



Cognitive Science 35 (2011) 638–681

Copyright © 2011 Cognitive Science Society, Inc. All rights reserved.

ISSN: 0364-0213 print / 1551-6709 online

DOI: 10.1111/j.1551-6709.2011.01174.x

Rules Versus Statistics: Insights From a Highly Inflected Language

Jelena Mirković,^a Mark S. Seidenberg,^b Marc F. Joanisse^c

^a*Department of Psychology, University of York*

^b*Department of Psychology, University of Wisconsin – Madison*

^c*Department of Psychology, The University of Western Ontario*

Received 23 October 2007; received in revised form 29 September 2010; accepted 10 October 2010

Abstract

Inflectional morphology has been taken as a paradigmatic example of rule-governed grammatical knowledge (Pinker, 1999). The plausibility of this claim may be related to the fact that it is mainly based on studies of English, which has a very simple inflectional system. We examined the representation of inflectional morphology in Serbian, which encodes number, gender, and case for nouns. Linguists standardly characterize this system as a complex set of rules, with disagreements about their exact form. We present analyses of a large corpus of nouns which showed that, as in English, Serbian inflectional morphology is quasiregular: It exhibits numerous partial regularities creating neighborhoods that vary in size and consistency. We then asked whether a simple connectionist network could encode this statistical information in a manner that also supported generalization. A network trained on 3,244 Serbian nouns learned to produce correctly inflected phonological forms from a specification of a word's lemma, gender, number, and case, and generalized to untrained cases. The model's performance was sensitive to variables that also influence human performance, including surface and lemma frequency. It was also influenced by inflectional neighborhood size, a novel measure of the consistency of meaning to form mapping. A word-naming experiment with native Serbian speakers showed that this measure also affects human performance. The results suggest that, as in English, generating correctly inflected forms involves satisfying a small number of simultaneous probabilistic constraints relating form and meaning. Thus, common computational mechanisms may govern the representation and use of inflectional information across typologically diverse languages.

Keywords: Morphology; Inflection; Connectionism; Rules; Statistical learning; Cross-linguistic

A long-standing debate in cognitive science concerns how to best characterize the knowledge underlying language acquisition and use. Grammar-based theories emphasize the role of rules and other abstract, symbolic representations (e.g., Pinker, 1999). Learning a

Correspondence should be sent to Jelena Mirković, Department of Psychology, University of York, York YO10 5DD, UK. E-mail: j.mirkovic@psych.york.ac.uk

language is equated with learning its grammar. Acquisition is possible only because humans are born with knowledge of linguistic universals (Universal Grammar), which constrain the child to entertain only some hypotheses about linguistic input (e.g., Crain, 1991). In contrast, the connectionist or parallel distributed processing approach views knowledge as represented by networks of simple neuron-like processing units. Language acquisition involves developing networks that support comprehension and production, guided by rich linguistic input and subject to general constraints on learning and knowledge representation. Thus, the grammar and connectionist approaches represent very different theories about the structure of language, how it is acquired and used, and how it relates to other human capacities (McClelland & Patterson, 2002; Seidenberg, 1997).

Much of the debate over these alternative theories has focused on investigations of the English past tense (e.g., Kielar & Joanisse, 2010; Kielar, Joanisse, & Hare, 2008; MacWhinney & Leinbach, 1991; McClelland & Patterson, 2002; Pinker, 1991; Pinker & Prince, 1988; Pinker & Ullman, 2002; Plunkett & Marchman, 1993; Rumelhart & McClelland, 1986), which has been offered as a prototypical example of linguistic knowledge that instantiates essential characteristics of language (Pinker, 1991, 1999). In Pinker and colleague's theory, two mechanisms are used for processing past tense forms: a rule-governed mechanism for generating regular past tenses, such as *walk-walked*, and a neurally and cognitively separate system for memorizing irregular past tenses, such as *swim-swam*. Numerous phenomena concerning normal and disordered behavior and their brain bases have been brought to bear on this distinction (Pinker & Ullman, 2002). For example, the associative memory system hypothesized to underlie the processing of irregular forms is sensitive to factors such as phonological similarity and frequency, whereas rule-governed forms are not (Pinker, 1991). Similarly, brain-injured patients are said to exhibit impairments indicating a double dissociation between rule and exception mechanisms (e.g., Pinker & Ullman, 2002).

A contrasting view holds that words are represented and processed within a single lexical system encoding multiple probabilistic constraints (Seidenberg & MacDonald, 1999; Seidenberg & McClelland, 1989). These constraints mainly derive from statistical relations that hold among words' semantic, phonological, and orthographic forms, and the contexts in which words occur. The constraints are represented by the weights on connections between units in a lexical network that performs tasks such as computing the meaning or sound of a word from print (e.g., Harm & Seidenberg, 2004) or generating the past tense of a verb, as in the classic *wug* test (e.g., Joanisse & Seidenberg, 1999). The same network is used in processing all words; thus, such models do not make the categorical distinction between rule-governed forms and exceptions that is central to the dual-mechanism theory. This property is consistent with the fact that there is considerable overlap between "regular" and "irregular" forms. For example, *kept*, the "irregular" past tense of *keep*, contains the final /t/ phoneme, which is the regular inflection in *walked* (see McClelland & Patterson, 2002, for discussion and other examples). Whereas the standard theory assumes a strong distinction between rule-governed forms and exceptions, the statistical theory treats these items as falling on a continuum reflecting degrees of consistency. This approach is sometimes termed a "single-mechanism" model in contrast to the "dual-mechanism," rules-and-exceptions theory. There is a single mechanism in the sense that the same network is used in processing

all words. However, the network encodes several types of information (semantic, phonological, and so on) which jointly determine a word's well-formedness.

The Rumelhart and McClelland (1986) past tense model, which only included phonological information, was followed by other models (e.g., Daugherty & Seidenberg, 1994; Joanisse & Seidenberg, 1999; MacWhinney & Leinbach, 1991; Plunkett & Juola, 1999; Plunkett & Marchman, 1993, and many others) that addressed concerns raised by Pinker and Prince (1988) and additional empirical phenomena. These models yielded two important findings. First, generating rule-governed forms and exceptions does not require two mechanisms; the same set of units and connections can generate both accurately. Second, the models support an alternative account of generalization: Whereas the ability to generalize (e.g., *wug-wugs*) has been taken as the classic evidence for the existence of rules (Berko, 1958; Pinker, 1991), in connectionist networks knowledge that is encoded on the basis of exposure to words can be used to process novel inputs. These points have been demonstrated with respect to both the past tense (e.g., Plunkett & Juola, 1999), and phenomena concerning the pronunciation of written words in English (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996) and Chinese (Yang, McCandliss, Shu, & Zevin, 2009). The phenomena are related because both systems are quasiregular, that is, containing rule-like patterns while at the same time admitting numerous partial regularities (Seidenberg & McClelland, 1989).

Although considerable attention has focused on these accounts of the past tense, other theories of morphological processing should also be acknowledged. Earlier research focused on whether morphologically complex words are decomposed into constituents during word recognition (e.g., Baayen, Dijkstra, & Schreuder, 1997; Caramazza, Laudanna, & Romani, 1988; Colé, Beauvillain, & Segui, 1989; Marslen-Wilson, Tyler, Waksler, & Older, 1994), and whether the way words are processed depends on typological differences between languages (e.g., Bentin & Frost, 1995; Boudelaa & Marslen-Wilson, 2001; Lukatela, Gligoričević, Kostić, & Turvey, 1980; Meunier & Marslen-Wilson, 2004). Behavioral phenomena such as the effect of lemma frequency (the summed frequency of all inflected forms of a word, e.g., the frequency of *car* + the frequency of *cars*) on lexical decision latencies have been taken as evidence for a distinct lemma level of morphological representation in the lexicon (e.g., Baayen et al., 1997; Colé et al., 1989; Taft, 1979) as have various priming effects (e.g., Frost, Forster, & Deutsch, 1997; Laudanna, Badecker, & Caramazza, 1992; Marslen-Wilson et al., 1994). Most of these researchers assume that some words are represented in terms of constituent morphemes and some as whole words. How a word is represented is thought to depend on factors such as semantic transparency (e.g., Marslen-Wilson et al., 1994), type of suffixation (Colé et al., 1989), or whether the words are known or novel (e.g., Caramazza et al., 1988).

Many of the same phenomena have also been considered from a connectionist perspective (e.g., Davis, Casteren, & Marslen-Wilson, 2003; Gonnerman, Seidenberg, & Andersen, 2007; Plaut & Gonnerman, 2000; Rueckl, Mikolinski, Raveh, Miner, & Mars, 1997; Seidenberg & Gonnerman, 2000). Morphological structures (and associated behavioral phenomena) can be seen as arising within a system that learns to map between form (phonology, orthography) and meaning (semantics), across one or more sets of weighted connections. These systems do not incorporate processing mechanisms such as lexical decomposition or

affix stripping that are specific to morphologically complex words. Such systems acquire the correlations that hold between form and meaning through experience with a large corpus of words. Classical morphemes (e.g., *walk-*, *-ed*) are an example of a consistent mapping between form (the *-ed* ending) and meaning (roughly, “in the past”) across many words. The strength of such correlations can vary significantly across words. For example, the *-mit* morphemes discussed by Aronoff (1976) (e.g., *admit*, *permit*, *submit*) behave in many ways like classical morphemes but lack a strong semantic component. The *short* in *shortage* or *dress* in *dresser* provide partial cues to meaning, but the words’ meanings are not a simple compositional function of the parts. Thus, there are degrees of consistency in mappings between form and meaning just as there are between present and past tense. The approach predicts that morphological effects in behavioral studies should be graded as observed in several studies (Feldman, 2000; Feldman & Prostko, 2002; Feldman, Soltano, Pastizzo, & Francis, 2004; Gonnerman et al., 2007; Seidenberg & Gonnerman, 2000). Such graded effects are harder to account for in models that assume that morphemes are discrete units manipulated by word-formation rules, as, for example, in compound formation (Haskell, MacDonald, & Seidenberg, 2003; Kim, Marcus, Pinker, Hollander, & Coppola, 1994).

1. A more complex case

The problem with the English past tense is that it is too simple to discriminate between theories. English inflectional morphology exhibits little of the complexity seen in other inflectional systems and in other aspects of language. Whereas the rules-and-exceptions approach may be sufficient (or approximately so) to account for the English system, a stronger test of the theory would be provided by determining whether it applies as well to a more complex inflectional system such as, for instance, verbs in Romance languages or nouns in Slavic. Whereas English verb inflection is limited to tense and number, many languages also encode gender, aspect, mood, voice, and other types of information inflectionally. Such systems present challenges for both rule-based and connectionist theories. The challenge for the rule-based approach is merely to describe such systems, which turns out to be difficult. For example, standard pedagogical grammars of a Slavic language, such as Serbian¹ (e.g., Mrazović & Vukadinović, 1991; Stevanović, 1986), characterize the system by a set of complex rules, with multiple conditions attached (discussed below). This stretches the rules versus exceptions principle far beyond what was sufficient for the much simpler English case, it raises questions about the distinction between “rule-governed” and “exception,” and it invites inquiry as to whether other formalisms might provide a better account.

At the same time, it is not obvious whether connectionist networks working with correlations between different types of information can encode a system of this complexity, with many regularities, subregularities, and exceptions, and intricate dependencies between different types of information. Equally important, would such models be informative about the nature of the system, for example, about the difficulty of learning different parts of the system, or about factors that influence the use of this knowledge? One concern about connectionist models is that they are difficult to analyze and so may produce correct output

without providing much insight about the underlying mechanisms (McCloskey, 1991, but also see Seidenberg, 1993; Seidenberg & Plaut, 2006).

In this research, we used a connectionist network to explore properties of Serbian noun morphology. The main goal was to examine whether principles that have been identified in developing connectionist models of English inflectional morphology would extend to this more complex system. It is possible that people who use typologically distinct languages have developed fundamentally different ways of representing and processing them. However, it is also possible, as generative linguists have long emphasized, that seemingly diverse languages may share deeper properties that reflect commonalities in how they are learned, processed, and represented in the brain. If a model based on the principles employed in previous research on English extended gracefully to the more complex system, it would suggest that there are commonalities between them which can be captured within this theoretical approach. It was also possible that the model might fail to capture basic facts about the system, either by failing to learn the system or by representing it in a way that clearly deviates from native speakers' knowledge. Either outcome would yield a broader view of the nature of inflectional systems.

1.1. Basic properties of Serbian noun morphology

Serbian is a south Slavic language with a rich and complex inflectional morphology. Verbs, for example, are marked for tense, person, number and, in some forms, gender. Serbian nouns are marked for one of seven cases, one of three genders, as well as one of two numbers. Case markings denote a noun's sentential function, allowing considerable word order freedom in sentences. For example, in the sentence:

Hana	Mariji	kutlačom	sipa	vruću	supu.
Hannah	to Maria	with a ladle	serves	hot	soup.

Hannah serves the hot soup to Maria with a ladle.

each noun is marked for case, such that the nominative case marking in *Hana* denotes the subject of the sentence, dative in *Mariji* denotes the recipient, accusative in *vruću supu* the object, and instrumental in *kutlačom* the instrument. (Note that the verb, *sipa*, is marked for number [singular], person [third], and tense [present].) As a result of case markings, the words in the sentence could be ordered in several other ways and still convey the same meaning (e.g., *Mariji Hana sipa vruću supu kutlačom*. *Kutlačom Hana Mariji sipa vruću supu*. *Vruću supu Mariji Hana sipa kutlačom*, to name just a few). Serbian also has extensive agreement: This includes number agreement between the verb and the subject of the sentence (as well as gender agreement in past participle), and in the case of noun phrases, case, gender, and number agreement between the modifier and the noun (as in *vruću supu* above, where *supu* (soup) is a feminine noun in accusative singular, and thus the adjective *vruću* (hot) is also marked as feminine, accusative singular). The extensive use of inflections to mark agreement is typical of Slavic languages.

