Experimental study of first and second language morphological processing

Kira Gor
University of Maryland

1. Introduction

Linguists are concerned with how language as a system works. Yet, some of them are especially interested in linguistic processing, or in other words, they wish to know what it is that people actually do when they produce and understand speech. And yet others set out to study linguistic processing in different populations of speakers – young children acquiring their first language (L1), child and adult second language (L2) acquirers, or children and adults with language impairments. The last two categories of researchers differentiate between a linguistic construct (rules postulated in a specific linguistic theory) and the actual mechanisms underlying speech production and perception, storage and retrieval of linguistic units, and conduct experiments to find out which theoretical constructs have a better fit to real speech processing observed in human speakers-hearers. In order to do so they utilize a variety of experimental techniques, which allow them to collect different types of speech data and perform analyses. This paper will review several experimental paradigms used in research on the processing of inflectional morphology in different languages, and also different populations of speakers. It will discuss the experimental methods, used in the studies, which contribute to the on-going debates concerning one of the most fundamental problems about human cognition, the problem of what kind of computations underlie linguistic processing and how the human mind deals with regularity and irregularity. Linguistic descriptions use symbolic rules, and formal linguistic theory assumes that linguistic processing, which is performed in the mind of the speaker-

---

1. As many research paradigms dealing with human language, the study of linguistic processing is guided by two important considerations: how authentic and natural the collected speech data are, and therefore, how truly they represent the actual linguistic behavior, and which kinds of instrumental and statistical analyses can be performed on such data. And as in other domains, there is often a conflict between the two considerations, notably, the more authentic the data, say, a long segment of naturally produced spontaneous speech, the least likely it will lend itself to instrumental analysis and quantification. And the opposite is also often true; the constraints placed by some experimental procedures make the use of authentic speech virtually impossible.
hearer, makes use of symbolic rules as well. At the same time, there is a whole direction in the study of language, which claims that symbolic rules are not part of the cognitive processes taking place in language production and perception, and these rules are only part of linguistic descriptions created by linguists. Instead of symbolic rules, human cognition relies on associative patterning in neural networks. According to this latter view, the linguistic mechanisms draw on the weights of connections between the nodes in a connectionist network, and not abstract symbolic rules. The studies conducted in support of these two conflicting views make use of several research methods. This paper will describe (1) an elicitation technique used in a real and nonce verb generation task, and (2) a lexical decision task (LDT), which measures reaction times to different verb-forms. Both these methods contributed significantly to promoting the understanding of morphological processing. The review of these methods will address specifically the role of different kinds of linguistic frequencies – type, token, whole-word, and stem-cluster – as well as phonological similarity, in morphological processing, and their use in linguistic experiments. The paper will focus primarily on design and planning of experiments on the processing of inflectional morphology with a cross-linguistic perspective in mind and special emphasis on first and second language acquisition. Thus, it will provide the basic background knowledge necessary for future research into the mechanisms of morphological processing, while at the same time, paying attention to the theoretical underpinnings of such research and the impact of the obtained results on the current agenda in the debates on the status of symbolic rules in the processing of inflectional morphology. The experimental studies discussed below deal with several languages, which vary in the richness of inflectional paradigms, English, Italian, and Russian.

2. Two theoretical approaches to morphological processing

There are two main points of view with regard to morphological processing, which draw on two understandings of linguistic mechanisms in general. English past tense inflection has become the experimental ground for testing the predictions of the two conflicting points of view concerning morphological processing with most of the data, especially, generated in connectionist modeling, coming from English. Past-tense verbs in English are inflected in two different ways. The vast majority of verbs forms their past tense by a general-purpose concatenative rule, ‘add -ed to the stem,’ which constitutes regular inflec-

2. It is also important to mention that the nativist position strongly supported by formal linguistics is based on the idea of innateness of language. Connectionist modeling places emphasis on learning and studies the impact of different kinds of input on language learning.

3. Nonce words are created by experimenters by manipulating the properties of real words; they are novel to the native speaker, do not have any meaning, but may be assigned a fictitious meaning within the experiment.

4. See the section Input Frequencies and Probabilities in Linguistic Processing below for the definitions of different kinds of frequencies.
First and second language morphological processing

A relatively small group of irregular verbs is inflected by some unpredictable stem change with or without the addition of some kind of inflection. These irregular verbs can cluster into ‘neighborhoods’ with similar stem allomorphy, such as sing-sang, ring-rang, spring-sprang. These properties of English, namely a categorical distinction between regular and irregular past-tense inflection, make it possible to investigate the issue whether regular and irregular processing are performed by two distinct mechanisms or just one, and whether regular, but not irregular inflection makes use of symbolic rule computation.

According to the dual-system approach (Clahsen 1999; Jaeger et al. 1996; Marcus et al. 1992, 1995; Pinker 1991, 1999; Pinker & Prince 1988, 1991, 1994; Pinker & Ullman 2002; Prasada & Pinker 1993; Ullman 1999), which is most consistently represented in Steven Pinker’s Words and Rules Theory, linguistic processing is subserved by two main systems – the mental lexicon and a computational system (Pinker 1999). Words in the mental lexicon are connected to each other based, e.g., on phonological or semantic similarity, and their storage and retrieval are performed in associative memory; symbolic rules are applied by computation. In the domain of inflectional morphology, irregular forms are stored in the mental lexicon while regular forms are computed on-line. There is no need to store regular inflected forms, as they will be assembled by the application of concatenative rules. However, no such rules can exist for the irregular forms, therefore they have to be memorized, and retrieved from the mental lexicon when needed. For example, English speakers do not need to store both the singular form of the noun tree and its plural form trees, since the plural is generated by the application of a regular rule: add -s to the singular stem. The noun child, however, needs to be memorized together with its irregular plural children. One important implication arising from such a standpoint is the assumption that irregular forms processed in associative memory will show frequency effects. The nature of these frequency effects will be discussed more in detail in the next section. Generally speaking, the underlying assumption in research on frequency effects is that the words that are activated more often have stronger memory traces, and as a result, are retrieved from memory faster. Contrary to irregular inflection, regular inflection will not show any frequency effects because symbolic rules are applied when the necessary criteria are met, regardless of the frequency of use. Since regularly inflected words are assembled on-line, they are not stored in undecomposed form (with inflections attached to the stems), and thus cannot show whole-word frequency effects. Thus, the dual-system approach makes clear predictions about the connection between input frequencies to the speaker (learner): frequency effects will be observable in irregular inflection, but not in regular inflection. One of such predicted effects includes shorter reaction times (RTs) in

---

5. Concatenation is addition of linguistic elements to one another, "stringing" them in a bead-like fashion.

6. Note, however, that some past-tense verb forms are identical to the basic stem, e.g., put-put. Such no change verbs are also irregular because they do not have the -ed inflection.

7. Allomorphy refers to variations in the phonological composition of morphemes; these phonologically different variants of the morpheme are called allomorphs.
a lexical decision task (LDT) to high-frequency than to low-frequency irregular word-forms, but no such effect for regular ones. The same strong predictions involve the role of phonological similarity in regular and irregular inflection – only irregular word-forms will show effects of phonological similarity, e.g., in a nonce word generation task. Speakers will use the irregular inflectional pattern only for the nonce words, which are very similar to the existing irregularly inflected words, whereas they will use the regular inflectional pattern regardless of phonological similarity. Over time, the dual-system approach moved from its strong version, according to which no regular inflected words are stored in the lexicon in undecomposed form, to a weaker version expressed in the Words and Rules Theory (Pinker 1999). The Words and Rules Theory admits that a certain number of high-frequency regular word-forms can be stored undecomposed, but that this fact does not affect the general processing model.

