. 2001. Morphology in language acquisition. In Spencer, Andrew & Zwicky, Arnold M. (eds.), *The handbook of morphology*, 374–389. Oxford: Blackwell Publishing.

Dąbrowska, Ewa. 2008. The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language* 58 (4). 931–951.

Dąbrowska, Ewa & Szczerbiński, Marcin. 2006. Polish children's productivity with case marking: the role of regularity, type frequency, and phonological diversity. *Journal of Child Language* 33(3). 559–597.

Dale, Philip S. & Bates, Elizabeth & Reznick, J. Steven & Morisset, Colleen. 1989. The validity of a parent report instrument of child language at twenty months. *Journal of Child Language* 16(2). 239–249.

Dale, Philip S. 1991. The validity of a parent report measure on vocabulary and syntax at 24 months. *Journal of Speech & Hearing Research* 34(3). 565–571.

Dressler, Wolfang U. 1985. Typological aspects of Natural Morphology. *Wiener linguistische Gazette* 36. 3–26.

Dressler, Wolfang U. 2001. The emergence of morphology - a Constructivist approach. *Hrvatska revija za rehabilitacijska istraživanja* 37(1). 23–36.

Dressler, Wolfang U. 2005. Morphological typology and first language acquisition: some mutual challenges. In Booij, Geert & Guevara, Emiliano & Ralli, Angela & Sgroi, Salvatore & Scalise, Sergio (eds.), *Morphology and linguistic typology, on-line proceedings of the Fourth Mediterranean Morphology Meeting* (MMM4), 7–21. University of Bologna. http://morbo.lingue.unibo.it/mmm/

Dressler, Wolfang U. & Karpf, Annemarie. 1995. The theoretical relevance of pre- and protomorphology in language acquisition. In Booij, Geert & van Marle, Jaap (eds.), *Yearbook of Morphology 1994*. 99–122. Dordrecht: Springer.

Dressler, Wolfang U. & Stephany, Ursula & Aksu-Koç, Ayhan & Gillis, Steven &. 2007. Discussion and conclusion. In Laaha, Sabine & Gillis, Steven (eds.), *Antwerp papers in linguistics 112*, *typological perspectives on the acquisition of noun and verb morphology*, 67–71. https://repository.uantwerpen.be/docman/irua/24dd68/63855.pdf

Engelmann, Felix & Granlund, Sonia & Kolak, Joanna & Szreder, Marta & Ambridge, Ben & Pine, Julian & Theakston, Anna & Lieven, Elena. 2019. How the input shapes the acquisition of verb morphology: Elicited production and computational modelling in two highly inflected languages. *Cognitive Psychology* 110. 30–69.

Fenson, Larry & Marchman, Virginia A. & Thal, Donna J. & Dale, Phillip S. & Reznick, J. Steven & Bates, Elizabeth. 2007. *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Brookes.

Frank, Michael C. & Braginsky, Mika & Yurovsky, Daniel & Marchman, Virginia A. (2021). *Variability and consistency in early language learning: the Wordbank Project*. Cambridge, MA: MIT Press.

Gagarina, Natalia. 2008. *First language acquisition of verb categories in Russian*. St. Petersburg: Nauka.

Granlund, Sonia & Kolak, Joanna & Vihman, Virve & Engelmann, Felix & Lieven, Eelena V. M. & Pine, Julian M. & Theakston, Anna L. & Ambridge, Ben. 2019. Language-general and language-specific phenomena in the acquisition of inflectional noun morphology: A cross-linguistic elicited-production study of Polish, Finnish and Estonian. *Journal of Memory and Language* 107. 169–194.

Hartshorne, Joshua K. & Ullman, Michael T. 2006. Why girls say 'holded' more than boys. *Developmental Science* 9(1). 21–32.

Hržica, Gordana 2011. *Glagolske kategorije aspekta, vremena i akcionalnosti u usvajanju hrvatskog jezika*. Zagreb: Filozofski fakultet, Sveučilište u Zagrebu (Doctoral dissertation).

Hržica, Gordana 2012. Daj mi to napisaj: preopćavanja glagolske osnove u usvajanju hrvatskog jezika. *Suvremena lingvistika* 38(74). 189–208.

Hržica, Gordana & Lice, Karolina 2013. Morfološke pogreške u uzorcima govornog jezika djece urednog jezičnog razvoja i djece s posebnim jezičnim teškoćama. *Hrvatska revija za rehabilitacijska istraživanja* 49(1). 65–77.