The regularities governing this system are complex and difficult to describe succinctly. Traditional grammars on this topic (e.g., Mrazović & Vukadinović, 1991; Stevanović, 1986) run to hundreds of pages. There are often several-way contingencies that determine the overt form of a word, making it difficult to describe the system in terms of simple rules and exceptions. For example, the vocative singular form of masculine nouns will have the suffix *-e* unless the stem ends in /j/, /z/, /tj/, /tɕ/, /dʒ/, /dz/, /j/, /ʎ/, or /ɲ/ when it will have *-u*; some of these nouns will also have variants ending in *-e*, as will some nouns with stems ending in /ar/ or /ir/. Also, some nouns with stems ending in /ts/ or /z/ will take the suffix *-u* (e.g., *Englez* (Englishman) and *Francuz* (Frenchman)). Some personal Muslim names with the stem ending in a velar consonant will take *-u* (e.g., *Abdulah*, *Alah*, *Refik*) (examples from Ivić, 1990). There is nothing to prohibit the creation of complex rules to describe data of this sort and indeed grammarians have tried. However, the system differs greatly from the characterization of grammatical rules that emerged from studies of the English past tense. First, the rules do not have the simplicity and transparency of the past tense rule, which provided some of its intuitive appeal. Second, again unlike English, there is disagreement about the proper statement of the rules for Serbian. Third, the Serbian rules make reference to many properties of words, including phonology (e.g., final stem phoneme, number of syllables) and semantics (e.g., ethnicity of the referent). However, the strong claim about grammatical rules based on English was that they are blind to semantics and phonology (Pinker, 1991, p. 531). Thus, Serbian can only be described as “rule-governed” but not under Pinker’s conception of a “rule.” Finally, descriptive grammars typically do not include information about quantitative aspects of the system, such as the frequencies with which different properties occur and co-occur. Although this practice is consistent with a “competence” approach, in which such usage information is excluded, it is highly relevant to statistical learning procedures of the sort thought to be used by infants (e.g., Gerken, 2002; Gerken, Wilson, & Lewis, 2005; Saffran, Aslin, & Newport, 1996) and generally in language processing (e.g., Dell, Reed, Adams, & Meyer, 2000; MacDonald, 1999; Trueswell, 1996).

The first part of this study involved an analysis of a large subset of Serbian nouns to obtain basic information about the structure of the system. The analyses of large corpora make it possible to identify properties of complex systems that would not otherwise be apparent. These corpus analyses show that Serbian inflectional morphology is indeed quasi-regular: It exhibits numerous regularities and subregularities, as well as seeming “exceptions.” This first step toward a more psycholinguistically relevant characterization of the system also yielded information about sources of complexity within it, and about the degrees of consistency and inconsistency in different regions of it.

In the second part of the study, we examined the Serbian system using a connectionist model employing statistical learning and processing via constraint satisfaction. A basic characteristic of these models is that the output that is produced in performing a task (e.g., producing a word) is that which best satisfies multiple probabilistic constraints picked up through learning. Such a model could be well-suited to encoding the myriad regularities and subregularities that pervade the system. We therefore sought to determine whether a simple network could learn a system of this sort and represent the contingencies among different types of information in a way that would support generalization.

The model also functioned as a discovery procedure: Given the complexity of the system and the sheer number of interacting factors, a working network might be required to discover many of its properties. In fact, the model led us to two main discoveries. One is about an overlooked source of complexity in the system, the fact that stems (roots) undergo morphologically conditioned phonological alternations. Learning these alternations turns out to be more difficult for the model than learning the inflections themselves. Second, we identified a novel statistical property of Serbian that was strongly related to the model's performance. This model-derived factor was then found to affect performance in a behavioral experiment with Serbian speakers.

The results of the corpus analysis, modeling, and the experiment suggest that Serbian inflectional morphology is a quasiregular system with many probabilistic constraints that jointly determine the form of a word. Taken with previous research suggesting that English inflectional morphology can also be characterized in this way (e.g., Joannis & Seidenberg, 1999; McClelland & Patterson, 2002), our results suggest that inflectional systems differ in complexity but are learned, represented, and processed in the same way.

2. Corpus analysis of Serbian nouns

Serbian nouns are marked for three morphosyntactic parameters: gender, number, and case. There are three genders (masculine, feminine, and neuter), two numbers (singular and plural), and seven cases (nominative, accusative, genitive, dative, instrumental, locative, and vocative). Because Serbian is a fusional language, several of these parameters are coded within a single suffix. As the examples in Tables 1 and 2 illustrate, the information about case and number is conveyed by the suffix. In contrast, gender information is conveyed partially through the distribution of suffixes (such that different genders have different suffixes in different cases) and partially through other phonological (e.g., stem-final phoneme, syllabic structure; see Mirković, Seidenberg, & MacDonald, 2008, for more details) and semantic properties (Mirković, MacDonald, & Seidenberg, 2005; Mirković et al., 2008). Importantly, however, knowing the case, number, and gender of a noun is not sufficient to

Table 1
Singular forms of nouns *medved* (bear), *krava* (cow), and *selo* (village)

Case	Number	Masculine	Feminine	Neuter
NOM.	Singular	medved	krav <u>a</u>	sel <u>o</u>
GEN.	Singular	medved <u>a</u>	krav <u>e</u>	sel <u>a</u>
DAT.	Singular	medved <u>u</u>	krav <u>i</u>	sel <u>u</u>
ACC.	Singular	medved <u>a</u>	krav <u>u</u>	sel <u>o</u>
INST.	Singular	medved <u>om</u>	krav <u>om</u>	sel <u>om</u>
LOC.	Singular	medved <u>u</u>	krav <u>i</u>	sel <u>u</u>
VOC.	Singular	medved <u>e</u>	krav <u>o</u>	sel <u>o</u>

Notes. All examples in Serbian are in International Phonetic Alphabet, unless italicized. See note 2 for abbreviations.

Table 2

Plural forms of nouns *medved* (bear), *krava* (cow), and *selo* (village)

Case	Number	Masculine	Feminine	Neuter
NOM.	Plural	medved <u>i</u>	krav <u>e</u>	sela <u>a</u>
GEN.	Plural	medved <u>a</u>	krav <u>a</u>	sela <u>a</u>
DAT.	Plural	medved <u>ima</u>	krav <u>ama</u>	sela <u>ima</u>
ACC.	Plural	medved <u>e</u>	krav <u>e</u>	sela <u>a</u>
INST.	Plural	medved <u>ima</u>	krav <u>ama</u>	sela <u>ima</u>
LOC.	Plural	medved <u>ima</u>	krav <u>ama</u>	sela <u>ima</u>
VOC.	Plural	medved <u>i</u>	krav <u>e</u>	sela <u>a</u>

Note. See note 2 for abbreviations.

Table 3

Singular forms of feminine nouns *duša* (soul) and *strast* (passion), which use different sets of suffixes across inflectional forms

Case	NOM.SG. in /a/	NOM.SG. in consonant
NOM.	du <u>ša</u>	strast
GEN.	du <u>še</u>	strast <u>i</u>
DAT.	du <u>ši</u>	strast <u>i</u>
ACC.	du <u>šu</u>	strast
VOC.	du <u>šo</u>	strast <u>i</u>
INST.	du/ <u>šom</u>	strast/ <u>cu</u>
LOC.	du/ <u>ši</u>	strast <u>i</u>

determine its correct phonological form; other factors modulate how this information is realized. For example, the accusative singular form of masculine nouns depends on the animacy of the noun (e.g., /delfin/(dolphin)-NOM.SG.², /delfina/-ACC.SG., whereas /klavir/(piano)-NOM.SG., /klavir/-ACC.SG.). In another example, if a feminine noun ends in a consonant in the nominative singular (e.g., /strast/(passion)-NOM.SG.), then its inflectional forms will be different than the ones for feminine nouns ending in /-a/ (Table 3). These examples show that the way case and number information is conveyed depends on both semantic (e.g., animacy) and phonological properties of nouns.

We conducted a corpus analysis to explore this system in more detail. Nouns in Serbian and other highly inflected languages are usually described in terms of declensions or inflectional paradigms (e.g., Comrie & Corbett, 1993). Declensions categorize nouns based on morphophonological properties, for example, the final phoneme of the stem and the set of suffixes they take (e.g., in Latin, the stem of the nouns in the first declension ends in *-a*, in the second declension in *-o*, etc, and each declension has a standard set of suffixes). However, because of the complexity of the system and numerous exceptions (actually subregularities determined by semantic and/or phonological factors), descriptive grammarians have not been able to reach a consensus about the number of declensions or the criteria for distinguishing them (e.g., Ivić, 1990; Stanić, 1949). The general problem of the adequate description of inflectional paradigms has been a widely investigated topic in linguistics (see, e.g., Carstairs-McCarthy, 1998, for discussion).

Following Maretić (1963), we based the corpus analysis on gender rather than declension. Aside from the disagreements about the identification of declensions, the major reason for focusing on gender is its important role in language use (comprehension and production): there is agreement in gender (as well as number and case) between, for example, a modifier and a noun (e.g., /lepī dečak/-MASC. (beautiful boy) vs. /lepa devojčica/-FEM. (beautiful girl)). Thus, a feature like gender, which is relevant for agreement, rather than declension, seemed more relevant to psycholinguistic issues. We note that some of the regularities described below overlap with some views of declensions in Serbian (e.g., Stanojčić, Popović, & Micić, 1989). Ultimately, however, given the nature of this inflectional system, using either gender or declension would yield similar patterns, that is, many inconsistencies and subregularities (see Ivić, 1990, and Milin, Filipović Đurđević, & Moscoso del Prado Martín, 2009, for a discussion of the relationship between declensions and paradigms).

In each gender, nouns were described in terms of two main factors. First, the genders differ with respect to the number of different suffixes inflectional forms can take. For example, in feminine nouns, there are two possible sets of suffixes and which one a noun will take depends on the final phoneme of its nominative singular form (Table 3).

The second factor is phonological properties of the stem (the word without its inflectional endings), for example, properties of the final phoneme or syllabic structure. Phonological properties of the stem interact with inflectional suffixes (as in English: *knife/knives* vs. *car/cars*) and so potentially create subregularities (e.g., *knives*, *leaves*).

2.1. Method

We analyzed a corpus of 300 nouns (2,648 inflected forms³) randomly drawn from the Frequency Dictionary of Contemporary Serbian (Kostić, 1999), which is itself derived from a corpus of about 2 million words (Kostić, 2001). This sample was also used as a part of the training corpus for the model presented below. Words longer than six syllables (approximately 0.5% of nouns in the 2 million corpus) and words with syllable onsets or codas containing more than two consonants (1.2% of nouns in the 2 million corpus) were excluded because they did not fit the syllabic template used in the modeling.

The analysis involved identifying various regularities and subregularities within the noun system. Grammar books provided some hints as to where to look, and the intuitions of a native speaker (J.M.) provided others. We were especially interested in characterizing statistical properties of the system. Where a particular pattern is noted, how often does it occur? How much of the system can be described in terms of recurring patterns? Are there true exceptions in the sense of forms that are not predictable from any factors?

2.2. Results

For the nouns in this sample, 46.41% were masculine, 38.75% feminine, and 14.84% neuter. These proportions are consistent with the ones obtained from the larger 2 million word corpus collected by Đ. Kostić (A. Kostić, personal communication): masculine: 44.99%;

feminine: 40.43%; and neuter: 14.57%. Detailed analyses of specific properties of nouns are presented below for each gender separately.

2.2.1. Masculine

The suffixes for the masculine nouns partition into two major types: one set is used with nouns ending in a consonant in the nominative singular (99.28% of masculine nouns; example 1(a) below, and Tables 1 and 2) and the other is used with nouns ending in /a/ in the nominative singular (0.72% of masculine nouns; 1(b) below). Interestingly, the nouns ending in /a/ actually pattern with (i.e., have the same set of suffixes as) the majority of feminine nouns (see the example in Tables 1 and 2). This type of cross-regularity contributes to the complexity of the system.

1. (a) medved (bear)-NOM.SG, medveda-GEN.SG., medvedu-DAT.SG., etc.
- (b) vodza (leader)-NOM.SG, vodze-GEN.SG, vodzi-DAT.SG, etc.

An important property of the masculine nouns of the 1(a) type is that if the noun is animate, its accusative singular form is the same as its genitive singular form (2(a) below); this holds regardless of phonological properties of the stem. Conversely, if the noun is inanimate, its accusative form is the same as its nominative singular form (2(b)). There are two partial regularities here. First, animacy is only predictive of the relationship between accusative and genitive or nominative singular forms; second, animacy is only relevant for masculine nouns.

2. (a) medved (bear)-NOM.SG, medveda-ACC.SG, medveda-GEN.SG.
- (b) klavir (piano)-NOM.SG, klavir-ACC.SG, klavira-GEN.SG.

The nouns of the 1(a) type partition into several subgroups resulting from the behavior of the stem to which the suffix is attached.

Simple forms. The nouns in this group (48.55% of masculine nouns) have a “simple” inflectional paradigm: All inflectional forms share the same stem (see Tables 1 and 2) and the only thing that changes is the suffix. These are items like *books* and *walked* in English, in which the stem is preserved in the suffixed form.

Extended plural forms. In the next largest group (20.29% of masculine nouns), which mainly consists of nouns with monosyllabic nominative singular forms, all singular case forms share the same stem (3(a)). All plural forms, however, have the same stem as singular forms, but in addition they have an infix preceding the inflectional suffix (3(b)). The form of the infix (/ov/ or /ev/) depends on the final consonant of the stem.

3. (a) broja (number)-GEN.SG, broju-DAT.SG.; plana (plan)-GEN.SG., planu-DAT.SG.
- (b) brojeva-GEN.PL, brojevima-DAT.PL.; planova-GEN.PL, planovima-DAT.PL.

Palatalized forms. In this group (11.59% of masculine nouns), the nouns end in a velar stop or fricative (/k/, /g/, /x/) in the nominative singular (4(a)); in inflectional forms ending

in a front vowel (/i/ or /e/) palatalization occurs, so that with the suffix /i/ the final phonemes /k/, /g/, /x/ change to /ts/, /z/, /s/, respectively, and with the suffix /e/ to /tʃ/, /ʒ/, /ʃ/, respectively (4(b)). However, the accusative plural is an exception: Even though its ending is /e/ no change occurs (4(c)).

4. (a) savetnik (advisor)-NOM.SG.
- (b) savetnitsi-NOM.PL., savetnitʃe-VOC.SG.
- (c) savetnike-ACC.PL.

Moving ‘a.’ This group (6.52% of masculine nouns) consists of nouns that have a “moving /a/” (Browne, 1993) in the nominative singular and genitive plural forms (5(a)). (In many cases, this is a consequence of diachronic change; e.g., Stevanović, 1986.) There is also a group of nouns (10.14% of masculine nouns) that in addition to the moving “a” exhibit some other phonological alternation, for example, voicing assimilation (5(b); 0.72% of masculine nouns), palatalization (5(c); 5.8% of masculine nouns), infixing in all plural forms (5(d); 2.9% of masculine nouns), or /l/ to /o/ alternation (see below) in addition to infixing in all plural forms (5(e); 0.72% of masculine nouns).