The single-system approach is built on the premises that there is no categorical distinction between two different processing mechanisms, the rule-based and the association-based ones (Bybee 1985, 1995, 2002; Langacker 1987, 1988; MacWhinney & Leinbach 1991; Plunkett & Marchman 1991, 1993; Rumelhart & McClelland 1986). All the words, uninflected, regularly and irregularly inflected, are stored in the mental lexicon and form associations based on phonological and semantic similarity. The frequency of these mappings between word-forms in the lexicon is crucial for their processing, e.g., for the speed of their retrieval. If a certain type of mapping, e.g., between the verb stem and the regularly inflected past-tense form, *repeat–repeated, discuss–discussed*, etc., is highly frequent in the network, such pattern of association gains in strength. This associative patterning has been modeled in numerous versions of connectionist networks, in which the weights of connections between the nodes representing word-forms are adjusted depending on the frequency of their activation. Obviously, the single-system approach predicts that both regular and irregular inflection will show frequency effects, since both regular and irregular word-forms are processed in the same associative network. It goes without saying, that phonological similarity lies at the core of single-system models, and its effects are predicted both for regular and irregular inflection. Associative patterning is a proposed alternative to symbolic rule computation, and the single-system approach seeks to show that our mind processes language without symbolic rules. Several research directions, most of them involved in connectionist modeling, espouse the single-system approach, including emergentism, a relatively recent direction, focusing on the emergence of language in an associative network in children acquiring their L1 as a product of input processing (Elman et al. 1996).

Since English does not possess a developed conjugational paradigm, and has only one large regular productive default class and one small irregular unproductive class in the past tense, in other words, it is impossible to divorce regularity, productivity, and high type frequency in it, research on other languages with complex inflectional morphology is in order. First of all, let us briefly define the terms regularity, productivity, type frequency, and default. Regularity in inflectional morphology refers to a pattern involving less changes to the shape of the morphemes involved. Regular inflectional processes can be described by a simple concatenative rule: add a certain ending to the stem.
past-tense and noun plural formation are examples of regular inflectional processes.\footnote{Of course, the -\textit{ed} inflection is pronounced differently depending on the final phoneme of the stem, compare stop-stopped, rob-robbed, and visit-visited.} Productivity refers to the fact that the pattern is used in new words, e.g., English uses the productive -\textit{ed} pattern in blog-blogged. Type frequency refers to the frequency of occurrence of a linguistic pattern, or in other words, to the size of a certain class of words using this pattern. And finally, default is characterized by the most open schema, and it is used "when all else fails" (Bybee 1995:452). It is important to maintain this distinction in the data, because the dual- and single-system theories connect regularity to different linguistic parameters. Thus, the dual-system approach claims that regularity and productivity are interconnected, and therefore, regular inflection has to be productive. Also, it does not explicitly differentiate between regular and default processing, often using the 'regular' and 'default' terms interchangeably. Contrary to that, the single-system approach draws a connection between productivity and high type frequency, and maintains that regularity, productivity, and default are all different parameters.

The cross-linguistic data on the processing of verbal morphology in German, Italian, Russian, Norwegian, and Icelandic, languages with a complex conjugational paradigm and often more than one regular verb class, by adults, children, as well as L2 learners, has shed some new light on the disputes (Chernigovskaya & Gor 2000; Gor 2003, 2004; Gor & Chernigovskaya 2001, 2003a, 2003b, 2005; Clahsen 1999; Matcovich 1998; Orsolini & Marslen-Wilson 1997; Orsolini, Fanari, & Bowles 1998; Ragnasdóttir, Simonsen, & Plunkett 1997; Simonsen 2000). All these studies have found the influence of frequency on the processing of all the verb classes, including the regular default classes, both in children and adults. It should be noted that a series of studies on the processing of German inflectional morphology supports the opposite, dual-system model of morphological processing and demonstrates the differences in the processing of different classes of German participles and noun plurals that are consistent across different experimental paradigms (Clahsen 1999). However, these studies do not maintain the distinction between regular and default inflection, which strongly undermines their claims. Research on the processing of Russian verbal morphology by native Russian adults and children, and adult American speakers of Russian as L2, which will be discussed more in detail below, has produced several results incompatible with the categorical distinction between regular and irregular morphological processing (Chernigovskaya & Gor 2000; Gor 2003, 2004; Gor & Chernigovskaya 2001, 2003a, 2003b, 2005). Thus, in nonce verb generation, all the groups of subjects generalized several conjugational patterns ranging in regularity; type frequency was one of the determining factors in the choice of the conjugational pattern – high type frequency patterns were generalized to novel verbs more readily; in a LDT, RTs were shorter for high-frequency forms in the conjugational paradigm. Generally speaking, the data on verbal processing in several languages other than English indicate that these languages have more than one generalizable regular inflectional pattern or several patterns differing in the degree of regularity, and do not support a categorical distinction between regular and irregular morphological processing.
3. Input frequencies and probabilities in linguistic processing

This section will briefly review the kinds of linguistic frequencies discussed in the literature, and the assumptions about the connection between frequencies and probabilities and their role in linguistic processing. This will prepare us to follow the experimental design of the studies using different experimental methods reviewed below. **Token frequency** is the frequency of the word-form in speech. In other words, token frequency tells us how often a language user encounters a certain word, either by producing it or hearing it produced by other speakers. E.g., all the instances when a given individual has heard the word *spoon* constitute token frequency of the word *spoon* in the input to this individual. Obviously, the real frequencies, either for an individual speaker or for a population of speakers, are unavailable; we can only rely on estimates, which are based on language corpora. At present, few large representative corpora are available for languages other than English. The largest and most used corpus of English, COBUILD, is constantly growing; in January 2002 it amounted to 450 million words. Another widely used source is the CELEX database, which in fact, comprises three different searchable databases, the Dutch, English, and German ones. One of the general challenges for the creators of such databases and language researchers alike is relatively limited availability of transcribed spontaneous speech data, and the necessity to rely extensively on written texts. This certainly limits the reliability of corpus-based estimates. Usually, the standard index for word frequency is the frequency of occurrence per 1 million word usages.

The rationale behind using token frequency counts in psycholinguistic experiments is that the more often the word stored in the lexicon is used, the stronger the memory traces for this word. This will result in greater access speed for this word. For simple undecomposable words consisting of a bare stem and no inflection token frequency is uncontroversial. However, the question arises concerning the psycholinguistic mechanisms underlying storage and access of complex inflected word-forms. It is hypothesized that **whole-word frequency**, or the frequency of the inflected word-form, should matter only for the complex words, which are stored undecomposed. This, in turn, implies that whole-word frequency effects (e.g., shorter RTs to high-frequency word-forms in a LDT) signal that the inflected word is stored in the mental lexicon, and not computed on-line by a symbolic rule. Then, **stem-cluster frequency**, or the combined frequency of all the word-forms with the same stem should reflect the role of the stem. Whole-word frequency effects should signal that the word is stored decomposed, and inflected by symbolic rule computation. To give an English example, whole-word frequency for the inflected verb *walks* is a fraction of the cumulative stem-cluster frequency for this verb, which includes the forms *walk, walks, walked*.

---

9. Information on COBUILD can be found at the website at [http://www.cobuild.collins.co.uk](http://www.cobuild.collins.co.uk).

10. The CELEX database is maintained by the Dutch Center for Lexical Information at the Max Planck Institute for Psycholinguistics, Nijmegen. The reader is referred to the website at [http://www.ru.nl/celex/](http://www.ru.nl/celex/) for further information.

11. Whole-word frequency is sometimes referred to as surface frequency and stem-cluster frequency as lemma frequency.
walked, and walking. A comparison of stem-cluster and whole-word frequencies was used in the study of English past-tense inflection (e.g., Alegre & Gordon 1999).

Type frequency refers to the frequency of occurrence, or in other words, size of a certain class of words in a language, and it is established based on dictionary counts. For example, one can estimate the type frequency of masculine and feminine nouns in French. Importantly, all the words in a class share some common feature(s), e.g., Russian 1st conjugation verbs use the thematic vowel ‘e’ in their non-past paradigm, while 2nd conjugation verbs use the vowel ‘i’. Therefore, type frequency can also refer to a pattern or rule, depending on the theoretical position. This ultimately means that the effects of type frequency in linguistic processing do not support the dual- or single-system approach by themselves unless we can independently demonstrate that the effects refer to a symbolic rule, or alternatively, to an associative pattern.