Jelaska, Zrinka & Bošnjak Botica, Tomislava. 2019. Conjugational types in Croatian. *Rasprave Instituta za hrvatski jezik i jezikoslovlje* 45(1). 47–74.

Katičić, Antigone. 2003. Early verb development in one Croatian-speaking child. In Dagmar Bittner, Wolfang U. Dressler & Marianne Kilani-Schoch (eds.), *Development of verb inflection in first language acquisition*, 239–267. Berlin: De Gruyter.

Kenstowicz, Michael & Kisseberth, Charles. 2014. *Generative phonology: description and theory*. Burlington: Elsevier Science.

Kilgarriff, Adam & Baisa, Vít & Bušta, Jan & Jakubíček, Miloš & Kovář, Vojtěch & Michelfeit, Jan & Rychlý, Pavel & Suchomel, Vít. 2014. The Sketch Engine: ten years on. *Lexicography ASIALEX* 1. 7–36.

Kirjavainen, Minna & Nikolaev, Alexandre & Kidd, Evan. 2012. The effect of frequency and phonological neighbourhood density on the acquisition of past tense verbs by Finnish children. *Cognitive Linguistics* 23(2). 273–315.

Kovačević, Melita. 2002. *Croatian corpus*, CHILDES. http://childes.psy.cmu.edu/data/Slavic

Krajewski, Grzegorz, Theakston, Anna L., Lieven, Elena V. M., & Tomasello, Michael 2011. How Polish children switch from one case to another when using novel nouns: Challenges for models of inflectional morphology. *Language and Cognitive Processes* 26(4–6). 830–861.

Kuvač Kraljević, Jelena & Olujić, Marina. 2018. Imageability and subjective frequency of the 500 rated nouns in the Croatian Lexical Database. *Suvremena lingvistika* 44(85), 73–90.

Lečić, Dario. 2015. Morphological doublets in Croatian: the case of the instrumental singular. *Russian Linguistics* 39(3), 375–393.

Ljubešić, Nikola & Klubička, Filip. 2014. {bs,hr,sr}WaC – Web Corpora of Bosnian, Croatian and Serbian. In Bildhauer, Felix & Schäfer, Roland (eds.), *Proceedings of the 9th Web as Corpus Workshop (WaC-9) @ EACL 2014*. 29–35.

Luce, Paul A. & Pisoni, David B. 1998. Recognizing spoken words: the neighbourhood activation model. *Ear and Hearing* 19(1). 1–36.

Maratsos, Michael. 2000. More overregularizations after all: new data and discussion on Marcus, Pinker, Ullman, Hollander, Rosen & Xu. *Journal of Child Language* 27(1), 183–212.

Marchman, Virginia & Plunkett, Kim & Goodman, Judith. 1997. Overregularization in English plural and past tense inflectional morphology: A response to Marcus (1995). *Journal of Child Language* 24(3). 767–779.

Marcus, Gary F. 1995. Children's overregularization of English plurals: a quantitative analysis. *Journal of Child Language* 22(2). 447–460.

Marcus, Gary F. & Pinker, Steven & Ullman, Michael & Hollander, Michelle & Rosen, John & Xu, Fei & Clahsen, Harald. 1992. Overregularization in language acquisition. *Monographs of the Society for Research in Child Development* 57(4), I–178.

Maslen, Robert J. C & Theakston, Anna L. & Lieven, Elena V. M., & Tomasello, Michael. 2004. A dense corpus study of past tense and plural overregularization in English. *Journal of Speech, Language, and Hearing Research* 47. 1319–1333.

McClelland, James L. & Rumelhart, David E. 1985. Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General* 114. 159–188.

Meunier, Fanny & Marslen-Wilson, William D. 2004. Regularity and irregularity in French verbal inflection. *Language and Cognitive Process* 19(4). 561–580.

Mirković, Jelena & Seidenberg, Mark S. & Joanisse, Marc F. 2011. Rules versus statistics: insights from a highly inflected language. *Cognitive Science* 35. 638–681.

Peti-Stantić, Anita & Anđel, Maja & Keresteš, Gordana & Ljubešić, Nikola & Stanojević, Mateusz-Milan & Tonković, Mirjana. 2018. Psiholingvističke mjere ispitivanja 3000 riječi hrvatskoga jezika: konkretnost i predočivost. *Suvremena lingvistika* 85. 91–112.