5. (a) lovtsa (hunter)-GEN.SG., lovats-NOM.SG, lovtsa-GEN.PL.
- (b) vrbats (sparrow)-NOM.SG, vrapsa-GEN.SG.
- (c) opstanak (survival)-NOM.SG, opstanski-NOM.PL.
- (d) ovan (ram)-NOM.SG, ovnovi-NOM.PL.
- (e) ugao (corner)-NOM.SG., uglovi-NOM.PL.

/l/ to /o/ alternation. In a small group of masculine nouns (2.17%), if /l/ is the final phoneme in the stem in some case forms it changes to /o/ (6).

6. pepeo (ash)-NOM.SG., pepela-GEN.SG., pepelu-DAT.SG.

Finally, in all masculine nouns ending in a consonant in the nominative singular, the instrumental singular form will depend on the final consonant of the stem: If it is an alveolar or palato-alveolar consonant the suffix will be /em/ (7); in all other cases it will be /om/.

7. broj (number)-NOM.SG., brojem-INST.SG.

As noted above, masculine nouns ending in /a/ in the nominative singular (example 1(b)) have the same set of suffixes as the majority of feminine nouns. Interestingly, many of these nouns represent various professions traditionally performed by men (e.g., *sudija* (judge), *vođa* (leader)), and they are syntactically masculine in singular forms, for example, they agree with adjectives in gender (e.g., *pravedni-MASC. sudija-MASC.* (a fair judge)). This illustrates that even when some information makes a part of the system particularly complex (such as the masculine nouns with typically feminine phonological forms), other

information (semantic, distributional) may actually provide constraints to grammatical properties of the word.

In summary, in terms of the number of suffix sets masculine nouns partition into two groups (1(a) and (b)), the ones with the nominative singular ending in a consonant (1(a)) being the overwhelming majority. In terms of morphophonological properties of the stem, we identified five major patterns but some with further subregularities. This suggests that the complexities of the system may lie more in stem properties than in inflections.

2.2.2. Feminine

In feminine nouns, there are also two sets of suffixes: one used with the nouns ending in /a/ in the nominative singular (90.35% of feminine nouns, see examples in Tables 1–3), the other with the nouns ending in a consonant (9.65% of feminine nouns, Table 3).

Among the feminine nouns ending in /a/ in the nominative singular, there are three subgroups which result from different morphophonological properties of the stem.

Simple forms. The nouns in this group (71.93% of feminine nouns) have the same stem in all inflectional forms and only the suffix changes (see Tables 1 and 2).

Consonant clusters. The stem of the nouns in this group (13.16% of feminine nouns) ends in a consonant cluster. Some of these nouns have a moving “a” in the genitive plural (8(a)), whereas others use the same form for the genitive plural and the dative and locative singular forms (8(b)).

8. (a) zem λ a (country)-NOM.SG., zema λ a-GEN.PL.
- (b) banda (gang)-NOM.SG., bandi-GEN.PL./DAT/LOC.SG.

Palatalized forms. In this group (5.26% of feminine nouns), the final phoneme of the stem is a velar stop or fricative (/k/, /g/, or /x/) which is palatalized if followed by the suffix /i/ (9).

9. fabrika (factory)-NOM.SG, fabritsi-DAT/LOC.SG.

Almost 10% (9.65% to be precise) of feminine nouns use a different set of suffixes than the three groups described above (Table 3). Many of these are abstract nouns (e.g., *strast* (passion), *nadmocnost* (superiority)). Most inflectional forms of the nouns in this group are the same—the majority ends with the suffix /i/ with no change in the stem (see examples in Table 3). However, these nouns are the only ones within the noun inflectional system that exhibit jotation, a phonological alternation involving softening/palatalization of the consonants preceding the glide /j/ (e.g., /d/ before /j/ becomes /dz/). This kind of alternation is relatively frequent in adjective and verb paradigms, but within the noun inflectional system it occurs only in this group of feminine nouns and is complex: It involves changing two phonemes, /st/ into /ft ζ / (10), and it happens only in one case, instrumental singular.

10. strast (passion)-NOM.SG, straft ζ cu-INST.SG.

In summary, like masculine nouns, feminine nouns partition into two groups based on the suffixes they take. However, in terms of morphophonological properties of the stem, feminine nouns are less complex than masculine as there are only three different patterns with fewer subregularities within them.

2.2.3. Neuter

Neuter nouns are characterized by only one set of suffixes. However, on the basis of phonological characteristics of the stem, we identified two groups. The nouns in one group (95.83% of neuter nouns) end in /o/ or /e/ in the nominative singular (see Tables 1 and 2), and a subgroup (12.5% of neuter nouns) of these have the moving “a” in the genitive plural (11(a)). The second group (4.17% of neuter nouns) consists of nouns that have an extended base in all inflectional forms except the nominative, accusative, and vocative singular: An affix (/n-/ or /t-/) is inserted between the stem and the inflectional suffix (11(b)).

11. (a) veslo (oar)-NOM.SG., vesala-GEN.PL.
 (b) ime (name)-NOM.SG., imena-NOM.PL.

In short, neuter nouns are the least complex of the three genders, as there is only one set of suffixes they use, and there are fewer patterns regarding morphophonological properties of the stem.

2.3. Discussion

The corpus analysis identified commonalities and differences across nouns in the three genders (see Table 4 for a summary). The principal similarity is that in all genders the complexities seem to lie more within stem alternations across inflectional forms than within suffixes. Further, all genders exhibit considerable complexity involving phonological and semantic factors. Some phonological alternations (e.g., extended stems) are common to all genders (examples 3, 11(b)), and some are found in one gender but not others. For example, some masculine nouns will undergo several phonological alternations, such as the ones in the moving “a” group (examples 5(b)–(e)). Similarly, jotation is only observed in feminine nouns (10). Finally, different semantic properties arise as constraints in different genders: animacy in masculine nouns, and abstractness in feminine (see also Mirković et al., 2005, 2008, for a discussion on semantic regularities across genders).

These data indicate that among the nouns of the language it is difficult to draw a clear distinction between “regular” (rule-governed) and “irregular” forms. From the rule perspective, Serbian looks quite different from English: There is a huge leap in the complexity of the rules and in the range of properties they are conditioned on. From the statistical perspective, the two systems look more similar: They are both quasiregular (Seidenberg & McClelland, 1989). There are consistent, rule-like patterns, conditioned by phonological and semantic factors, and there are cases that deviate from these central patterns but not arbitrarily.

Table 4
Summary of morphophonological patterns with type counts

Gender	NOM. SG. ending ^a	Stem Properties	Example	Inflectional Neighborhood Size ^b
Masculine	Consonant	Simple forms	medved	48.55
		Extended plural	broj	20.29
		Palatalized forms	savetnik	11.59
		Moving ‘‘a’’		
		Only	lovats	6.52
		+ Palatalization	opstanak	5.80
		+ Extended plural	ovan	2.90
		+ Voicing assimilation	vrabats	0.72
		+ /l/ to /o/ and extended plural	ugao	0.72
		/l/ to /o/	pepeo	2.17
Feminine	/-a/			
	/-a/	Simple forms	krava	71.93
		Consonant clusters		
		GEN.PL.=DAT/LOC.SG.	banda	10.53
		Moving ‘‘a’’	zemlja	2.63
		Palatalized forms	fabrika	5.26
Neuter	Consonant		strast	9.65
	/-o/, /-e/	Simple forms	selo	83.33
		Moving ‘‘a’’	veslo	12.50
		Extended forms	ime	4.17

^aIt determines the set of suffixes the noun takes.

^bPercentage of lemmas within each gender.

3. Computational model

In this section, we describe a computational model we used to explore Serbian noun inflection. Two models of Serbian noun morphology were proposed in earlier research. The Satellite Entries Hypothesis of Lukatela et al. (1980) suggests that nouns are represented in a satellite-like fashion, such that the nominative form is the ‘‘access’’ form and other case forms are at equal distance from the nominative form. The basis for this model was a set of experiments on Serbian nouns that showed no effect of frequency in cases other than the nominative (Lukatela et al., 1980). However, later studies (Kostić, 1991, 1995; Todorović, 1988) failed to replicate these effects. Based on these findings, Kostić (1991, 1995) proposed the Informational Approach, which captures facts about the noun system in terms of information load, calculated from the frequency of noun forms and their syntactic complexity (measured in terms of the number of functions and meanings cases can have). This measure was found to correlate with lexical decision latencies (Kostić, 1991, 1995). We think this approach captures some important factors that contribute to word complexity. However, we were more interested in characterizing the system in terms of the factors encoded in the inflections (gender, number, and case), which capture aspects of language relevant for language use, such as agreement, determining the number of referents, and sentential function. Some of these factors are closely related to the Kostić informational load measure. For

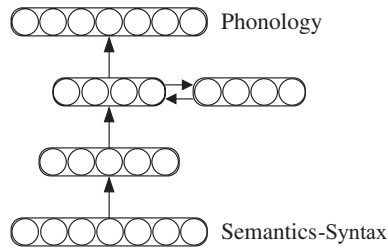


Fig. 1. Architecture of the Serbian morphology model. Ellipses represent groups of units organized into layers. Arrows indicate directional weighted connections used to pass information among layers. The input layer is connected to the phonological layer via two hidden layers consisting of 250 and 100 units, respectively. The second hidden layer, which directly feeds into the output layer, is connected in both directions to a “recurrent” layer consisting of 100 units. The inclusion of this layer permitted the network to produce sequential outputs when presented with a static input.

example, “case” implicitly encodes information about the range of sentence structures a word can enter into, which affects informational load. Furthermore, we also wanted to investigate the extent to which the proper inflection of a form can be determined by semantic and phonological factors.

3.1. Model architecture

The network (Fig. 1) consisted of four layers of units that were connected using weighted connections. The input layer consisted of 419 binary units that coded a localist representation uniquely identifying each word in the corpus: 407 of the units were lemma nodes,⁴ 3 were gender nodes, 2 were number nodes, and 7 were case nodes. Each word was identified by activating (setting to 1.0) one input node from each group (lemma, gender, number, and case). For instance, the word /kravom/(cow)-FEM.INST.SG. was represented by activating the units corresponding to the *krava* (cow) lemma, the feminine gender, the singular number, and the instrumental case; values of all other units were set to zero. Similarly, the word /kravu/(cow)-ACC.SG. used the same lemma, gender, and number input values, but instead the accusative case unit was set to 1.0 and the instrumental case unit was set to zero. The model learned to map this input specification to the word’s overt phonological form.

The output layer represented word phonology by producing a sequence of syllables over time. The syllable representation was slot based and vowel centered, such that each syllable was fit to a CCVCC frame (C = consonant, V = vowel; the sonorants /r/, /l/ and /n/ were coded in the vowel slot when they occurred as vocalic, and in consonant slots otherwise; Stanojčić et al., 1989). Each phoneme was represented as a vector of 16 binary units where each unit represented a binary phonetic feature (Table 5). The phonological representation was based on the standard description of Serbo-Croatian phonemes (e.g., Corbett, 1987; Stanojčić et al., 1989). Empty slots were coded by setting all features in that slot to zero (e.g., the final consonants in the syllable /kra/).

In contrast to standard feed-forward networks that produce a static output representation, the present network contained recurrent connections within the second hidden layer which

Table 5
Phonological representation used in the network

	Cons.	Vocalic	Obstruent	Sonorant	Lateral	Contin.	Non-contin.	Voiced	Voiceless	Nasal	Labial	Coronal	Palatal	High	Distrib.	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	1	0	0	0	0	0
d	1	0	1	0	0	0	1	1	0	0	1	0	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	1	0	0	1	0	0	0	0	0
v	1	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0
ts	1	0	1	0	0	1	1	0	1	0	1	0	0	0	0	0
s	1	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0
z	1	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0
ʒ	1	0	1	0	0	1	0	1	0	0	1	0	0	0	1	0
ʃ	1	0	1	0	0	1	0	1	0	0	0	0	0	0	1	0
dʒ	1	0	1	0	0	1	1	0	1	0	1	0	0	0	1	0
dʒ	1	0	1	0	0	1	1	0	1	0	1	0	1	0	0	0
tʃ	1	0	1	0	0	1	1	0	1	0	1	0	0	1	0	0
tʃ	1	0	1	0	0	1	1	0	1	0	1	0	1	0	0	0
x	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	1
m	1	0	0	1	0	0	1	1	1	1	1	0	0	0	0	0
n	1	0	0	1	0	0	1	1	0	1	1	0	0	0	0	0
ɲ	1	0	0	1	0	0	1	1	0	1	1	0	1	0	0	0
l	1	0	0	1	1	1	0	1	0	0	1	0	0	1	0	0
ʎ	1	0	0	1	1	1	0	1	0	0	0	1	1	0	1	0
r	1	0	0	1	0	1	0	1	0	0	1	0	0	0	0	0
j	1	0	0	1	0	1	0	1	0	0	1	0	1	0	0	0
i	0	1	0	0	0	1	0	1	0	0	0	1	0	0	0	0
e	0	1	0	0	1	1	0	1	0	0	0	1	0	0	0	0
a	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0	1
o	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0	1
u	0	1	0	0	1	1	0	1	0	0	1	0	0	1	0	1

Note. Cons., consonantal; Contin., continuant; Non-contin., non-continuant; Distrib., distributed.

permitted it to generate a series of discrete outputs for a given static input. This approach permitted us to simulate the production of multisyllabic words of varying lengths within a relatively small output frame. This also had the benefit of allowing the network to use the same units and connections to produce words of many lengths.

The network simulated a simplified version of the task of producing speech: It was given information about an intended utterance as input and had to produce the corresponding phonological form as output. This architecture also seemed appropriate for exploring questions about learning a complex inflectional system. The model was not designed or intended to capture detailed aspects of children's performance in learning inflectional morphology; rather, it was used to identify properties of the system that might affect acquisition and skilled performance. The model was provided with knowledge of genders, cases, and other information represented on the input layer, which must also be learned by the child (see Mirković et al., 2005, regarding gender as an emergent category). In addition, a localist "lemma" system is used to represent a word's meaning, an implementational compromise rather than a theoretical claim. In reality, we assume individuals learn to map sounds and meanings in a way that allows lemma-like structures to emerge as an interlevel, mediating code (see Seidenberg & Gonnerman, 2000, for discussion). We employed this compromise because of the unavailability of reliable feature-based semantic representations for Serbian nouns (see, e.g., McRae, Cree, Seidenberg, & McNorgan, 2005), but it should not be taken as a claim that the system demands the use of this type of representation. This representation does not capture similarity relations among lemmas and other semantic properties, which can be captured using distributed semantic representations. Such representations seem highly relevant to properties of this inflectional system, for example, the role of a feature such as animacy in determining inflections on masculine nouns. In summary, localist representations were used on the input layer to allow us to explore other properties of the inflectional system, setting aside the issue of how lemmas, genders, and cases are themselves learned and represented.