And finally, frequency of use has had limited application in research on the processing of inflectional morphology. This measure is ‘intermediate’ between token and type frequencies, as it refers to all the uses (‘tokens’) of a certain class of words in speech. This measure can be obtained by recording all the instances of the use of a certain class of words in an observational experiment.12 A similar measure, labeled the number of uses, made it possible to obtain a more realistic assessment of the input to L2 learners of Russian (Chernigovskaya & Gor 2000; Gor & Chernigovskaya 2005). In a population of speakers with limited input and incomplete access to the linguistic system, such as children acquiring L1 or L2 and adult L2 learners, this combined measure is potentially useful. The number of uses was also a basis for the assessment of the rule weight in the Rule Competition Model of child L1 acquisition of English past-tense morphology (Yang 2002).

The role of frequencies in linguistic processing is not restricted to stronger memory traces for high-frequency word-forms, which increase the speed of access to linguistic units stored in the lexicon and their retrieval. Note that this understanding focuses on the role of token frequency; however, the hypothesized role of type frequency is based on somewhat different assumptions. Thus, the strength of association between two word-forms in the lexicon depends on the frequency of activation of the connection between them. It is believed that human speakers store the information about different kinds of linguistic frequencies they encounter. The information about both the frequency of linguistic events and the frequency of combinations of events is hypothesized to be built into the probabilistic mechanisms used both in speech production and reception (Bod et al. 2003). The influence of probabilities associated with type frequency manifests itself in nonce verb generation, when the most frequent pattern has the highest generalization rates (Chernigovskaya & Gor 2000). Thus, input frequencies and probabilities based on them influence linguistic processing and are psycholinguistically interconnected.

The idea that rules have probabilities in young children acquiring their native language, whose linguistic system has not yet stabilized and who operate with several competing grammars, was explored by Charles Yang (2002). Yang proposed the Rule Competition Model of child L1 acquisition of English regular and irregular verbal morphology, which is

12. The study of the American learners of Russian discussed below relied on an estimate based on a simulation using the instructional materials (Chernigovskaya & Gor 2000).
a radical departure both from the dual- and single-system approaches. The new branch of linguistics, Probabilistic Linguistics, explicitly explores the idea that the language faculty is inherently probabilistic and studies frequency effects (Bod et al. 2003). Based on the results of several experiments on the processing of Russian verbal morphology by child and adult L1 speakers as well as adult American learners of Russian (Chernigovskaya & Gor 2000; Gor, 2003, 2004; Gor & Chernigovskaya 2001, 2003a, 2003b, 2005), the Rules and Probabilities Model for both native and non-native processing of complex Russian verbal morphology was proposed (Gor 2003, 2004). This model is similar to Yang’s Rule Competition Model when it addresses L2 morphological processing, however, the idea that rules have probabilities in native adult processing of Russian verbs, both in production and perception, takes the Rule Competition Model one step further. In conclusion, several research agendas point in the same direction, notably, that the categorical distinctions between symbolic rules and the lexicon, between regular and irregular morphological processing, and the strong claim that in linguistic processing, symbolic rules are immune to probabilities need reassessment in the light of new data.

4. Real and nonce verb generation task

Research on several languages, including English, Italian, Russian, German, Norwegian, and Icelandic, with adults, children, and L2 learners, uses the same experimental paradigm, a verb generation task, to study the processing of verbal morphology (Chernigovskaya & Gor 2000; Gor 2003, 2004; Gor & Chernigovskaya 2001, 2003a, 2003b, 2005; Clahsen 1999; Matcovich 1998; Orsolini & Marslen-Wilson 1997; Orsolini, Fanari, & Bowles 1998; Prasada & Pinker 1993; Ragnasdóttir, Simonsen, & Plunkett 1997; Simonsen 2000). The experiments adapt an elicitation technique first developed in a study by Bybee and Slobin (1982): The experimenter works individually with each subject and elicits the expected verb-form by acting out a short quasi-dialog. In some cases, instead of oral presentation, adult subjects receive a written list of sentences and are asked to fill in the blank with the appropriate form of the provided verb in writing (Orsolini & Marslen-Wilson 1997). For obvious reasons, with native adults only nonce verbs are used for the testing material, as with real verbs one would expect ceiling effects in this population of speakers, while with children and L2 speakers, whose internalized verbal system is incomplete, both existing and nonce verbs are used. The aim of such experiments is to study the factors, which determine the generalization rates, or the rates of responses using a particular conjugational pattern. The list of such possible factors includes type frequency, regularity, default, and phonological similarity to the existing verbs, such as the rhyming effect. The following sections will discuss the methodology and the main findings of the studies involving two languages, Italian and Russian.
4.1 Nonce verb generation task with adult L1 speakers of Italian: The role of phonological similarity

The experiment on Italian nonce verb generation explored the effect of two variables on the relative generalizability of inflectional processes: (1) regular versus irregular conjugation class membership, and (2) phonological similarity to the existing Italian verbs (Orsolini & Marslen-Wilson 1997:22).13 This study with adult native speakers of Italian as subjects was designed similarly to an earlier study on English past-tense inflection (Prasada & Pinker 1993), and replicated its task conditions, since the goal was to verify if the results obtained in the English study held for Italian. The original study by Prasada and Pinker was done in the framework of the dual-system approach and tested the hypothesis that in nonce verb generation, the regular -ed rule would be applied regardless of the similarity of the nonce verb to the existing regular verbs. Unlike the regular rule, irregular conjugational patterns will be applied only to the nonce verbs bearing a close resemblance to the existing irregular verbs. And indeed, the obtained results confirmed the hypothesized predictions. Both the English and Italian study explored the influence of phonological similarity on regular and irregular nonce verb generation, and given that in Italian the differences between irregular and regular past-tense inflection are smaller than in English, a less categorical disjunction between regular and irregular inflection was expected for Italian past-tense verbs. The choice of the stimulus material reflects the goals of the study. The nonce verbs created for the experiment belonged to irregular conjugation II (infinitive ending in -ere) and regular conjugation III (infinitive ending in -ire) classes (Orsolini & Marslen-Wilson 1997:22ff.). Three phonological similarity conditions were defined to match the conditions of the Prasada and Pinker study. In Condition 1 (high similarity), the verbs had minor phonological changes in the initial syllable and closely resembled the existing verbs. For example, the nonce verb frendere is very similar to the existing irregular verbs prendere and rendere. In Condition 2 (root similarity) the nonce verbs did not match closely any of the real Italian verbs, however, the initial part of the verb as well as the final root segment resembled the existing Italian verbs. For example, the initial syllable imm-, which is used by verbs like immigr-are and immet-ere, was used for the nonce irregular verb immund-ere. This item has the highly frequent -nd root ending typical of irregular conjugation II verbs. And finally, in Condition 3 (low similarity) the nonce verbs did not use the initial or final segments of any existing Italian verbs, an example of such a verb is maffeere. The likelihood that a certain conjugational pattern is chosen may depend not only on how close the created pseudo-verb is to the existing verbs, but also on how many similar-sounding verbs there are in the language. Thus, to match the study of English past-tense generation (Prasada & Pinker 1993), the “rhyming rate” of pseudo-regular and pseudo-irregular verbs was measured for each similarity condition and comparable rhyming rates were obtained for most Italian stimuli sets. The rhyming rate was defined as the average number of existing verbs, which rhymed.

13. In addition to the nonce verb elicitation task discussed here, the article reports the results of a cross-modal immediate repetition priming task experiment. For lack of space, they will remain outside the scope of this paper. The reader is referred to the original article for details (Orsolini & Marslen-Wilson 1997).
with the novel verbs. The way phonological similarity and the size of the neighborhood (similarly-sounding verbs) are calculated depends on the language in question. Note that, generally speaking, the strength of a neighborhood may be represented as a figure obtained by subtraction of the number of enemies from the number of “friendly” neighbors (see Ullman 1999:56). The enemies are similarly-sounding verbs with a different conjugational pattern, or in case of English past-tense inflection, with a different stem-past mapping, e.g., for the stem-past pair sing-sang, ring-rang would be a friendly neighbor, while bring-brought would be an enemy. This type of computation was used by Michael Ullman in an experiment, which studied the influence of phonological similarity on acceptability ratings of English regular and irregular stems and past-tense forms by adult native speakers (Ullman 1999).