Pinker, Steven. 1998. Words and rules. *Lingua* 106(1–4). 219–242.

Pinker, Steven. 1999. *Words and rules: the ingredients of language*. New York: Basic Books.

Pinker, Steven & Prince, Alan. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28(1–2). 73–193.

Plunkett, Kim & Juola, Patrick. 1999. A connectionist model of English past tense and plural morphology. *Cognitive Science* 23(4). 463–490.

Plunkett, Kim & Marchman, Virginia. 1991. U-shaped learning and frequency effects in a multi-layered perception: implications for child language acquisition. *Cognition* 38(1). 43–102.

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

Räsänen, Sanna H. M. & Ambridge, Ben & Pine, Julian. M. 2016. An elicited-production study of inflectional verb morphology in child Finnish. *Cognitive Science* 40. 1704–1738.

Rowland, Caroline & Fletcher, Sara & Freudenthal, Daniel. 2008. How big is big enough? Assessing the reliability of data from naturalistic samples. In Behrens, Heike (ed.), *Corpora in language acquisition research: history, methods, perspectives*, 1–24). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Rychlý, Pavel. 2007. Manatee/Bonito-A modular corpus manager. In Sojka, Petr & Horák, Aleš (eds.), *RASLAN 2007 Recent advances in Slavonic natural language processing*. 65–70. Brno: Masaryk University.

Savičiūtė, Egle & Ambridge, Ben & Pine, Julian M. 2018. The roles of word-form frequency and phonological neighbourhood density in the acquisition of Lithuanian noun morphology. *Journal of Child Language* 45(3). 641–672.

Silić, Josip & Pranjković, Ivo. 2005. *Gramatika hrvatskoga jezika*. Zagreb: Školska knjiga.

Stephany, Ursula & Christofidou, Anastasia. 2009. The emergence of nominal inflection in Greek. In Stephany, Ursula & Voeĭkova, Maria D. (eds.), *Development of nominal inflection in first language acquisition*. 217–264. Berlin, New York: De Gruyter Mouton.

Stephany, Ursula & Voeĭkova, Maria D. 2009. Introduction. In Stephany, Ursula & Voeĭkova, Maria D. (eds.), *Development of nominal inflection in first language acquisition: A cross-linguistic perspective*, 1–14. Berlin: Mouton de Gruyter.

Tardif, Twila & Gelman, Susan & & Xu, Fan. 1999. Putting the "Noun Bias" in context: a comparison of English and Mandarin. *Child Development* 70(3). 620–635.

Theakston, Anna L. & Lieven, Elena V. & Pine, Julian M. & Rowland, Caroline F. 2005. The acquisition of auxiliary syntax: BE and HAVE. *Cognitive Linguistics* 16(1). 247–277.

Thornton, Anna M. 2019. Overabundance: a canonical typology. In Franz Rainer & Francesco Gardani & Wolfgang U. Dressler & Hans Christian Luschützky (eds.), *Competition in inflection and word-formation*. 223–258. Studies in Morphology 5. Dordrecht: Springer.

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

Vitevitch, Michael S. & Stamer, Melissa K. 2006. The curious case of competition in Spanish speech production. *Language, Cognition and Neuroscience* 21(6). 760–770.

Voeĭkova, Maria D. & Dressler, Wolfgang U. 2002. *Pre- and protomorphology: early phases of morphological development in nouns and verbs*. München: Lincom.

Voeĭkova, Maria D. & Gagarina, Natalia. 2012. Early syntax, first lexicon and the acquisition of case forms by two Russian children. In Maria D. Voeĭkova & Wolfgang. U. Dressler (eds.), *Pre- and protomorphology: early phases of morphological development in nouns and verbs*. 115–131. Munich: Lincom Europa.

Xanthos, Aris & Laaha, Sabine & Gillis, Steven & Stephany, Ursula & Aksu-Koç, Ayhan & Christofidou, Anastasia & Gagarina, Natalia & Hrzica, Gordana & Ketrez, F. Nihan & Kilani-Schoch, Marianne & Korecky-Kröll, Katharina & Kovačević, Melita & Laalo, Klaus & Palmović, Marijan & Pfeiler, Barbara & Voeikova, Maria D. & Dressler, Wolfgang U. 2011. On the role of morphological richness in the early development of noun and verb inflection. *First Language* 31(4). 461–479. https://doi.org/10.1177/0142723711409976.