3.2. Training

The network was trained on a corpus of 3,244 inflected forms derived from 407 nouns drawn from the Frequency Dictionary of Contemporary Serbian (Kostić, 1999). Of the 407 nouns, 300 were randomly drawn from the Serbian frequency dictionary resulting in a total of 2,648 inflectional forms (also used in the corpus analysis above). A set of additional 107 nouns (a total of 596 forms) was added to facilitate more detailed analyses of the network⁵. As each noun in Serbian can have between 1 and 14 inflectional forms (seven cases \times two numbers), in most cases the training items for any given noun were only a subset of these, reflecting the fact that some inflectional forms did not occur in the corpus of 2 million words used to generate the Frequency Dictionary, even though such forms are legal Serbian nouns. In addition to these "missing" items, we held back 90 items (one inflectional form for each of the 90 nouns) which we used for a generalization test described below.

Words in the training set were selected on the basis of the following criteria: The maximum length was six syllables and each syllable could not exceed the CCVCC frame. Word

frequencies were taken from the Serbian frequency dictionary, and then logarithmically transformed based on the formula

$$p = \frac{\log(f + 1)}{\log(f_{\max})},$$

where f represents the frequency of the word and f_{\max} the frequency of the most frequent word in the corpus (/broj/(number)-MASC.NOM.SG., with 1,426 occurrences in a corpus of 2 million words). This frequency compression scheme preserved the general statistical structure of Serbian, while assuring that the network was periodically exposed to low-frequency words.

Network training proceeded as follows: In each training trial, the network was presented with a randomly selected word on the input layer and was required to produce the correct phonological form at the output. For example, the word /krava/(cow)-FEM.NOM.SG. was presented to the input layer in the format described above and the network was required to produce the phonetic features of the phonemes /k/, /r/, /a/, /v/, and /a/ at the output, syllabified as /kra/ and then /va/. The network was trained such that the representation at the input layer was clamped (frozen to the set values) for the entirety of each training trial. Activation propagated from the input to the output, and each syllable was activated at the output layer in two time-step increments. For a six-syllable word, for example, the phonemes of the first syllable were activated for time steps 3 and 4, the second syllable was activated at time steps 5 and 6, the third at time steps 7 and 8, and so on. Each training trial concluded when the last syllable was output for two time steps. The network was trained using the backpropagation through time learning algorithm (Hinton, 1989), which compared target activation values in the output layer with the obtained activations for these units at each time step, and adjusted each connection weight in a way that gradually minimized differences between obtained and desired activation values for each unit. Activation was compared using a cross-entropy error measure (Hinton, 1989). Error tolerance was set to 0.1, learning rate was 0.005; the initial random weight range was ± 0.01 .

3.3. Results

3.3.1. Learning: Overview

Five models with the same architecture were trained as replications or “subjects.” The replications differed only with respect to the random initial weights and the sample of words presented during training. Data presented are averages across the five simulations. Items were scored using a nearest neighbor criterion based on Euclidean distance, such that all the phonemes had to be closer to the correct phoneme than to any other phoneme. Training was concluded when all words in the training set were identified correctly by this criterion (Fig. 2). On average, the training set was learned by 2.3 million iterations (range: 1.9–3 million). The learning curve has a characteristic sigmoid shape: slow initial learning, with a rapid increase such that 90% of items were learned within the first 800,000 iterations. The model then needed about the same amount of training to learn the remaining items. Thus, the task was learnable.

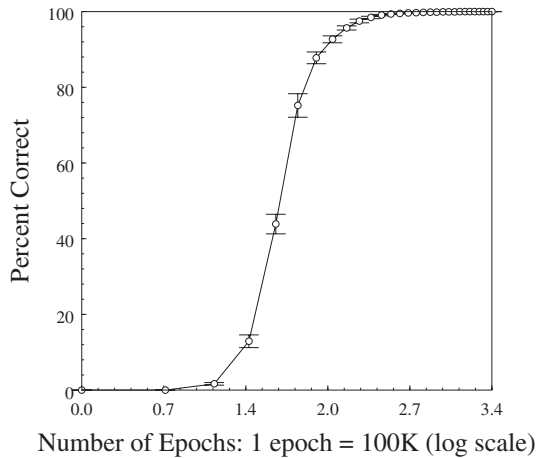


Fig. 2. Percentage of correctly produced items during learning (averaged across five runs).

3.3.2. Learning: Analyses of errors

We then examined which forms were the most difficult to learn and yielded poorest performance. This is relevant to determining how the model learned the training set and facilitated developing measures that could be related to human performance. For these analyses, we examined all errors that each model produced after approximately 600,000 iterations, at which point accuracy averaged 75% (2,427⁶ correctly produced items). We first analyzed the errors to identify factors that contributed to the difficulty of the items, and then using a multiple regression we examined the relative contribution of these factors to the model's performance.

Errors were categorized as follows: (a) suffix errors: the stem was produced correctly but the suffix was incorrect; (b) stem errors: the suffix was produced correctly but the stem contained one or more incorrect phonemes; and (c) suffix + stem errors: both the stem and the suffix contained one or more incorrectly produced phonemes. As in the corpus analysis, the stem was defined as the word without its inflectional endings. For example, for the noun *medved*(bear) (Tables 1 and 2), *medved* is considered the stem.

Stem errors were the most common: An average of 90% of incorrectly produced items contained an error in the stem. A smaller proportion of incorrectly produced items (8% on average) contained an error in the suffix only. Finally, an average of 2% of the incorrectly produced items contained an error both in the stem and in the suffix. The nature of suffix and stem errors is examined more closely below.

3.3.2.1. Suffix errors: The data presented below are the percentage of incorrect items calculated relative to the number of items of each type in the training corpus, averaged across five runs of the model. For example, in the second run of the simulation, the model produced 10 genitive plural feminine items incorrectly. There were 104 items of this type in the training corpus, and thus 10% of genitive plural feminine items were produced incorrectly. This way of describing the data represents more clearly the relative difficulty of each type of form. Thus, even though there was the same number of items (10) produced incorrectly in the

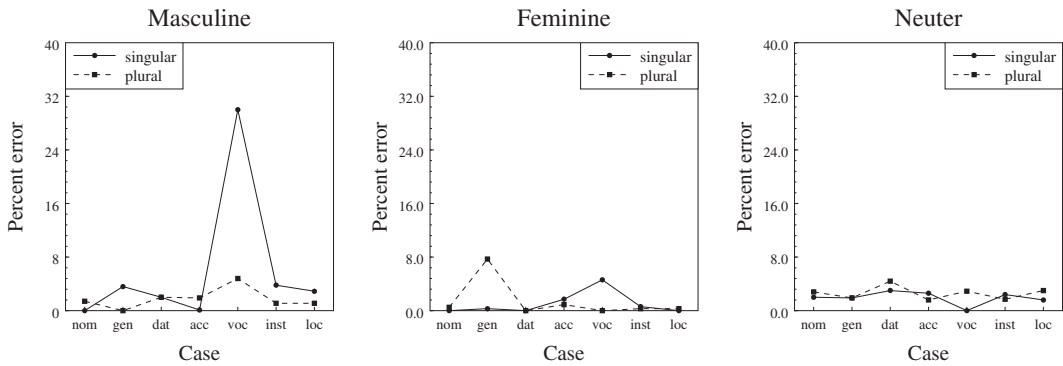


Fig. 3. Percentage of items produced with an incorrect suffix (and a correct stem) in the partially trained model (averaged across five runs).

vocative singular masculine items, this was a much larger percentage (50%) than in the case of genitive plural feminine forms, as there were only 20 vocative singular masculine items in the training corpus compared with 104 genitive plural feminine items.

The distribution of items produced with the correct stem and an incorrect suffix across different inflectional forms is presented in Fig. 3 and further analyzed below. On average, there were more suffix errors in masculine (4%) than neuter (2%) and feminine (1%) gender. There were also more errors in singular (3%) than plural (2%) forms.

At this point in training, the vocative singular produced the largest proportion of suffix errors (30%) in the masculine gender. This form produces the largest proportion of suffix errors by far (Fig. 3). Vocative case is relatively rare; only 0.8% of all masculine noun tokens in the 2 million corpus are in the vocative singular case. In addition, unlike in other case forms, there are two possible suffixes in this form: /e/ (55%), /u/ (40%), or no overt suffix because the nominative singular form is used.

Errors were further analyzed to determine whether they were phonemic (i.e., the most similar phoneme to the target was produced) or overgeneralizations from suffixes for other similar forms. We focused on the most likely overgeneralizations here, such as alternative suffixes for the same gender + case + number form, suffixes from the same gender and case but different number, or from the same case and number but different gender (see examples in Table 6). For example, for the vocative singular form in masculine nouns the suffix /a/ is considered to be a true phonemic error as neither feminine nor neuter vocative singular forms nor masculine vocative plural take this suffix (cf. Table 6). However, suffixes /u/, /i/, and /o/ can be described as overgeneralizations from, for example, the alternative masculine vocative singular suffix (/u/), masculine vocative plural (/i/), or feminine vocative singular form (/o/). Owing to the relatively high similarity in the phonological representation of vowels (see Table 5), in some cases it was not possible to unambiguously establish whether the errors were overgeneralizations or true phonemic errors, and thus we refer to these errors as ambiguous. For example, if /i/ is produced instead of the correct suffix /e/, it is not clear whether this is an overgeneralization from the masculine vocative plural or a pure phonemic error, as /i/ is the closest to /e/ in the phonological space.

Table 6
Examples of suffix overgeneralization errors for the target *medvede* (bear)-
MASC.VOC.SG

	Type
medvedu	Alternative suffix for the same gender + case + number e.g., MASC.VOC.SG.: <i>konju</i> (horse)
medvedi	Same gender + case, different number e.g., MASC.VOC.PL.: <i>medvedi</i> (bear), <i>konji</i> (horse)
medvedo	Same case + number, different gender e.g., FEM. VOC. SG.: <i>kravo</i> (cow)

Note. See note 2 for abbreviations.

Table 7
Examples of suffix errors in the training corpus in the partially trained model

	Correct Form	Phonemic Error	Ambiguous Error	Case	Number	English Gloss
Masculine	profesore	profesora	profesori	VOC.	Singular	<i>professor</i>
	majstore		majstori	VOC.	Singular	<i>repairman</i>
	smehu		smeho	VOC.	Singular	<i>laughter</i>
Feminine	drugaritse	drugaritsa		VOC.	Singular	<i>friend</i>
	unuko		unuku	VOC.	Singular	<i>granddaughter</i>
	opni		opne, opna	GEN.	Plural	<i>membrane</i>
	ve3bi		ve3be	GEN.	Plural	<i>exercise</i>

Note. See note 2 for abbreviations.

In the vocative singular of masculine nouns on average 77% of incorrectly produced items with a suffix error had an ambiguous error (see examples in Table 7). Of these, 41% were produced with the suffix /o/ (which is also the most similar phoneme to the target suffix /u/), 37% with the suffix /i/ (which is also the most similar phoneme to the target suffix /e/), 12% with /e/, and 10% with /u/. This shows that the model has learned that the vocative singular suffix in masculine nouns is a vowel, and that it is biased toward producing vowels that are most similar to the target vowel. Owing to the high similarity of the target vowels and the feminine vocative singular suffix (/o/) and masculine vocative plural suffix (/i/), it is not clear whether these errors are overgeneralizations or merely phonemic (speech) errors.

Vocative singular was also one of the two forms with the highest percentage of suffix errors in feminine nouns (5%). This form is also rare (1.3% of all feminine tokens in the 2 million corpus), and unlike the majority of other inflectional forms it has two possible endings, /o/ and /e/. Again, we coded the incorrectly produced suffixes as a phonemic (/a/) or ambiguous (overgeneralization/phonemic) error (/u/, /e/) (see examples in Table 7). The majority of items (82%) had an ambiguous error. In 83% of the items with ambiguous errors instead of the target suffix /o/ /u/ was produced, which is a vocative singular suffix for many masculine nouns, but also differs from the target suffix in only one feature.

For feminine nouns, the genitive plural form had the most difficult inflectional suffix: 8% of feminine genitive plural forms was produced with an incorrect suffix at this point in

training. All incorrectly produced items were from feminine nouns taking the suffix /i/ in this form, thus deviating from the majority of feminine nouns that take /a/ (see examples 8(b) and 10 in the corpus analysis). The corpus analysis indicates that 20.2% of feminine nouns belong to this group. We performed the same analysis as before. The two incorrectly produced suffixes were /e/ and /a/, which could both be overgeneralizations from other feminine forms: /a/, the main genitive plural ending for feminine nouns, as well as neuter, or /e/, the genitive singular ending for feminine nouns. The majority of incorrectly produced items (91%) had the suffix /e/, which is also the most similar phoneme to the target suffix /i/ (see examples in Table 7). As in the vocative singular errors, this shows that the model has learned that the suffix for this form is a vowel, and it produces the vowel most similar to the target suffix in the phonological space.

Finally, neuter nouns had the smallest proportion of suffix errors of the three genders, and they were equally distributed across inflectional forms (Fig. 3). This is not surprising because the nouns of this gender have a single set of suffixes, which are consistent across items (Table 4).

Summary. The above analyses demonstrate that the model has relatively little difficulty with the suffixes. In the errors it produces, unsurprisingly the model has most difficulty with low-frequency inflectional forms, such as the vocative singular. In addition, what seems to make these difficult is that they have several alternative endings. In the case of the genitive plural in feminine nouns this is particularly the case. Interestingly, the genitive plural suffix in the subgroup of feminine nouns with the alternative ending /i/ is highly reliable: All genitive plural forms in the nouns ending in a consonant in the nominative singular have the suffix /i/ (example 10 in the corpus analysis), and 68% of the genitive plural forms of nouns with a consonant cluster end in /i/ (example 8(b)). Thus, if only phonological cues are taken into account, the suffix in this inflectional form seems to be highly reliable, yet the model finds it difficult. A possible reason for this is the fact that a relatively small group of nouns behaves in this way (only 20.2%), and despite the reliability of the suffix within the group there are still more nouns with the suffix /e/, when all feminine nouns are taken into account. It is possible that the performance of the model on this group of nouns would improve with distributed semantic representations, as all nouns ending in a consonant are abstract, and this would help the model's performance.