An equal number of pseudo-irregular (ending in -ere) and pseudo-regular (ending in -ire) verbs were created, 20 in each of the three phonological similarity conditions, thus the total number of experimental stimuli was 120. The stimuli were presented as infinitives with fictitious definitions, e.g., the nonce verb effadere was defined as “to cook nervously.” This study used the written format for data collection, in which the subjects had to fill in the blanks in the sentences by generating the past participle of the provided verb. Here is an example of such a sentence, where the verb effadere was to be used: *The fish soup was not very tasty because Maria had ________ and had forgotten to put salt in it.*

20 adult native speakers of Italian, paid volunteers, took part in the experiment. The subjects’ responses were further categorized as regularizations (i.e., the past participles produced using regular conjugation III and II\(^{14}\) thematic vowels and suffixes), hyperregularizations (the forms with a default conjugation I vowel and suffix in response to conjugation II and III verbal stimuli), and irregularizations (the forms with stem changes and irregular suffixes). Based on the obtained results, the study arrived at the conclusion that, indeed, the sharp dichotomy between regular and irregular mechanisms of morphological processing postulated by the dual-system theory cannot handle the data on Italian verb generation. First, the rate of irregularizations, or the use of the irregular conjugation II pattern was much higher for Italian than for English. Irregular conjugational patterns were generalized at a high rate even in the root similarity condition, where there was no close similarity to the prototype. And second, regular conjugation III inflection in subjects’ responses exhibited the influence of phonological similarity, which should not have been present in regular verb generation. The study failed to support the single-system account either, thus leaving the answer about the type of processing observed in Italian verb generation open. One important contribution of this study to the dual- versus single-system debates and our understanding of the mechanisms of morphological processing is that it demonstrates that the study of one single language is insufficient to support any universal claim about linguistic processing, and that cross-linguistic data are in order.

\(^{14}\) It should be noted, that not all conjugation II verbs are irregular in Italian.
4.2 Real and nonce verb generation task with adult L1 and L2 speakers of Russian and L1 children: The role of type frequency

While the Italian study discussed above dealt with the role of phonological similarity, the experiments on Russian verb generation focused on the interaction of regularity and type frequency, or class size. They addressed the opposite claims about the role of frequency in regular versus irregular inflection put forward by the proponents of the dual- and single-system approaches. According to the dual-system account, regular inflection, unlike irregular inflection, is unaffected by input frequencies. The single-system account maintains that both regular and irregular inflection are influenced by input frequencies. A series of experiments on Russian verb generation has prompted the Rules and Probabilities Model (Gor 2003, 2004), which challenges both the dual- and single-system approaches. The model is based on the following claims: (1) In languages with numerous inflectional patterns regularity can become a gradual parameter, and as a result, a categorical distinction between regular and irregular inflection is unjustified. (2) In the processing of regular inflectional morphology, speakers make use of abstract rules. (3) Rule application is influenced by the probabilities based on linguistic frequencies. In other words, the data on several populations of L1 and L2 speakers of Russian suggest that regular inflection is influenced by input frequencies, but it is rule-based rather than associative memory-based. The following sections will describe the experiments, which demonstrated the role of input frequencies and probabilities in regular morphological processing. Another series of experiments, not discussed here (Chernigovskaya & Gor 2000; Gor & Chernigovskaya 2001), showed that regular morphological processing in Russian relies on rule application rather than associative patterning (Gor 2004). The next section will provide the basic information about Russian verb conjugation and different kinds of frequencies used in designing the experiments with native and non-native speakers of Russian. It will exemplify the kind of issues the researcher deals with when working with material on an individual language.

4.2.1 Russian verbal morphology and the resources on Russian verb frequencies

The debates on the processing of English past-tense morphology early on identified a major ‘weakness’ of the English data, notably, the impossibility to separate regularity, productivity, high type frequency, and the default. Indeed, 96% of English verbs are regular and productive, while only 4% are irregular and unproductive. Another feature of the English past tense is that it has only one regular conjugation rule, and therefore regular and default processing are merged together. These limitations led researchers to look at other languages with rich inflectional morphology. As it turns out, Russian has strong advantages over English in that it has numerous conjugation classes differing in the degree of regularity and size. The on-going collaborative project on the processing of Russian verbal morphology by different groups of L1 and L2 speakers, children and adults, as well as children and adults with language disabilities, conducted at the University of Maryland, USA

---

15. For a more detailed account of the verb generation task with L1 children and L2 adults see Gor and Chernigovskaya, 2003, 2004.
and Saint Petersburg University, Russia capitalized on these features of the Russian conjugational system. This section will discuss the major features of the Russian conjugational system necessary in order to follow the experimental design and evaluate the proposed argumentation. It will also review the resources on verb frequencies used in the study.

According to the one-stem verb system created by Roman Jakobson (Jakobson 1948), Russian has 11 verb classes, each with its own suffix, or verbal classifier. The eleventh class has a zero suffix, and is subdivided into smaller subclasses depending on the quality of the root-final consonant. This is a small class, especially given the variety of conjugational patterns it includes, and there are well under 100 basic stems in it (Townsend 1975). The conjugational patterns of some of the sub-classes of the non-suffixed stems have idiosyncratic features, and thus form verb clusters, which can be compared to the neighborhoods of English irregular verbs, or alternatively, characterized by the minor rules. The remaining 10 suffixed classes are identified by the suffix: -aj-, -ej-, -a-, -e-, -i-, -o-, -ova-, -avaj-, -nu-, and -zha-. The suffix determines all the parameters of the conjugational paradigm, including the choice of the thematic vowel in the inflections (-e- for 1st and -i- for 2nd conjugation), and different types of stem changes (see Table 1). When the endings are added to the stem (which includes the optional prefix, the root, and the suffix), an automatic truncation rule works at the juncture of the stem and the ending. If the stem ends in a vowel and the ending begins in a vowel, the first vowel is truncated. The same is true for the consonants: the first one is deleted. Past tense and infinitive endings begin with a consonant, and non-past tense endings begin with a vowel, therefore, stem-final vowels will be deleted in non-past tense forms, and consonants will be deleted in past tense and infinitive forms. The examples below illustrate the automatic truncation processes taking place when the inflections are added to the stems in two verbs classes, the -aj- and -a- classes, used in the experiments.

Table 1. Non-automatic morphological processes in the stems included in the experiments

<table>
<thead>
<tr>
<th>Verb classes and sub-classes</th>
<th>-aj-</th>
<th>-a-</th>
<th>-ej-</th>
<th>-e-</th>
<th>(i)j-*</th>
<th>-i-</th>
<th>-ova-</th>
<th>-avaj-</th>
<th>(o)j-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consonant</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stress shift17</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Suffix alternation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vowel alternation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vowel deletion</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The (i)j- and (o)j- stems are the sub-classes of zero-suffixed stems, each of them has an idiosyncratic ("irregular") feature in the conjugational pattern.

16. The consonant “zh” represents any palatal consonant – a hushing or “j” – and is not part of the suffix.

17. This paper does not discuss stress shifts in the obtained data. Unlike all the other morphological processes that can be predicted given the verb stem, stress shifts need to be lexically encoded. However, the pattern of stress shift within the paradigm is fixed.
gul’-aj- + t’ = gul’at’ (infinitive, “walk”)
gul’-aj- + u = gul’aju (1st person sg, non-past tense, “I walk”)
pis-a- + t’ = pisat’ (infinitive, “write”)
pis-a- + u = pisju (1st person sg, non-past tense, “I write”)

Note that the infinitive does not contain the information about the verb class, because the \( j \) of the suffix may have been truncated and is unrecoverable. This means, that a novel infinitive ending in \(-at’\) may be interpreted either as an \(-aj-\) or \(-a-\) verb. This property of the Russian verbal system makes it possible to research the role of probabilities (and phonological similarity as well) in verb generation. Also, note that in th \(-a-\) stem non-automatic consonant mutation takes place in addition to automatic truncation.