An interesting finding in the analysis of suffix errors is the lack of errors in the accusative singular of masculine nouns. The suffix in this inflectional form depends on the animacy of the referent: Nouns denoting animate referents (25% of masculine nouns in our sample) take the same suffix as the genitive singular form (/a/), whereas inanimates have the same form as the nominative singular, that is, what we consider the stem. As discussed below, the nominative singular form in masculine nouns is the most diverse in terms of the word ending (any consonant but one). The fact that animate nouns take a suffix (/a/) in this form makes them the majority group, which facilitates learning their suffix, even though only 25% of masculine nouns denote an animate referent.

In summary, these results show that even in the partially trained model, suffix errors were relatively rare in all three genders. The most difficult cases were the ones that combined low frequency and a one-to-many mapping from the input representation to phonology. These

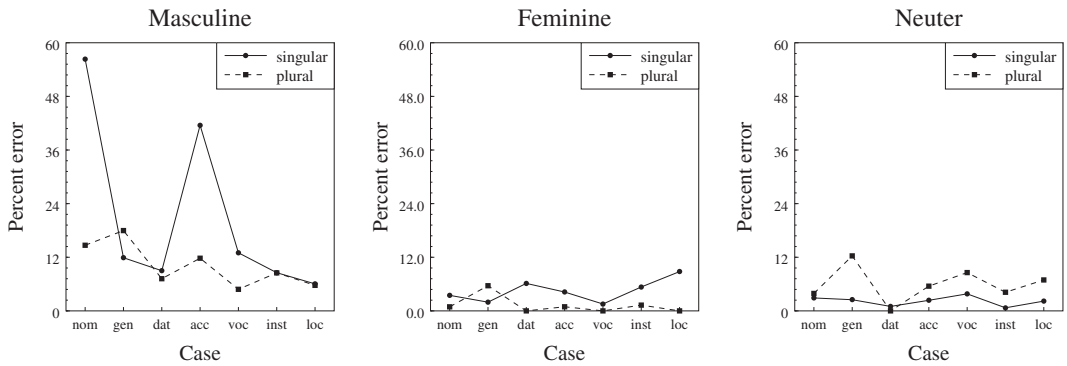


Fig. 4. Percentage of items produced with a morphological error in the stem (and a correct suffix) in the partially trained model (averaged across five runs).

results are similar to the effects of frequency and consistency of spelling-sound correspondences in English (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Jared, 2002; Seidenberg & McClelland, 1989). This suggests that the same processes may be involved in learning both types of information.

3.3.2.2. Stem errors: As with suffix errors, there were more stem errors in masculine (25%) than in feminine (16%) or neuter (15%) genders in the partially trained model. There were also more stem errors in singular (21%) than plural (16%) forms.

To identify properties of the morphophonological alternations that influenced model performance, we first analyzed the items with stem errors according to whether the incorrectly produced phoneme(s) were morphological (in the part of the stem that changes across inflectional forms) or nonmorphological (in the part of the stem that doesn't change across inflectional forms). For example, for the target /dinamitsi/(dynamics)-FEM.LOC.SG., with a palatalized final stem phoneme (see example 9 in the corpus analysis), the output /denamitsi/ was coded as a nonmorphological stem error, whereas /dinamiti/ as a morphological stem error. In slightly less than half of the items with stem errors (46%), all phonemes but the one(s) involved in morphophonological alternations were produced correctly (i.e., contained morphological errors only). As we were interested in establishing what properties of the inflectional system the model finds difficult, we focused the analyses below on this subset of items.⁷

The distribution of morphological stem errors across different forms in the three genders is presented in Fig. 4. The suffix in all these items was produced correctly. Nominative and accusative singular forms were the most difficult forms in masculine gender: 56% and 42%, respectively, of items in these inflectional forms were produced with a morphological stem error. In the majority of these cases (66%), only the final stem phoneme was produced incorrectly: It was either omitted, or a similar (but incorrect) phoneme was produced (see examples in Table 8). A possible reason for this finding is that the nominative singular form of masculine nouns has no typical ending. Unlike feminine and neuter nouns where the final phoneme in the nominative singular is typical (i.e., present in the majority of items) for that

Table 8

Examples of stem errors in the nominative and accusative singular of masculine nouns in the partially trained model

Correct form	Output	Case	Number	English Gloss
drozd	droz	NOM.	Singular	<i>thrush</i>
ludak	ludat	NOM.	Singular	<i>madman</i>
kurs	kur	NOM./ACC.	Singular	<i>course</i>
plan	plab, plad	NOM./ACC.	Singular	<i>plan</i>
živats	živats, živats	NOM./ACC.	Singular	<i>nerve</i>
potomak	potomgak, potomdak	NOM.	Singular	<i>descendant</i>

Note. See note 2 for abbreviations.

particular gender (/a/ and /o/, respectively), any consonant can appear as the final phoneme in the nominative singular form of masculine nouns.⁸ The same is true for both nominative and accusative singular forms of inanimate masculine nouns. In addition, as shown in the corpus analyses, the majority of inflectional forms across genders have only one or two possible endings. As the model learns the structure of the system, it encodes the fact that in the majority of items word endings are redundant. However, the nominative singular (and accusative singular in inanimate nouns) in masculine nouns deviates from this pattern and this makes it particularly difficult for the model to learn.

The remainder of the incorrectly produced nominative and accusative singular items belongs to masculine nouns with phonological alternations involving a change in the syllable structure across inflectional forms (see example 5 in the corpus analysis). For example, the phoneme /v/ in /živats/(nerve)-MASC.NOM.SG. (see Table 9) is the onset of the second syllable (ži-vats) in only 3 of 14 forms, where it is a part of the first syllable (živ-tsa, živ-tsu, živ-tsi-ma, etc.). The internal inconsistency across inflectional forms is reflected in two main types of errors: duplication errors, where the phoneme is produced in both syllables, for example, /op-stan-nak/ for the target /op-sta-nak/(survival)-MASC.NOM.SG. (see example 5(c) in the corpus analysis), and phonemic errors in consonantal slots of the second syllable, for example, /ži-fats/ for the target /živats/(nerve)-MASC.ACC.SG. or /no-fats/ for the

Table 9

Inflectional forms of the masculine noun *živac* (nerve); “-” indicates syllable boundary, provided in the Frequency Dictionary of Serbian (Kostić, 1999)

Case	Singular	Plural
NOM.	živ-vats	živ-tsi
GEN.	živ-tsa	živ-va-tsa
DAT.	živ-tsu	živ-tsi-ma
ACC.	živ-vats	živ-tse
INST.	živ-tsem	živ-tsi-ma
LOC.	živ-tsu	živ-tsi-ma
VOC.	živ-t/é	živ-tsi

Note. See note 2 for abbreviations.

target /no-vats/(money)-MASC.ACC.SG. This type of error reflects the fact that the model encodes and relies on the similarity structure of different inflectional forms of the word, and thus it has most difficulty with forms with the least amount of overlap. To use the example of /zivats/, it is the nominative and accusative singular and the genitive plural that the model will find most difficult (/zi-vats/, /zi-va-tsa/ vs /ziv-tsa/, /ziv-tsu/, /ziv-tsem/, /ziv-tsi/, /ziv-tsima/, and /ziv-tse/). This is particularly true for items with several phonological alternations (examples 5(b–e) in the corpus analysis). In addition, all items with these properties belong to smaller neighborhoods (Table 4), that is, there is also little support from items behaving in a similar way.

The genitive was the most difficult plural form in all three genders, with 18%, 6%, and 12% of masculine, feminine, and neuter nouns in this form, respectively, produced with a morphological stem error. A common property of this inflectional form in all three genders is a group of nouns that exhibit the moving ‘‘a’’: 70%, 100%, and 96% of the incorrectly produced items in masculine, feminine, and neuter genitive plural, respectively, were from this group. This phonological alternation involves a movement of the vowel /a/ within a consonant cluster in the stem, for example: /zivtsa/(nerve)-MASC.GEN.SG., but /zivatsa/-GEN.PL. (Table 9). This alternation changes the syllabic structure of the word, for example: /ziv-tsa/-GEN.SG. contains two syllables, whereas /zi-va-tsa/-GEN.PL. contains three syllables. Importantly, the structure of the syllables is different: /ziv-/tsa/ versus /zi-/va-/tsa/. This is unlike the majority of other forms without this alternation where the syllabic structure across inflectional forms is preserved (compare Tables 1 and 2 with Tables 9 and 10). The model makes similar errors as in the nominative/accusative case, with incorrect phonemes in consonantal slots of the syllables involved in the change, for example, for the target /zi-va-tsa/(nerve)-MAS.GEN.PL. it produces /zi-za-tsa/, /ziz-tsa-tsa/ (see Table 11 for more examples). In all three genders, there are only small inflectional neighborhoods that behave in this way, so there is little support from similarly behaving items.

Finally, in feminine nouns, locative and dative singular were the most difficult singular forms. In the majority of the forms produced with a morphological stem error (93%), only the final stem phoneme was incorrect, and these items involved palatalization in this

Table 10
Inflectional forms of the feminine noun *igla* (needle) and neuter noun *ve-slo* (oar); ‘‘-’’ indicates syllable boundary

Case	Feminine		Neuter	
	Singular	Plural	Singular	Plural
NOM.	i-gla	i-gle	ve-slo	ve-sla
GEN.	i-gle	i-ga-la	ve-sla	ve-sa-la
DAT.	i-gli	i-gla-ma	ve-slu	ve-sli-ma
ACC.	i-glu	i-gle	ve-slo	ve-sla
INST.	i-glom	i-gla-ma	ve-slom	ve-sli-ma
LOC.	i-gli	i-gla-ma	ve-slu	ve-sli-ma
VOC.	i-glo	i-gle	ve-slo	ve-sla

Note. See note 2 for abbreviations.

Table 11

Examples of stem errors in the genitive plural in the training corpus in the partially trained model

	Correct Form	Output	Case	Number	English Gloss
Masculine	maslat/aka	maslatsaka, maslataka	GEN.	Plural	<i>dandelion</i>
	pritisaka	pritisaka, priritaka	GEN.	Plural	<i>pressure</i>
Feminine	zema/áa	zemjaja, zemjala, zem/ána, zem/ara	GEN.	Plural	<i>country</i>
	igala	izala, izara, ivara, ibara, irara	GEN.	Plural	<i>needle</i>
Neuter	iskustava	iskusvava, iskufava, iskosvava	GEN.	Plural	<i>experience</i>
	vesala	vesara, vesara, vezala, vet/ala	GEN.	Plural	<i>oar</i>

inflectional form. The model produced either the unpalatalized stem phoneme (e.g., /fabriki/, /prugi/, /svrhi/ for targets /fabritsi/(factory), /pruzi/(railroad), /svrsi/(prupose)), or a phoneme very similar to it (e.g., /fabriti/). Again, there were relatively few items with this property in the training corpus.

3.3.2.3. Inflectional neighborhood size: One consistent finding in the above analyses is that the items produced with a morphological stem error are from groups of nouns with relatively few members, that is, nouns where there are relatively few other nouns behaving in the same way across inflectional forms. We calculated the proportion of nouns behaving in similar ways across inflectional forms for all three genders and refer to this measure as *inflectional neighborhood size* (Table 4). This is a measure of consistency in the model's input–output mappings; for example, across the majority of inflectional forms of feminine nouns only the suffix changes, that is, there is a high degree of consistency of the input (lemma, gender, case, number) to output (phonology) mapping (e.g., Tables 1 and 2). However, in 5.26% of feminine nouns in addition to the suffix change the mapping involves a change in the stem-final phoneme in some inflectional forms (the palatalized forms in Table 4, example 9 in the corpus analysis). This is analogous to the consistency of mapping in orthography to phonology in reading; for example, the orthographic string *ost* is sometimes pronounced as /o/ as in *cost* and sometimes as /ou/ as in *host*, whereas the string *ike* is always pronounced as /aik/. What we refer to as *inflectional neighborhood size* is just one way to measure the degree of consistency in the mapping, that is, the proportion of items in a corpus behaving in the same way across inflectional forms (friends). This type of measure has been a useful tool in exploring other phenomena described within the connectionist framework, and it captures well how the models of this kind learn and represent information (e.g., Jared, 2002; Jared, McRae, & Seidenberg, 1990). The analyses above indicate that this may be a useful measure to describe relevant features of a morphological system as well.

We explored the effect of consistency in mapping captured by inflectional neighborhood size by correlating this measure with the proportion of different types of stem errors. If this measure captures important aspects of morphological structure, it should be negatively correlated with the proportion of morphological stem errors but it should not be correlated with the proportion of phonemic stem errors. This would indicate that the larger the proportion of items behaving in a morphologically similar way, the less difficult (in terms of morphological properties) the items are for the model to learn. This is exactly what we found: The

correlation between the proportion of morphological stem errors and inflectional neighborhood size across the three genders is negative and significant ($r = -.54, p < .05$): the larger the neighborhood the fewer the errors. There is no significant correlation between inflectional neighborhood size and the proportion of nonmorphological stem errors ($r = .25, p = .35$). This shows that the model's performance improved when morphological alternations were present in a larger proportion of items. This finding reflects one of the critical aspects of models of this kind: The network's performance is influenced by all words in the training corpus, and there is no distinction between rule-governed and exceptional cases. What matters is a graded similarity among items and how much support different mappings get from items behaving in a similar way.

Summary. The above analyses establish that the model had more difficulty with the properties of stems rather than suffixes. This result is somewhat surprising, because discussions of systems such as Serbian morphology typically emphasize the complexity of the inflections, not processes that affect the stems (e.g., Kostić, 1995; Lukatela et al., 1980). Our result suggests that suffixes are relatively consistent across different inflectional forms, whereas there are more subregularities among the stems (see Table 4). For example, in masculine nouns, we identified only two possible sets of suffixes, whereas there were 10 different subgroups of nouns based on morphophonological properties of stems.

The model was also sensitive to the complexity of the morphophonological properties that each gender entailed. The properties of the input to form mapping varied greatly in each gender. Masculine nouns have a more complex mapping than feminine or neuter nouns (see Table 4). For example, an accusative singular in neuter nouns maps to a single suffix directly attached to the stem, whereas in masculine nouns the accusative singular can take a suffix /a/, a suffix /u/, or no suffix, and in addition there can be a deletion of a phoneme (as in /ovan/-ovna/) or a voicing assimilation (as in /vrabats/-vraptsa/). This difference in consistency across genders appeared to strongly influence network performance. Masculine nouns yielded the greatest error rate both with respect to stem and suffix errors, even though they have the highest type frequency. In contrast, performance was better with feminine nouns since even though they are slightly lower in type frequency than masculine nouns, their input-form mapping is more consistent. For example, there are only five subgroups based on stem properties. Neuter nouns are the least complex in terms of morphophonological consistency, as they involve only one set of suffixes and only three subgroups based on stem properties; however, performance for these items was poorer than for feminine nouns, reflecting their lower type frequency.

The effects of frequency and consistency are also seen in the analyses of individual inflectional forms. For example, the largest proportion of suffix errors was found in the vocative singular of masculine nouns which can take three different forms—a suffix /e/, a suffix /u/, or no suffix and in addition is one of the least common forms.

3.3.3. *Learning: Analyses of correctly produced forms*

To examine more closely the effects of the factors identified in the error analyses, we explored how they influence performance on the correctly produced items in the fully trained networks, using stepwise multiple regression. Cross-entropy error reflects the

distance of the network's output to perfect output, so for correctly produced items this measure can be used to assess the relative difficulty of learning. We used the cross-entropy error averaged across the five networks as the dependent variable. The predictor variables in the regression consisted of two measures of frequency previously found to influence word recognition in other languages (Baayen et al., 1997): log surface (form) frequency (the frequency of each inflectional form) and log lemma frequency (the summed frequency of all inflectional forms). We also included word length (number of syllables), which is also known to influence human performance cross-linguistically (e.g., Balota et al., 2004; Barca, Burani, & Arduino, 2002; Cuetos & Barbon, 2006). Finally, we included a measure of morphological consistency, inflectional neighborhood size (Table 4).

All of these factors accounted for a significant proportion of variance in the model's performance (cross-entropy error) (Table 12). The model's performance was influenced more strongly by inflectional neighborhood size, a measure of consistency of input to phonology mapping, than by lemma or surface frequency. This result confirms the findings obtained in the analyses of the incorrectly produced items and underscores the importance of consistency of the input to output mapping in the processing of morphologically complex words.

3.3.4. Generalization

The modeling results so far indicate that a relatively simple artificial neural network can learn a complex morphological system. Examining the items that were most difficult for the model to learn was informative as to how the model solved the task as well as what factors shaped its performance, including word frequency and a measure of mapping consistency. These findings suggest that the model has encoded statistical properties of the system rather than merely memorized each word. The sensitivity to inflectional neighborhood size, in particular, provides evidence that it has picked up generalizations across items related to morphological structure. However, a stronger test of the model's ability to generalize based on prior learning is provided by testing the trained network on items that were not included in the training set, a "wug" test (Berko, 1958). Generalization performance also provides an additional basis for assessing factors relevant to the model's performance.

Generalization was tested using a corpus of 90 inflected items that were explicitly withheld from the model's training set. For example, in training the model was exposed to the nominative and dative singular and nominative and accusative plural forms of the noun *jastog* (lobster), but not to the genitive plural form. Withholding this last item from the training corpus permitted us to test how the network could generalize what it had learned

Table 12
Predictors of average cross-entropy error in learning

Variable	Unique Variance (%)
Inflectional neighborhood size	8.5
Log(lemma frequency)	4.6
Log(surface frequency)	3.0
Word length	1.4

Note. $p < .001$ for all variables.

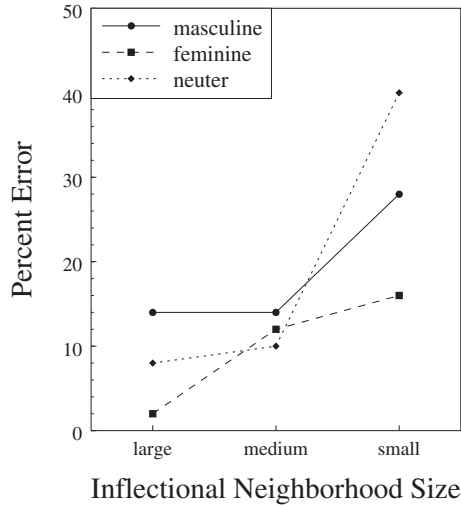


Fig. 5. Percentage of errors in the generalization test (averaged across five runs).

about this noun, and about other masculine genitive plurals. The holdback set consisted of 30 items per gender. As the analysis of learning performance indicated that the model was influenced by inflectional neighborhood size, we chose items from small (less than 10%), medium (10%–40%), and large (above 40%) neighborhoods in each gender. Lemma frequency was closely matched across genders and categories. Each network was tested on all of the holdback items at the conclusion of training, when it was correctly producing all items in the training corpus. Items were scored using the same method described above.

Correct output was 85% averaged across runs of the network.⁹ This value should be considered relatively high given that the input–output mapping is not strictly deterministic and so the model could produce output that is plausible but nonetheless incorrect. Also, the model was not provided with all information that is clearly relevant to human performance (see below). This indicates that the network extracted the regularities present in the training corpus allowing it to produce the forms not seen during training to a great degree of accuracy. We performed an ANOVA on error rates across networks as the dependent variable, and gender and inflectional neighborhood size as independent variables (Fig. 5). There was a main effect of gender, $F(2, 8) = 22.02, p < .01$, with more errors in neuter (19.34%) than masculine (18.87%) and feminine (10%) gender. There was also a main effect of inflectional neighborhood size, $F(2, 8) = 29.79, p < .001$, with the most errors in small neighborhoods (28%) relative to medium (12%) and large (8%). There was also a significant gender \times inflectional neighborhood size interaction, $F(4, 16) = 4.50, p < .05$.

The results show that neuter nouns were more difficult in generalization, whereas in learning, masculine nouns were the most difficult. This may indicate that even though the model was able to learn neuter nouns with relative ease, perhaps because of their low-type frequencies it did not encode the nouns of this gender in enough detail to support generalization.

This analysis also underlines the importance of inflectional neighborhood size. Interestingly, the most difficult were neuter nouns (the smallest percentage of nouns in the corpus)

Table 13
Overregularizations in the generalization test

	Correct Form	Output	Case	Number	English Gloss
Neuter	jastut/etom	jastut/om	INST.	Singular	<i>small pillow</i>
	devojt/eta	devojt/a	GEN.	Singular	<i>young girl</i>
Masculine	pisca	pisac	ACC.	Singular	<i>writer</i>
	savetnika	savetnik	ACC.	Singular	<i>advisor</i>
	kovat/a	kovat/	ACC.	Singular	<i>blacksmith</i>
	diplomate	diplomata	GEN.	Singular	<i>diplomat</i>
Feminine	beležaka	bele/ka	GEN.	Plural	<i>note</i>

Note. See note 2 for abbreviations.

in the small neighborhoods (i.e., cases where a morphophonological pattern occurred relatively rarely within a gender): The model produced 40% of these items incorrectly in the test corpus. This indicates that for generalization, the size of a neighborhood of items that behave in a similar way is more important than the frequency of individual items, given that in the test corpus there was no significant difference in lemma frequency in different genders. It is also possible that the model's performance on these nouns was relatively poor because it lacked semantic information which they may have in common. For example, some of the nouns in this group refer to animate offspring, for example, /odojt/e/(infant), /kut/e/(puppy), and /sirot/e/(orphan)/.

We also examined the errors the model produced in the generalization set. On average 32% of errors were unambiguously overregularizations (see examples in Table 13). The majority of them (82%) were with items from small neighborhoods. For example, in feminine and neuter nouns, the largest proportion of nouns are "simple forms" (Table 4): inflectional forms derived by changing the suffix without any transformation of the stem. This is exactly what the model produced for some items which actually require stem transformations, for example, /jastut/e/(small pillow) in the instrumental singular form requires the extension of the stem with /-t-/ before the adding of the suffix, /jastut/etom/, but the model produced it incorrectly by only adding the suffix to the stem (see Table 13).

Most errors in masculine nouns were a different type of overregularization: The accusative singular form of animate nouns was produced as though they were inanimate (Table 13). Taken another way, the model produced a nominative singular form, which is the same as the accusative singular in inanimate but not in animate nouns. This appears to represent another example of overgeneralizing based on a more frequent pattern: Inanimate nouns accounted for 75% of masculine nouns in this form in the training corpus. Since animacy information was not available to the model, it defaulted to the inanimate form in some cases.

3.4. Discussion

The principal goal of this research was to gain a better understanding of a complex inflectional system, and how it relates to a broader controversy centering on how morphologically complex words are represented and processed. The corpus analysis provided frequency information that is not included in grammatical descriptions. Those data suggest that the

system is characterized by several major regularities within each gender. The complexities of the system lie in morphophonological properties of stems. Even though there are a variety of possible stem transformations within each gender, there are many subregularities. In addition, some seemingly arbitrary morphological exceptions may exhibit additional semantic regularities (e.g., animacy in the accusative singular form of masculine nouns, abstractness in feminine nouns ending in a consonant).

We also found that a connectionist network can encode this complex inflectional system. The task on which the network was trained was solved in a way that supports both learning and generalization. Performance of the model was shaped by factors including form and lemma frequency, word length, and a measure of input–output mapping consistency. The model performed well overall on the generalization test, with the absence of some types of semantic information (e.g., animacy) creating systematic overgeneralizations. The results suggest that the same principles that have been applied in studies of English inflectional morphology (e.g., Joanisse & Seidenberg, 1999; McClelland & Patterson, 2002) extend to more complex systems, such as Serbian. Importantly, the converse does not seem to be true: Mechanisms (such as simple rules and lists of exceptions) that appear to be plausible candidate descriptions of the English system seem ill-matched to a more complex system like Serbian. Thus, the utility of the statistical learning approach becomes clearer when the more complex system is brought into the picture. The results suggest that it would be fruitful to pursue this theoretical framework further, in particular, as the basis of models that address how children acquire knowledge of the system.

4. Inflectional neighborhood size: Comparing computational and behavioral data

Analyzing the model's performance led to the identification of a measure of consistency that shaped its performance, which we termed inflectional neighborhood size. This is a measure of consistency with respect to the proportion of items that have similar morphophonological properties in terms of the mapping of the input to the phonological output across inflectional forms. This property affected both learning and generalization. In the final part of this study, we examined the relationship between the model's performance and people's. We were particularly interested in determining whether inflectional neighborhood size was relevant. It is possible that the model solved the problem of generating correctly inflected forms in a way that differed greatly from native speakers' own solutions. We therefore examined whether the inflectional neighborhood size affected human performance in a naming experiment in which this factor was varied systematically, and we compared subjects' performance with the model's.

4.1. Human performance

4.1.1. Method

4.1.1.1. Materials: There were 150 words in the experiment, 50 of each gender. We chose two neighborhoods in each gender that were different in size but as similar as possible in

Table 14
Properties of different inflectional neighborhoods

Gender	Neighborhood Size	Inflectional	Mean Log	Mean Log
		Neighborhood Size (% lemma within gender)	Form Frequency	Lemma Frequency
<i>Masculine</i>	Large	43.19 (48.55)	1.99 (2.05)	4.14 (4.40)
	Small	8.65 (6.52)	1.95 (2.25)	3.92 (4.99)
<i>Feminine</i>	Large	67.43 (71.93)	2.15 (2.20)	4.34 (4.53)
	Small	7.88 (5.26)	2.33 (2.77)	3.94 (5.28)
<i>Neuter</i>	Large	60.65 (83.33)	2.11 (2.20)	3.98 (4.22)
	Small	14.15 (4.17)	1.89 (2.62)	3.28 (5.93)

Notes. We present data from the training corpus used in the model, with the data from the sample used in the corpus analysis in parentheses. The latter is a more representative corpus since it was randomly drawn from the Frequency Dictionary of Contemporary Serbian (Kostić, 1999; see the description of the training corpus for more details).

Table 15
Properties of the stimuli used in the behavioral experiment (means and SDs)

Gender	Neighborhood Size	Number of Syllables	Summed Log Bigram Frequency	Log Form Frequency	Log Lemma Frequency
<i>Masculine</i>	Large	3.28 (0.84)	17.41 (4.29)	1.69 (1.13)	3.44 (1.47)
	Small	2.88 (0.97)	16.12 (3.96)	1.39 (1.10)	3.44 (1.47)
<i>Feminine</i>	Large	3.12 (0.78)	17.69 (4.31)	1.53 (1.17)	3.59 (1.37)
	Small	3.16 (0.90)	16.76 (3.91)	1.84 (1.17)	3.58 (1.46)
<i>Neuter</i>	Large	3.08 (0.49)	16.56 (2.33)	1.13 (1.20)	2.61 (1.61)
	Small	3.04 (0.85)	16.38 (4.25)	1.06 (0.96)	2.56 (1.42)

mean surface and lemma frequency (Table 14): The large neighborhood in all three genders included simple forms (see Table 4), and the small were moving ‘a’ only in masculine gender, palatalized forms in feminine, and extended forms in neuter.

The stimuli were drawn from the Frequency Dictionary of Contemporary Serbian (Kostić, 1999). Stimulus lists were matched pairwise for the onset and number of syllables, both within and across genders. If the exact onset match could not be found, then we used the word with a very similar onset phoneme. The stimuli were also matched for summed log bigram frequency (Table 15). Within gender, the stimuli in the two categories were also matched for log surface frequency and log lemma frequency (Table 15). However, the difference in frequency was significant across genders (there was a main effect of gender for log lemma frequency, $F(2, 144) = 6.40, p < .01$, and for log form frequency, $F(2, 144) = 3.44, p < .05$), with no significant interactions with inflectional neighborhood size), and thus frequency was used as a covariate in the subsequent analyses.

4.1.1.2. Participants: Forty-six undergraduates from the Department of Psychology, University of Belgrade, Serbia, participated in the study as a part of course requirements. Two additional participants were run, but they were excluded from analyses based on very high error rates (falling greater than 2 SDs above the group mean).

4.1.1.3. *Procedure:* Stimuli were presented individually on the center of a 12-in. screen of an Apple Macintosh iBook, in 24-point font (Charcoal Cyrillic), in white on a black background, using Psycscope (Cohen, MacWhinney, Flatt, & Provost, 1993). At the beginning of each trial, a fixation cross appeared on the screen for 550 ms, after which the word “stimulus” appeared. Subjects were asked to read the word aloud as quickly and accurately as possible. The word was on the screen until the subject started pronunciation, such that activation of the voice key made the word disappear. The experimenter scored responses as correct and incorrect using a button box, which also initiated the next trial. Stimulus presentation was randomized for each subject. Subjects’ responses were also recorded on audio tape, which was used to verify accuracy *post hoc*. The experimental session started with ten practice items (three masculine, three neuter, and four feminine nouns) that were not used in testing.