Overall, the Russian verbal system possesses the following features:

– Numerous verb classes varying in size;
– Developed conjugational paradigm;
– No sharp division between regular and irregular classes;
– Several regular classes in addition to default;
– Infinitives of many verb classes have unrecoverable stems due to the truncation of the stem-final consonant before consonantal endings.

Table 1 lists the morphological processes shaping the conjugational patterns of the stems chosen for the experiments. It does not include automatic consonant or vowel truncation, which occurs at the juncture of the stem and the ending.

Table 2 contains information about the type frequencies of the stems included in the experiment and about their productivity. Since no such data were available, we performed the computations ourselves on the Grammatical Dictionary of the Russian Language (Zalizniak 1980) with approximately 100,000 entries. In the first row of Table 2 corresponding to the Russian language, the first set of numbers in each column represents the results of our verb counts. These counts contain all the verbs belonging to a particular conjugation class and include the prefixed verbs.\(^{18}\) For the small unproductive classes, we also provide the number of unprefixed basic stems below (based on Townsend 1975; and Davidson et al. 1996). The second and third rows contain two kinds of data on input frequencies specifically obtained for L2 learners taking part in the verb generation experiment. If we assume that the type frequencies found in the Russian language are available to adult native speakers through their linguistic experience, the situation with th L2 learners is much more complicated, since it is impossible to exactly estimate the input they have received during their exposure to L2. Therefore, we decided to focus on the L2 learners, for whom we could produce the most accurate estimates of the input frequencies, the American students of Russian who had completed one year of study of Russian at the University of Maryland. We simulated all the language activities in which they were engaged, in class and at home, based on the set of instructional materials they were using. Thus, the estimates assume that the learners completed all the assignments, which was certainly not

\(^{18}\) Natalia Sliusar from the Department of General Linguistics at St. Petersburg State University performed these computations.
Table 2. Type frequency of the verb classes included in the experiments: Native and second language input

<table>
<thead>
<tr>
<th>Verb classes</th>
<th>-aj-</th>
<th>-a-</th>
<th>-ej-</th>
<th>-e-</th>
<th>(ij)-</th>
<th>-i-</th>
<th>-ova-</th>
<th>-avaj-</th>
<th>(o)j-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russian</td>
<td>11814</td>
<td>940</td>
<td>608</td>
<td>328</td>
<td>160</td>
<td>7019</td>
<td>2816</td>
<td>94</td>
<td>98</td>
</tr>
<tr>
<td>Type frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Input to L2 learners</td>
<td>55 (86)</td>
<td>14 (24)</td>
<td>0 (4)</td>
<td>8 (12)</td>
<td>3 (3)</td>
<td>52 (80)</td>
<td>13 (34)</td>
<td>2 (7)</td>
<td>2 (5)</td>
</tr>
</tbody>
</table>

19. Strictly speaking, the term "L2 input" in this case is inaccurate, since the assessment of L2 input frequencies included both what the learners heard and produced themselves, in other words, both the input and the output.

20. The first figure corresponds to the number of verbs in the active vocabulary, and the second figure (in parentheses) to all the verbs from the active and passive vocabulary combined.

21. Both these corpora are available at the following URL: http://www.sfb441.uni-tuebingen.de/b1/en/korpora.html. The description of the Uppsala Corpus can be found in the Frequency Dictionary of Modern Russian (Lönngren 1993), which is actually based on this corpus. The description of the Tübingen corpus exists only in the electronic version, and can be accessed at: http://www.sfb441.uni-tuebingen.de/b1/en/korpora.html#interview.

exactly true, however, it was a good approximation of their linguistic input, because the curriculum expects them to follow the textbook very closely. Naturally, in more advanced speakers, who have more exposure to the L2 outside the classroom and the textbooks, such estimates become less accurate.

At the time of the experiments, there existed two searchable internet-based Russian corpora, the Uppsala Corpus and the Tübingen Corpus, both available for on-line query at the site of the University of Tübingen. The Uppsala Corpus contains 1 million words and comprises literary texts, press, and scientific texts from the second half of the 20th century. The Tübingen Corpus contains only 600,000 words, however, it is very valuable, since it contains only the oral interview data (presumably edited) — these are the interviews published in Russian press of the end of the 20th century. According to the estimates of Lönngren, a corpus of 1 million words is reasonably representative for the words with the frequency of 80 per million and higher (Lönngren 1993), but not for lower frequency ones. The list of such words is published in the Frequency Dictionary and includes 213 verbs (Lönngren 1993). We used the data on these 213 verbs, and also performed on-line queries to select the experimental material for our lexical decision task experiment.
At present, a new representative internet-based corpus of Russian is used to obtain both stem-cluster and whole-word frequencies (Sharoff 2002).22

4.2.2 Material and method
The series of experiments on Russian real and nonce verb generation by adult native and non-native speakers of Russian and Russian children compared the rates of stem recognition for the real and nonce verbs belonging to four different verb classes, the -aj-, -a-, -i-, and -ova- (see Tables 1 and 2 for the conjugational features and type frequencies of these classes).23 Stem recognition, as a measure, does not count all the smaller errors in the application of the conjugational pattern, it is only concerned with the choice of one conjugational pattern over the other(s), e.g., -aj- versus -a-, etc. The -aj- class is the regular high type frequency default class representing the ‘Vowel + J’ conjugational pattern, which is characterized by no irregular morphophonemic processes, only the truncation of the ‘J’ before consonantal endings.24 The -a- and -i- classes both represent the ‘Vowel + Ø’ pattern, which is characterized by some irregularities, such as consonant mutations and stress shifts.25 At the same time, the -i- class is high type frequency productive, while the -a- class is low type frequency unproductive. The -ova- stem is regular, in a sense that all the forms in the paradigm are predictable, and productive, but it has a rare feature – suffix alternation – in its paradigm.

The token frequencies, or more exactly, stem-cluster frequencies of the verbs included in the testing material, which reflect the frequency of the stem in all the forms of a particular verb that occurred in the database, were obtained from The Frequency Dictionary of Russian Language by Zasorina (1977). This dictionary contains approximately 40,000 words and is based on a 1,000,000-word corpus of written Russian language including fiction, scientific texts, and newspaper and journal articles. The real verbs chosen for the experimental material belonged to two frequency ranges: very high-frequency and high-frequency, and the nonce verbs were created by manipulating the initial segments of the very high-frequency verbs (see Table 3). Note that the labels ‘very high-frequency’ and ‘high-frequency’ are relative, with the ‘high-frequency’ category corresponding rather to the mid-frequency range. The choice of the frequency ranges was dictated by the following consideration: For L1 children, both very high-frequency and high-frequency ranges should include the familiar verbs; for beginning L2 learners, only very high-frequency verbs should be familiar. The verbs were balanced in average token frequency across the classes. This produced 60 stimuli, to which 20 fillers were added, in two different versions of the experiment these were either an additional set of 20 nonce verbs, which paralleled


23. In fact, these experiments are part of the on-going project, which includes other populations of speakers, such as children with language disabilities, patients with aphasia, and now schizophrenic patients. For the comparative data on normally developing children and children with language disabilities see Gor and Chernigovskaya 2004.