4.1.2. Results and discussion

Two stimuli were removed from the analyses because of a computer error. Both the incorrect stimulus and its match item were removed so that the properties of the stimuli (onset, word length, frequency measures) remained unchanged. Analyses proceeded as follows: Reaction times for correct responses within ± 3 standard deviations of the group mean were submitted to a gender \times inflectional neighborhood size analysis of covariance with the log form frequency and log lemma frequency as covariates. The analysis revealed main effects of inflectional neighborhood size, $F(1, 45) = 6.31, p < .05$, and gender, $F(2, 90) = 25.23, p < .001$, as well as the interaction of these two variables, $F(2, 90) = 6.89, p < .01$ (Fig. 6). Planned comparisons showed that items from smaller neighborhoods took longer to process in masculine and neuter gender (for neuter gender, this tendency was marginally

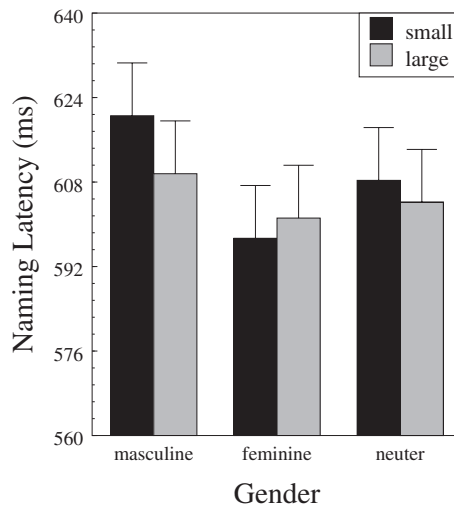


Fig. 6. Mean naming latencies (adjusted for covariates) for items in large and small neighborhoods in the three genders.

significant, $p = .052$). Interestingly, the effect of inflectional neighborhood size in feminine nouns was the opposite: Items from a larger neighborhood took longer to process (the effect here was also marginally significant, $p < .06$).

This pattern of results indicates that masculine nouns produced longer naming times than neuter and feminine nouns, similar to what was found in the model during learning. In addition, the items from small neighborhoods in masculine and neuter gender took longer to process than items from large neighborhoods, again confirming what was found in the model, both in learning and generalization. However, in feminine nouns, we found the opposite pattern. This result is less puzzling if the properties of other words in the language are considered. The phonological pattern typical for the nouns from the smaller neighborhood, palatalization, occurs in other parts of the language (e.g., masculine nouns, verbs), so it is possible that its type frequency (or neighborhood size) in the language as a whole is much higher than in feminine nouns alone, and this may have made them easier to process.

4.2. Performance of the model

We next performed a more fine-grained analysis of the effect of inflectional neighborhood size in the model to more closely compare it with human performance. We selected a set of items from the same neighborhoods used in the naming experiment such that their frequency and length were matched across different neighborhood sizes, yielding a test corpus of 539 inflectional forms (287 from the large neighborhoods and 252 from the small neighborhoods), derived from 80 lemmas in total: 12 masculine, 14 feminine, and 14 neuter nouns per neighborhood size (Table 16), with log lemma frequency and log form frequency equated across neighborhood size within gender. Log lemma frequency, log form frequency, as well as word length differed significantly across genders; to account for this, these factors were included as covariates in subsequent analyses.

All inflectional forms of these nouns in the training corpus were part of the test set. Cross-entropy error for items in the fully trained model was used as the dependent variable, and each run of the model was taken as a “subject.” We ran an ANCOVA, with inflectional neighborhood size and gender as independent variables, and word length, log form frequency, and log lemma frequency as covariates. Both factors as well as their interaction reached significance: inflectional neighborhood size, $F(1, 4) = 124.88$, $p < .001$; gender,

Table 16
Properties of the stimuli used in the computational experiment (means and SDs)

Gender	Neighborhood Size	Number of Lemmas	Number of Forms	Number of Syllables	Log Form Frequency	Log Lemma Frequency
<i>Masculine</i>	Large	12	90	3.33 (0.78)	1.44 (1.23)	4.14 (1.35)
	Small	12	75	2.48 (1.08)	1.63 (1.31)	4.05 (1.17)
<i>Feminine</i>	Large	14	102	3.08 (0.59)	2.08 (1.53)	4.78 (1.56)
	Small	14	96	2.80 (0.67)	2.09 (1.53)	4.83 (1.49)
<i>Neuter</i>	Large	14	95	2.71 (0.89)	1.77 (1.52)	4.49 (1.82)
	Small	14	81	2.97 (0.72)	1.62 (1.55)	4.17 (1.83)

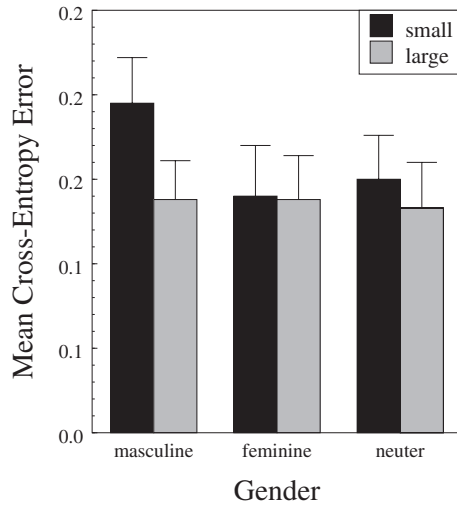


Fig. 7. Mean cross-entropy error (adjusted for the covariates and averaged across five runs) for items from small and large neighborhoods across the three genders.

$F(2, 8) = 33.16, p < .001$; inflectional neighborhood size \times gender, $F(2, 8) = 41.59, p < .001$. As indicated in Fig. 7, masculine nouns produced the largest cross-entropy error followed by neuter and feminine nouns. This replicates the findings reported in the analyses of the model's performance during learning. Furthermore, the items from small neighborhoods produced on average more error than the items from large neighborhoods. This was modulated by the interaction with gender, such that the difference between the large and the small neighborhoods was greatest in masculine and smallest in feminine nouns. This is consistent with the observation that palatalization, the phonological pattern typical of the smaller neighborhood of feminine nouns, occurs in masculine nouns as well, which would weaken the effect in feminine nouns.

4.3. Discussion

In exploring how a connectionist model solved the task of learning a complex inflectional system, we discovered that a measure of consistency of mapping, inflectional neighborhood size, influences the model's performance. We conducted a naming experiment with native Serbian speakers and found that this factor also influences human performance. An unexpected finding was that the effect of inflectional neighborhood size was reversed in feminine nouns. These results suggest that performance on the feminine nouns from the small neighborhood was affected by items elsewhere in the lexicon that patterned similarly. This increases the size of the group of words that exhibit the same morphophonological pattern, which in turn may have made the feminine nouns from the small neighborhood easier to process relative to items in the corresponding masculine and neuter conditions. Importantly, the similarity in the findings between the model and human participants indicates that the

way the network learned the small subset of the Serbian noun system was similar in important ways to how humans process the system.

4.3.1. Comparison to a probabilistic rule-based model

Inflectional neighborhood size is also related to important work by Albright and Hayes (2003) (henceforth AH), who have implemented a model with general rule-induction procedures in learning morphological systems. Complex inflectional systems, such as Serbian's, provide a strong test for their approach. AH's model derives morphological rules by making pairwise comparisons between inflected forms of a lemma. The purpose of this iterative process is to derive a set of rules of increasing generality. The more general rules resemble the informal notion of rule (as in the theory of Pinker and colleagues); however, the model also induces rules that are narrower in scope. In the limit—suppletive forms, such as *go-went*—a rule can apply to a single case. All rules in the AH model operate over phonological units. Each derived rule gets a confidence score depending on the extent to which it successfully predicts inflectional variants for a given phonological context. For example, in the English past tense, the phonological context [ŋ] often produces the [I] to [Λ] transformation of the present to the past form (e.g., *swing-swung*, *sting-stung*). This can be described as a rule: [I] → [Λ] / —[ŋ]_[+past], the accuracy of which can be calculated by identifying the number of verbs within a lexicon with a specified phonological context in the present tense (e.g., [ŋ] ending) and the number of items where the required morphological change (e.g., [I] → [Λ]) occurs in the past tense. The ratio of these two numbers represents *raw confidence*, that is, accuracy with which the rule applies in a particular corpus. Thus, if all items with a defined phonological context in the present tense involve the specified morphological change in the past tense, the rule has very high confidence values. Phonological contexts in which a particular morphological transformation works well are termed islands of reliability. The performance of the model is assessed by first applying the derived rules to produce inflectional forms for novel items, and then comparing the confidence measure associated with the rules that produced the items with human grammaticality judgments of the same items.

The AH model is similar to ours in some important respects. In both approaches, regularities are encoded by a single computational mechanism; neither incorporates an explicit distinction between “rule-governed” and “exception.” Connectionist models are also known to pick up regularities at different grain sizes (see Seidenberg & McClelland, 1989, for discussion). Both approaches emphasize the importance of similarity (islands of reliability, inflectional neighborhoods) in deriving generalizations. Given the overlap between the approaches, we would expect them to account for many of the same phenomena. We were also interested in how they differ. In the future, it would be important to systematically compare AH and connectionist style models. What follows is a preliminary step in this direction.

We trained the AH model¹⁰ on the same training corpus used in the simulation presented above. The model's task was to derive rules describing the mapping between the nominative singular form of the lemmas and all other inflectional forms in the training corpus. The model was tested in two ways. First, we explored whether the derived rules can produce the correctly inflected forms of the Serbian nouns and specifically whether the

rules captured variables to which language users are sensitive, such as word length and inflectional neighborhood size. Second, we assessed whether the rules were sensitive to inflectional neighborhood size, using the same items as used in assessing our model.

The same training corpus as in the simulations presented above was used as input to the AH model. In “training,” the AH model-derived rules that described the transformation of the 407 lemmas in the nominative singular form to the other inflectional forms in the training corpus. The first test consisted of applying the derived rules to the 407 lemmas in the nominative singular form and analyzing the resulting inflectional forms. For the 407 lemmas in the training corpus, the model produced 8,715 unique inflectional forms. These items were scored such that each produced inflectional form that matched an item in the training corpus was scored as correct; 23.4% of the produced items were correct according to this criterion. A further 6.7% of the produced items were grammatical but were not in the training corpus.¹¹ The items scored as correct accounted for 97.4% of the items in original training corpus, and the remaining 2.6% of items were not generated by the derived rules.

These findings indicate that the rules derived by the model capture some properties of Serbian nouns, insofar as they can produce most items in the training corpus. However, the majority of items that the rules produced were ungrammatical (69.9%). The average confidence for the rules that produced the correct items was 0.15 (range: 0.0004–0.53) compared with 0.018 (range: 0.0004–0.26) for the average confidence of ungrammatical items. Despite the difference in the average confidence scores between these items, there is a large overlap in the distributions, and so it remains unclear how the model can distinguish between the grammatical and ungrammatical items.

We also assessed the extent to which performance of the AH model was influenced by factors known to influence human behavior. The model cannot account for frequency effects because frequency is not represented in the model. In a multiple regression with confidence values associated with each correctly produced form as the dependent variable, inflectional neighborhood size accounted for a significant proportion of variance, $r^2 = .03$, $p = .002$, but length did not.

In the second test, we assessed the model on the same inflectional forms that were used in comparing our model to human performance (see section 4.2). We extracted all inflectional forms of the 80 lemmas used in the test that were correctly produced by the AH model (97% of items in the test corpus) and ran an ANOVA with gender and inflectional neighborhood size as factors. This yielded gender as the only significant factor, $F(2, 74) = 10.84$, $p < .001$, with masculine nouns produced by rules with the lowest confidence values (Fig. 8). This indicates that the model captured phonological properties that differentiated genders. However, there was no effect of inflectional neighborhood size, unlike the human data.

In summary, these findings suggest that the AH model successfully captures some phonological properties of the Serbian noun system. However, the model differs from human participants in several ways. First, the rules produce both grammatical and ungrammatical forms (mainly the latter). Some kind of filtering mechanism would have to be introduced to eliminate the ungrammatical forms. Second, the model is not sensitive to word length or frequency, which affect human performance. This information could be added in a

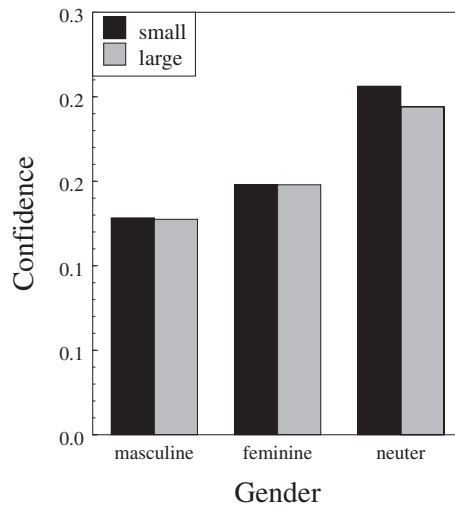


Fig. 8. Adjusted confidence values from the MinGenLearner model.

“performance” module; our approach obviates this need. Finally, the AH model is specifically designed to derive rules about sublexical structure. In our model, the sensitivity to sublexical properties is an emergent property of the process of learning to produce words. Importantly, connectionist models of the kind used here will develop sensitivities to regularities at various grain sizes, and inflectional neighborhood size is just one of them. In short, we see a model such as AH’s as a kind of competence theory of what neural networks such as ours learn. The model captures some of the same structure, but it excludes information that is relevant to how humans learn. The model is powerful enough to capture many properties of Serbian, but it requires rules that apply to single cases. It requires the capacity to compare alternative forms, whereas our models learn on the basis of feedback about the correctness of the computed output. Other similarities and differences between the approaches merit further investigation.

5. General discussion

The idea that the English past tense is a quintessential linguistic rule has been widely popularized in recent years. As linguistic subsystems go, it is rather atypically simple. It is possible that the English past tense represents the distillation of essential properties of language, which are merely harder to discern in more complex cases, such as Serbian morphology. The alternative is that English inflectional morphology is atypically simple and thus does not support broader generalizations about the nature of linguistic knowledge. (It is also possible that English and Serbian inflectional morphology are simply different.) Our investigations of Serbian suggest that it is similar to English, not because both are felicitously described by rules, but because both are quasiregular. As such, they can both be characterized by systems that learn statistical mappings between codes. In doing so, the model we

developed provided new insights about properties of the Serbian system and language processing. For other examples of insights about the nature of morphological knowledge and language processing that can be gained by studying systems other than the English past tense, see, for instance, Boudelaa and Marslen-Wilson (2001); Meunier and Marslen-Wilson (2004); Orsolini and Marslen-Wilson (1997); and Plaut and Gonnerman (2000).