24. In other words, its suffix includes a combination of one of the vowels and a ‘J’.

25. In the ‘Vowel + Ø’ pattern, a single vowel constitutes the suffix.
Table 3. The testing material for the verb generation task

<table>
<thead>
<tr>
<th>Verb Class/Infinitive/Basic Stem/Gloss</th>
<th>Very High-Frequency</th>
<th>High-Frequency</th>
<th>N nonce</th>
</tr>
</thead>
<tbody>
<tr>
<td>-aj- 'gul’at’ (gul’-aj-), &quot;to walk&quot;</td>
<td>N = 5</td>
<td>N = 5</td>
<td>N = 5</td>
</tr>
<tr>
<td></td>
<td>Average Fr. 173.2</td>
<td>Average Fr. 10.8</td>
<td></td>
</tr>
<tr>
<td>-a- 'pisat’ (pis-a-), &quot;to write&quot;</td>
<td>N = 5</td>
<td>N = 5</td>
<td>N = 5</td>
</tr>
<tr>
<td></td>
<td>Average Fr. 174.4</td>
<td>Average Fr. 13.4</td>
<td></td>
</tr>
<tr>
<td>-i- 'nosit’ (nos-i-), &quot;to carry&quot;</td>
<td>N = 5</td>
<td>N = 5</td>
<td>N = 5</td>
</tr>
<tr>
<td></td>
<td>Average Fr. 170.2</td>
<td>Average Fr. 11.0</td>
<td></td>
</tr>
<tr>
<td>-ova- 'probovat’</td>
<td>N = 5</td>
<td>N = 5</td>
<td>N = 5</td>
</tr>
<tr>
<td>(prob-ova-), &quot;to try&quot;</td>
<td>Average Fr. 149.8</td>
<td>Average Fr. 11.0</td>
<td></td>
</tr>
</tbody>
</table>

The high-frequency verbs in the material, or 20 verbs belonging to other verb classes: -e-, -ej-, -avaj-, (i)j-, and (o)j-. The inclusion of a certain number of distractors from five more classes complicated the subjects’ task by making the choice of the appropriate conjugational pattern less obvious and thus preventing the subjects from quickly developing response strategies. Two different versions of the experiment were created. In the first one the initial stimulus verb was presented to the subject in its infinitive form, while in the second one it was presented in the past-tense plural form. The subjects were required to generate the non-past 3rd person plural and 1st person singular forms of the verbal stimuli. All the verbs were embedded in simple carrying sentences, which together with follow-up questions formed a quasi-dialogue:

**Infinitive**

Experimenter: I want to ________.
Subject: Me too, I want to ________.
Experimenter: And what are you doing today?
Subject: Today I ________. ²⁶
Experimenter: And Mary and Peter?
Subject: Today they ________.

**Past Tense**

Experimenter: Yesterday they ________, And what are they doing today?
Subject: Today they ________.
Experimenter: And you?
Subject: Today I ________.

The task was performed orally, however, the adult subjects were provided the written script for each mini-dialog. Audio recordings of the sessions were further analyzed to establish the stem recognition and correct response rates. The verbs were presented in a quasi-random order with no two verbs belonging to the same stem following each other. Half of the subjects received the version of the text with all the stimuli presented in an inverse order to control for the possible effects of the order of presentation, such as fatigue or habituation to the task.
The subjects for this set of experiments were:
Experiment 1A – 15 adult native speakers of Russian.\textsuperscript{27}
Experiment 1B – 37 volunteer students at the University of Maryland at the end of
their second semester of Russian in two groups of 20 and 17 subjects.\textsuperscript{28}
Experiment 1C – 20 Russian children with normal language and cognitive develop-
ment, and no hearing problems. There were 5 children aged 4, 9 children aged 5, and 6
children aged 6 in the group of subjects.

4.2.3 Results and discussion
This section will compare and discuss the data obtained in all the three experiments,
1A, 1B, and 1C. The results obtained for native Russian adults and American learners
of Russian are represented in Figure 1.\textsuperscript{29}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Rates of stem recognition in a verb generation task with adult L1 and L2 speakers (infini-
tive stimuli)}
\end{figure}

\textsuperscript{27} The data on adult L1 speakers were analyzed by Megan Malinowski and became part of her
Master’s Thesis.

\textsuperscript{28} Since the groups of 20 and 17 subjects were tested on different versions of the test, their data were
not pooled together, but were compared to the matching data – 20 subjects were compared with 20 Rus-
sian children, while 17 subjects were compared with 15 Russian adults. The data on 17 L2 learners were
analyzed by Natalia Romanova and became part of her Master’s Thesis.

\textsuperscript{29} Since not all the adult and child L1 learners completed both versions of the test with the past-
tense and infinitive stimuli, the comparisons will be made for the available data sets, either past tense
or infinitive.
The rates of stem recognition for native Russian adults were predictably much higher than for L2 learners. Their data show the ceiling effect with the real infinitive stimuli, both very high- and high-frequency. In the averaged data, the rate of stem recognition was the lowest for the -a- class, since the infinitive is ambiguous due to stem consonant truncation, and the stem can be attributed either to the default -aj- class or the -a- class. The -i- class, which has a conflicting status – it is a high type frequency productive class, but has some irregularities in its conjugation – also caused some problems. The American learners, for whom most of the verbs, even some very high-frequency ones, were novel, showed much lower stem recognition rates on all the classes, except for the -i- class. What is remarkable, however, is that L2 learners were not able to rely on the default pattern when dealing with the 'ambiguous' infinitives of the -aj- and -a- classes. They chose one or the other pattern roughly equally often. This distribution reflects the composition of the experimental material – equal proportion of the -aj- and -a- verbs – rather than the type frequencies found in Russian.

Obviously, the comparison of the data on real and nonce verbs pooled together masks the results on novel verb generation, hence the necessity to isolate the nonce verb generation data (see Figure 2). Remarkably, while native speakers dealt with the novel -aj- and -a- verbs as expected based on type frequencies for these classes – they gave strong preference to the -aj- pattern – L2 learners did not show a strong preference, and overall, their stem recognition rates for these two 'ambiguous' stems were low. At the same time, they surpassed native speakers in processing the -i- verbs. This, again, shows that they were less influenced by the default 'Vowel + j' conjugational pattern. The -ova- class with suffix alternation was predictably more often recognized by L1 than L2 speakers given that the latter ones did not have sufficient exposure to that class.

Figure 2. Rates of stem recognition for nonce verbs in L1 and L2 adult speakers (infinitive stimuli)
First and second language morphological processing

Figure 3. Rates of stem recognition for nonce verbs in L1 children and L2 adults (past-tense stimuli)

Figure 3 compares the data on verb generation in adult L2 learners and native Russian children broken down by the age. First, it clearly demonstrates a developmental tendency in L1 children, notably, they strongly rely on the default -aj- pattern at age 4, and then gradually move away from the default to the ‘Vowel + Ø’ pattern. Also, at age 5 the -ova-conjugal pattern stabilizes. The comparison of the child and adult L2 data demonstrates that L2 learners do not fit with any of the age groups: they perform like 4-year-olds on the -aj- verbs, like 6-year-olds on the -a- verbs, and outperform all the age groups on the -i- and -ova- verbs.

4.2.4 Conclusion
In conclusion, the verb generation experiment showed that the knowledge of probabilities influenced the processing of regular verbal morphology in all the three populations of speakers – L1 adults and children, as well as L2 adults. The generalization rates in novel verb generation depended on the type frequency, as well as the degree of regularity for different conjugal patterns. When dealing with ambiguous verbs, all the speakers preferred the high type frequency default conjugal pattern. At the same time, the
processing of verbal morphology in L2 learners differed in substantial ways from that of adult native speakers – L2 learners were less biased towards the default than L1 speakers. One of the major differences between the type frequencies available to adult L1 and L2 speakers is the leveling of the size of verb classes in L2 compared to the full verbal system. For example, the the -aj- class is roughly 12 times larger than the -a- class in Russian language, but only 3 times larger in the input to L2 learners (see Table 2). Therefore, the fact that L2 learners relied on default processing less than native speakers can be attributed in part to the reduced verbal system that is available to them. The study demonstrated that beginning L2 learners do not use the same probabilities as native speakers. Also, L2 rates of stem recognition did not match any of the child L1 age groups. At the same time, L1 children demonstrated a clear developmental tendency of moving away from the default pattern, which is acquired first. The Rules and Probabilities Model (Gor 2003, 2004) took into account these results, and also the results of another set of experiments with adult L1 and L2 speakers of Russian (Chernigovskaya & Gor 2000). The model claimed that both in L1 and L2 acquisition, and, more importantly, in endstate adult L1 processing of inflectional morphology, rules are applied with varying degrees of probabilities, the latter depending on input frequencies. The important part of the argumentation is the claim that the underlying mechanism in novel verb generation is based on rules and not phonological associations. With regard to this set of native and non-native data, one can argue that all the speakers used the regular default ‘Vowel + j’ rule with different probabilities depending on the suffix vowel, the ‘a’ versus the ‘i’, in accordance with the type frequencies of the verb classes. In this sense, rules and probabilities interacted in adult and child L1 speakers and L2 learners. Certainly, only native adults could make use of the probabilities associated with the full Russian verbal system.

5. Lexical decision task

5.1 Lexical decision task with adult native speakers of English

Research on the processing of verbal morphology in English has been concerned with the question whether regularly inflected verbs are stored decomposed or in their inflected form. The most representative study, in fact, demonstrated that English regularly inflected verbs with the frequency above 6 per million are stored undecomposed, which deals a considerable blow to the claim of the dual-system approach that all regular verbs are inflected by symbolic rule computation, since the 6 per million threshold puts even less frequent inflected verb-forms on the list of stored items in English (Alegre & Gordon 1999). This comprehensive study had a broader focus and included a series of five lexical decision task (LDT) experiments investigating the effect of whole-word frequency of different classes of inflected words, not just past-tense verb forms, on reaction times (RTs). Thus, the inflected items included in the experiments consisted of plural nouns (spheres, cellars, etc.), and different inflected verb forms (pretended, yelling, conveys) with adjectives (arrogant,
exquisite) representing uninflected words. As discussed above, this research paradigm is based on the assumption that sensitivity of RTs to whole-word or stem-cluster frequency can signal whether the inflected word is stored whole or decomposed into morphemes. Whole-word frequencies will influence RTs if the inflected word is stored undecomposed. Stem-cluster frequencies will influence RTs if the word is stored decomposed, and the frequencies of all the inflected forms carrying the same stem are added together. Then, if the words are matched for stem-cluster frequency, but differ in whole-word frequency, and these differences have an effect on RTs in a LDT, one can argue that the inflected words from the experimental set are stored undecomposed. The experiments treated frequency as a continuous rather than a categorical variable and compared RTs to inflected and matched uninflected words. The experimental material in each of the 5 experiments was similar in composition: the target stimuli constituted one fourth of the testing material with fillers constituting three fourths of the material. For example, in Experiment 1, 66 experimental regularly inflected items were matched with 66 simple (uninflected) words, 66 “inflected” nonwords, and 66 simple nonwords (Alegre & Gordon 1999:45–46). The experimental procedure also remained the same across the five experiments. The LDT was programmed and performed using the Micro Experimental Lab (MEL) software on a PC environment. The main experiment was preceded by several practice trials; in the main part of the test, all the experimental and filler items were presented in a different random order to each subject. The participants received the instruction to press the “1” key on the keyboard if the stimulus appearing on the screen was a word, and “2” if it was a nonword. The stimuli were preceded by a fixation point, an asterisk, in the middle of the screen, and then appeared centered around the asterisk position for 1.5 seconds (s) or until the subject responded. All the slow answers which exceeded 1.5 s were discarded from further analyses. The experiments provided feedback to the participants, and when the answer was incorrect, the message “wrong answer” appeared on the screen.

The main part of experimental design was the establishment of stem-cluster and whole-word frequency ranges for each experiment depending on the hypothesis it tested, and finding the stimuli that belonged to the required range based on Francis and Kucera (1982). A caveat is in order when using this research method: Not only is it extremely difficult to find the verbs matched in stem-cluster frequency, but considerably differing in the whole-word frequency of a specific form in the paradigm, it is not always clear, which forms should be included in the count for stem-cluster frequency. For example, in Russian, a language with a developed conjugational paradigm, should one include all the four participles, two gerunds, and two imperative forms in addition to the non-past tense and past-tense forms in the verb’s stem-cluster frequency count? The decision will considerably affect the frequency range. Despite all these methodological difficulties,

31. The terms nonce word, pseudo-word, and nonword used by different researchers refer to the same type of word, an artificially created word with no meaning, which has some of the formal properties of existing words.

32. Note that even for English the issue of how to calculate verb stem-cluster frequencies is far from being transparent. For example, a study, which measured RTs in a verb generation task, did not include
this research paradigm generates meaningful results. Thanks to a careful selection of the frequency ranges, and control for stem-cluster frequency of the stimuli varying in whole-word frequency, the Alegre and Gordon (1999) study produced the results necessary for the understanding of storage and retrieval of morphologically complex regular items. Some of their findings have direct bearing on the polemics surrounding English past-tense inflection. Thus, the study reliably demonstrated whole-word frequency effects above the threshold of 6 per million in regularly inflected words, which can be taken as an indication that they are stored undecomposed, as inflected word-forms, and not assembled on-line by the application of a symbolic rule, as predicted by the dual-system account.

5.2 Lexical decision task with adult native speakers of Russian

The computation of stem-cluster frequency remains a theoretical problem for Russian, a language with rich verbal morphology, both inflectional and derivational. Thus, it is not clear that all the participles and gerunds should be included in stem-cluster counts. However, in designing the LDT, we decided to pursue a different issue, and thus were able to avoid this methodological difficulty. The goal of the experiment was to prove that type frequency, or the frequency of the conjugational pattern, also affects RTs in a LDT. Combined with the argument that the conjugational pattern in question represents a symbolic rule rather than an analogy based on phonological similarity, this result would support the claim that rules have probabilities in adult native processing. In a series of previous studies we explored the role of type frequency as the frequency of the verb class (Chernigovskaya & Gor 2000; Gor & Chernigovskaya 2005). In the LDT, we wanted to test another hypothesis that different forms in the paradigm, which involve the use of a similar rule in terms of stem allomorphy, or in other words, display the same degree of regularity, but with different inflections, will produce different latencies in a LDT depending on the frequency of the form in the paradigm. Our analysis of the frequencies of the 213 most frequent Russian verbs confirmed the commonly known fact that the 3rd person singular verb form is the most frequent of the non-past tense forms, while the 2nd person singular is one of the least frequent forms. The infinitive has the frequency similar to the one found in the 3rd person singular, however, it is hypothesized to have a special status in terms of storage and retrieval, since it is the ‘citation’ form, and therefore, could have a faster access speed. Table 4 demonstrates that the distribution of frequencies discussed above holds for both corpora we were using.

The purpose of the LDT was to measure the speed of recognition of visually presented inflected verbs. The experiment tested the hypothesis that RTs depend on the type frequency of the forms in the conjugational paradigm. The subjects in the experiment see the words appearing on the screen and need to press two different buttons on the the frequencies of regular past-tense forms in the stem-cluster frequency counts. This was done on the grounds that the frequencies of irregular past-tense verbs are not included in their stem-cluster frequency due to stem allomorphy, and the approach should be consistent for all the verbs. However, this led to a paradoxical situation when past-tense frequency could be higher than stem-cluster frequency (Prasada et al. 1990).
Table 4. Frequency distribution of the finite verb forms in the Uppsala and Tübingen Corpora (percentages based on the average for 213 most frequent verbs)