The main limitations of our model concern the explicit representation of gender, case, and lemma as input. We assume that these types of information are also learned, but we have not demonstrated this here. Elsewhere we have shown that a similar model could produce correctly inflected forms without an explicit, localist representation of gender (Mirković et al., 2005, 2008), because grammatical gender is correlated with semantic and phonological properties of words (see also Corbett, 1991; Kelly, 1992). Similarly, lemma representations may reflect the shared phonological and semantic information structure among the words in the inflectional paradigm of a noun. On this account, factors such as gender and lemma emerge as hidden-layer representations in systems that map between semantics and phonology. Case information relates to the sentence structures in which words occur. Interestingly, there is some behavioral evidence that it is the individual functions the cases serve in different contexts that are relevant for adult processing (e.g., Kostić, 1995) and not an abstract notion of case. In general, using localist representations as input limits model performance in important respects; encoding semantic information at a finer grain would allow a model to capture additional regularities (e.g., involving animacy). Thus, further investigations can be aimed at exploring the use of distributed representations, as well as modeling more closely how children acquire such systems.

In summary, we were interested in the structure of Serbian noun inflection, whether this system could be learned by a connectionist network, and whether the network's representation of this system corresponded in important ways to that of skilled adult language users. The results suggest that Serbian conforms better to the quasiregular concept than to a rules-and-exceptions dichotomy. The model demonstrated that such a system can be learned by a connectionist network, given a suitable specification of the input and sufficient training examples. The model's learning was shaped by factors such as frequency and morphophonological similarity that also influence the performance of adult speakers. In traditional grammar-based theories factors such as these are considered relevant only for a subgroup of morphologically complex words, the "irregulars" (e.g., Pinker, 1991; Pinker & Ullman, 2002); however, we have demonstrated that in Serbian, these factors have broader effects. Thus, the model provides support for the view that morphological structures emerge from the convergence of other codes, obviating the need for a discrete, isolable level of morphological representation (see also Plaut & Gonnerman, 2000; Rueckl & Raveh, 1999; Seidenberg & Gonnerman, 2000).

Acknowledgments

Research supported by a grant from the National Institute of Mental Health (RO1-MH58723), when it supported basic research on language. MFJ was supported by a

New Investigator award from the Canadian Institutes for Health Research, and the Canadian Language and Literacy Research Network. We thank Aleksandar Kostić for making available the electronic version of the Frequency Dictionary of Contemporary Serbian, and Jelena Morača, Slobodan Marković and the members of the Laboratory for Experimental Psychology, University of Belgrade, Serbia, for their help with the behavioral study. The computational model was implemented using software developed by Michael Harm, now at Google, Inc., whom we also thank.

Notes

1. Languages referred to as Serbian, Croatian, and Bosnian were previously referred to as Serbo-Croatian.
2. The following abbreviations will be used in the article: SG., singular; PL., plural; FEM., feminine; MASC., masculine; NEUT., neuter; NOM., nominative; GEN., genitive; DAT., dative; ACC., accusative; INST., instrumental; LOC., locative; VOC., vocative.
3. Each noun can have 14 inflectional forms: two numbers \times seven cases.
4. We use the term “lemma” here to denote a localist representation of the word meaning common to all of its inflectional forms.
5. This additional set comprised of nouns from inflectional neighborhoods with fewer members as described in the corpus analysis before. This enabled us to perform a more detailed analysis of the effects of inflectional neighborhood size presented next. The items were drawn from the Serbian frequency dictionary such that their morpho-phonological properties matched the properties of the smaller inflectional neighborhoods in all three genders, keeping the relative proportions of different neighborhoods similar to the ones in the corpus analysis. Pilot simulations trained on only the 2,648 words indicated that the additional items from smaller inflectional neighborhoods did not significantly affect the network’s ability to learn the items in the training set.
6. As a result of an error in coding, four lemmas were excluded from the analyses.
7. The analysis of nonmorphological errors the model produced is informative regarding how the model captures phonological regularities of Serbian and how this relates to human speech errors.
8. We examined the final phonemes of masculine nominative singular forms in the corpus of 2 million words and of 25 possible consonants, only one (*/dʒ/*) was not found.
9. Two items in the test corpus were excluded because of a coding error.
10. The program to run the model, MinGenLearner, was kindly provided by Bruce Hayes at <http://www.linguistics.ucla.edu/people/hayes/learning/#Software>.
11. Given that the training corpus was derived from the Frequency dictionary of Serbian, that is, a sample of natural language, it did not contain all grammatical forms of a given lemma.

References

- Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, *90*, 119–161.
- Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, MA: MIT Press.
- Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language*, *37*, 94–117.
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283–316.
- Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavioral Research Methods, Instruments, & Computers*, *34*, 424–434.
- Bentin, S., & Frost, R. (1995). Morphological factors in visual word identification in Hebrew. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 271–292). Hillsdale, N. J.: Erlbaum.
- Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150–177.
- Boudelaa, S., & Marslen-Wilson, W. D. (2001). Morphological units in the Arabic mental lexicon. *Cognition*, *81*, 65–92.
- Brown, W. (1993). Serbo-Croat. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic languages* (pp. 306–387). London and New York: Routledge.
- Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28*, 297–332.
- Carstairs-McCarthy, A. (1998). Paradigmatic structure: Inflectional paradigms and morphological classes. In A. Spencer & A. M. Zwicky (Eds.), *Handbook of morphology* (pp. 13–44). Oxford, England: Blackwell.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope: A new graphic interactive environment for designing psychological experiments. *Behavioral Research Methods, Instruments, & Computers*, *25*, 257–271.
- Colé, P., Beauvillain, C., & Segui, J. (1989). On the representation and processing of prefixed and suffixed words: A differential frequency effect. *Journal of Memory and Language*, *28*, 1–13.
- Comrie, B., & Corbett, G. G. (Eds.) (1993). *The Slavonic languages*. London and New York: Routledge.
- Corbett, G. (1987). Serbo-Croat. In B. Comrie (Ed.), *The world's major languages* (pp. 391–409). London and New York: OUP.
- Corbett, G. G. (1991). *Gender*. Cambridge, UK: Cambridge University Press.
- Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences*, *14*, 597–650.
- Cuetos, F., & Barbon, A. (2006). Word naming in Spanish. *European Journal of Cognitive Psychology*, *18*, 415–436.
- Daugherty, K., & Seidenberg, M. S. (1994). Beyond rules and exceptions: A connectionist approach to inflectional morphology. In S. D. Lima, R. L. Corrigan, & G. K. Iverson (Eds.), *The reality of linguistic rules* (pp. 353–388). Amsterdam; Philadelphia: J. Benjamins.
- Davis, M. H., Casteren, M. van, & Marslen-Wilson, W. D. (2003). Frequency effects in processing inflected Dutch nouns: A distributed connectionist account. In R. H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 427–462). Berlin: Mouton de Gruyter.
- Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*, 1355–1367.
- Feldman, L. B. (2000). Are morphological effects distinguishable from the effect of shared meaning and shared form? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*, 1431–1444.
- Feldman, L. B., & Prostko, B. (2002). Graded aspects of morphological processing: Task and processing time. *Brain and Language*, *81*, 12–27.
- Feldman, L. B., Soltano, E. G., Pastizzo, M. J., & Francis, S. E. (2004). What do graded effects of semantic transparency reveal about morphological processing? *Brain and Language*, *90*, 17–30.

- Frost, R., Forster, K. I., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23, 829–856.
- Gerken, L. A. (2002). Early sensitivity to linguistic form. *Annual Review of Language Acquisition*, 2, 1–36.
- Gerken, L. A., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, 32, 249–268.
- Gonnerman, L. M., Seidenberg, M. S., & Andersen, E. S. (2007). Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General*, 136, 323–345.
- Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, 111, 662–720.
- Haskell, T. R., MacDonald, M. C., & Seidenberg, M. S. (2003). Language learning and innateness: Some implications of compounds research. *Cognitive Psychology*, 47, 119–163.
- Hinton, G. E. (1989). Connectionist learning procedures. *Artificial Intelligence*, 40, 185–234.
- Ivić, P. (1990). Sistem padežnih nastavaka imenica u srpskohrvatskom književnom jeziku [The system of case suffixes in nouns in standard Serbo-Croatian]. In P. Ivić (Ed.), *O jeziku nekadašnjem i sadašnjem* (pp. 286–309). Belgrade: BIGZ-Jedinstvo.
- Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language*, 46, 723–750.
- Jared, D., McRae, K., & Seidenberg, M. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, 29, 687–715.
- Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology following brain injury: A connectionist model. *Proceedings of the National Academy of Sciences USA*, 96, 7592–7597.
- Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, 99, 349–364.
- Kielar, A., & Joanisse, M. F. (2010). Graded effects of regularity in language revealed by N400 indices of morphological priming. *Journal of Cognitive Neuroscience*, 22, 1373–1398.
- Kielar, A., Joanisse, M. F., & Hare, M. L. (2008). Priming English past tense verbs: Rules or statistics? *Journal of Memory and Language*, 58, 327–346.
- Kim, J. J., Marcus, G. F., Pinker, S., Hollander, M., & Coppola, M. (1994). Sensitivity of children's inflection to grammatical structure. *Journal of Child Language*, 21(1), 173–209.
- Kostić, A. (1991). Informational approach to the processing of inflected morphology: Standard data reconsidered. *Psychological Research*, 53, 62–70.
- Kostić, A. (1995). Information load constraints on processing inflectional morphology. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 317–344). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kostić, D. (1999). *Frekvencijski rečnik savremenog srpskog jezika [Frequency dictionary of the contemporary Serbian language]*. Yugoslavia: Institute for Experimental Phonetics and Speech Pathology and Laboratory for Experimental Psychology, University of Belgrade.
- Kostić, D. (2001). *Korpus srpskog jezika [Corpus of Serbian language]*. Yugoslavia: Institute for Experimental Phonetics and Speech Pathology and Laboratory for Experimental Psychology, University of Belgrade.
- Laudanna, A., Badecker, W., & Caramazza, A. (1992). Processing inflectional and derivational morphology. *Journal of Memory and Language*, 31, 333–348.
- Lukatela, G., Gligorijević, B., Kostić, A., & Turvey, M. T. (1980). Representation of inflected nouns in the internal lexicon. *Memory & Cognition*, 8, 415–423.
- MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In B. MacWhinney (Ed.), *The emergence of language* (pp. 177–196). Mahwah, NJ: Erlbaum.
- MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition*, 29, 121–157.

- Maretić, T. (1963). *Gramatika hrvatskoga ili srpskog jezika [Grammar of the Croatian or Serbian language]*. Zagreb: Matica hrvatska.
- Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, *101*, 3–33.
- McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, *6*, 465–472.
- McCloskey, M. (1991). Networks and theories: The place of connectionism in cognitive science. *Psychological Science*, *2*(6), 387–395.
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavioral Research Methods*, *37*, 547–559.
- Meunier, F., & Marslen-Wilson, W. (2004). Regularity and irregularity in French verbal inflection. *Language and Cognitive Processes*, *19*, 561–580.
- Milinić, P., Filipović, Đurđević, D., & Moscoso del Prado Martín, F. (2009). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language*, *60*, 50–64.
- 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*, 139–167.
- Mirković, J., Seidenberg, M. S., & MacDonald, M. C. (2008). Acquisition and representation of grammatical categories: Grammatical gender in a connectionist network. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1954–1959). Austin, TX: Cognitive Science Society.
- Mrazović, P., & Vukadinović, Z. (1991). *Gramatika srpskohrvatskog jezika za strance [Serbo-Croatian grammar for foreigners]*. Sremski Karlovci: Izdavačka knjižarnica Zorana Stojanovića.
- Orsolini, M., & Marslen-Wilson, W. D. (1997). Universals in morphological representation: Evidence from Italian. *Language and Cognitive Processes*, *12*, 1–47.
- Pinker, S. (1991). Rules of language. *Science*, *253*, 530–535.
- Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: Basic Books.
- Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*, 73–193.
- Pinker, S., & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*, 456–463.
- Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, *15*, 445–485.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56–115.
- Plunkett, K., & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science*, *23*, 463–490.
- Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, *48*, 21–69.
- Rueckl, J. G., & Raveh, M. (1999). The influence of morphological regularities on the dynamics of a connectionist network. *Brain and Language*, *68*, 110–117.
- Rueckl, J. G., Mikolinski, M., Raveh, M., Miner, C. S., & Mars, F. (1997). Morphological priming, connectionist networks, and masked fragment completion. *Journal of Memory and Language*, *36*, 382–405.
- Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *272*, 1926–1928.

- Seidenberg, M. (1993). Connectionist models and cognitive theory. *Psychological Science*, 4(4), 228–235.
- Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, 275, 1599–1604.
- Seidenberg, M. S., & Gonnerman, L. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, 4, 353–361.
- Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23, 569–588.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.
- Seidenberg, M. S., & Plaut, D. C. (2006). Progress in understanding word reading: Data fitting versus theory building. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 25–49). Hove & New York: Psychology Press.
- Stanić, M. (1949). *Tipovi imeničkih deklinacija našeg jezika [Different types of noun declensions in our language]*. Zagreb: Jugoslavenska akademija znanosti i umetnosti.
- Stanojčić, Ž., Popović, L., & Micić, S. (1989). *Savremeni srpskohrvatski jezik i kultura izražavanja [Contemporary Serbo-Croatian grammar and style manual]*. Belgrade, Novi Sad: Zavod za udžbenike i nastavna sredstva.
- Stevanović, M. (1986). *Savremeni srpskohrvatski jezik: gramatički sistemi i književnojezička norma [Contemporary Serbo-Croatian language: grammatical systems and norms]*. Beograd: Naučna knjiga.
- Taft, M. (1979). Recognition of affixed words and the frequency effect. *Memory & Cognition*, 7, 263–272.
- Todorović, D. (1988). Hemispheric differences in case processing. *Brain and Language*, 33, 365–389.
- Trueswell, J. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, 35(4), 566–585.
- Yang, J. F., McCandliss, B. D., Shu, H., & Zevin, J. (2009). Simulating language-specific and language-general effects in a statistical learning model of Chinese reading. *Journal of Memory and Language*, 61, 238–257.