<table>
<thead>
<tr>
<th>Number</th>
<th>Infinitive</th>
<th>Non-Past Tense</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Singular</td>
<td>Plural</td>
</tr>
<tr>
<td>Person</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uppsala Corpus</td>
<td>18.8</td>
<td>2.5</td>
</tr>
<tr>
<td>Tübingen Corpus</td>
<td>20.4</td>
<td>6.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number</th>
<th>Past Tense</th>
<th>Imperative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Masculine</td>
<td>Feminine</td>
</tr>
<tr>
<td>Uppsala Corpus</td>
<td>20.9</td>
<td>10.1</td>
</tr>
<tr>
<td>Tübingen Corpus</td>
<td>14.6</td>
<td>7.2</td>
</tr>
</tbody>
</table>

button box or keys on the keyboard depending on whether they think that the word is a real word or non-word. For morphologically simple undecomposable words the speed of access would reflect their frequency, more frequent words will be accessed faster. For morphologically complex words the RT will reflect its whole-word frequency if it is accessed undecomposed, however, if it is decomposed in order to be accessed, the speed of decomposition would depend both on the frequency of the stem, which will be controlled by using the same verb in several inflected forms, and the frequency of the inflection rule or pattern. In order to make the LDT task reasonably challenging for the subjects, the experimenter needs to make the testing material sufficiently varied by including not only a matching number of non-words, but also numerous word and non-word fillers. Since the purpose of a LDT is to measure the latency of the first reaction to the visual verbal stimulus, the experiment needs to be fast-paced. This paper will discuss the methodology of the first version of the LDT experiment, LDT1, which was designed on a Mac platform using PsyScope software. A more detailed account of this experiment can be found in Gor 2003. Many thanks to Natalia Sliusar for help with the preparation of the testing material and design of the experiment with PsyScope.
-i- and -e- classes have the same conjugational patterns involving the stem allomorphy similar to the one found in the -a- class, but distributed differently across the paradigm; however, the -i- class is large and productive, while the -e- class is small and unproductive. And finally, the -ova- class is large and productive, and has suffix alternation in its paradigm. According to the dual-system approach criteria, the -aj-, -i-, and -ova- classes are regular, because they are productive. The high-frequency verbs belonged to the 10–25 per million frequency range as found in the Frequency Dictionary (Lönngren 1993) based on the Uppsala Corpus (this count reflects the stem-cluster frequency), while the low-frequency verbs were not found in the dictionary at all, and therefore had the stem-cluster frequency of less than 1 per million. We used three forms of each verb, the infinitive, 2nd person singular and 3rd person singular, and a set of matching nonce verbs, created by manipulating the initial segments of the real verbs used in the experiment, usually, the initial consonant. Thus, we included 150 real (5 verbs × 3 forms × 5 classes × 2 frequency ranges) and 150 matching nonce verbs. Since it is impossible to match the verbs belonging to the five classes exactly in length given that the length of their suffixes differs, the experiment controlled for the following factors: all the verbs were prefixless and belonged to the imperfective aspect. Their length varied from 2 to 3 syllables and from 7 to 9 letters. The fillers were 15 high-frequency and 15 low-frequency nouns matched in stem-cluster frequency and length with the verbs used in three cases, the Nominative singular (the citation form), Dative plural, and Instrumental singular. For each of the real nouns a matching nonce noun was created. The number of nouns was 90 (15 nouns × 3 forms × 2 frequency ranges) plus 90 matching nonce nouns, thus bringing the total number of stimuli in the experiment to 480.

5.2.2 Method
The presentation of the stimuli to the subjects and the measurement and recording of the RTs were controlled using PsyScope, a software designed for the Mac environment. PsyScope eliminates the programming stage thus allowing the researcher to work within the graphic environment, and visually represents the experimental design. It was developed by a team of researchers at Carnegie Mellon University and made available for research purposes through the site at http://psyscope.psy.cmu.edu/. One of the factors we needed to control for was theumber of repetitions of the stem. Remember that the subject was exposed to three different word-forms of the same verb or noun, and the RTs can be expected to decrease with each repetition of the same lexeme within the same experiment. This is a well-attested effect observed in repetition priming studies. We dealt with this problem in the following way: The total number of 480 stimuli was divided in three

---

34. As already noted above, the labels for frequency ranges do not carry any absolute values, in fact, the high-frequency category includes the mid-frequency range, while the low-frequency category includes very low-frequency verbs.

35. A detailed manual can be downloaded from this site, which takes the beginning researcher through all the necessary steps to set up the experiment. At present, PsyScope does not work with Mac OSX, but it is currently being updated to be fully compatible with the new operating system.
blocks with 160 items in each, and the material was balanced across the blocks with the form of each word occurring only once within the block. Then, we prepared six different versions of the test with all the possible sequences of the blocks, and presented a different version of the test to each of the participants in groups of six. With this design, the repetition effects remained present for each individual subject, but were counterbalanced across the subjects. A practice session with 12 stimuli preceded the main three blocks. The stimuli appeared on the screen in random order and disappeared after 600 milliseconds (ms). The subject pressed one of the buttons on the button box for a word and another for a nonword, at which moment the inter-stimulus interval was initiated.36 This interval varied in duration between 700 and 1100 ms to avoid the effect of habituation to the fixed interval when the subject shows a better readiness towards the end of the trial. If the subject failed to press any button, after 6000 ms the inter-stimulus interval was initiated automatically. PsyScope records the RTs and the type of response (the word/nonword decision) in a chart that is easily imported into another program, such as Excel or SPSS to be further analyzed. The first pilot series was conducted with 8 volunteers, adult native speakers of Russian.

5.2.3 Results and discussion
The RTs of each of the 8 subjects were averaged to see if there were any outliers in the group, and indeed, the RTs of one participant were considerably longer than those of the other participants (see Figure 4), and the data for this participant were eliminated from further analyses.

The first result of this pilot study was the confirmation of the expected effect of word frequency on response latencies, since indeed, the high-frequency verbs produced shorter RTs than the low-frequency verbs across all the five verb classes (see Figure 5).

![Figure 4. Average RTs of individual participants on a lexical decision task](image)

36. The use of the button box makes it possible to measure RTs with the accuracy of up to 1 ms.
Figure 5. Reaction times to high-frequency, low-frequency, and nonce verbs

The role of type frequency was most evident in high-frequency verbs, where the difference between the latencies found in the 2nd person singular, 3rd person singular, and the infinitive is the greatest (see Figure 6).

Given the frequency range used in the experiment, one can expect type frequency to affect the RTs to a lesser degree in low-frequency verbs where all the forms in the paradigm have the frequency of less than 1 per million. In nonce verbs, the infinitive required slightly more time to process than the 3rd person singular form, but the difference between the 2nd person singular and the 3rd person singular was maintained. At the same time, the LDT1 experiment showed no effect of the verb class on RTs in visual perception of inflected verbs.

This experiment was replicated at Saint Petersburg State University, Russia using Presentation software on a PC platform (Sliusar 2003). In the modified version, LDT2, the testing material was the same as in LDT1. A group of 28 adult native speakers of Russian took part in this experiment, 16 women and 12 men aged 15–50. The analysis of variance demonstrated that for all the stimuli combined (high-, low-frequency, and nonce

37 The modified LDT2 experiment was part of the Master’s Thesis by Natalia Sliusar.
verbs), four out of the five tested factors are statistically significant – frequency, number of repetitions, form in the paradigm, and length in letters (Sliusar 2003:71 ff.). The verb class factor was again not significant. The analysis of the three frequency ranges separately revealed that the number of repetitions was a highly important factor in the nonce verb category, and to a lesser degree, in the low-frequency category. The length in letters factor emerged only in the low-frequency category, while the form in the paradigm (2nd versus 3rd person singular versus the infinitive) factor became highly significant only in the high-frequency category.

To summarize the results of both versions of the LDT, the type frequency of the form in the paradigm is a significant factor affecting RTs for the verbs with stem-cluster frequency of 10 per mln and above (the high-frequency category in the experiment). Since the 2nd person forms have lower type frequency in speech and take longer to process, one can conclude that token frequency affects the recognition of inflected words. The role of the other factors, such as word frequency, the number of repetitions, and length in letters, is not a new finding in a LDT, but their significance in the present study validates the other finding. Let us now address briefly the kind of processes involved in verb recognition in our LDT. The issue in question is whether inflected forms are stored in the mental lexicon undecomposed as whole words, or decomposed into stems and inflections. The lexical decision about the verb stimuli will be based on whole word access in the first case and decomposition in the second. One can argue based on the average frequencies for the forms in the paradigm calculated for the 213 most frequent verbs in the Uppsala and
Tübingen Corpora (see Table 4), that infinitives and the 3rd person singular forms each account for approximately 20% of the stem-cluster frequency. Therefore, for the verbs with stem-cluster frequency of 10–25 per mln, the frequency of the infinitive and 3rd person singular will be roughly in the 2–5 per mln range. The 2nd person singular will be 10 times less frequent, and therefore will fall below 1 per mln, as well as all the frequencies of the forms of low-frequency verbs with stem-cluster frequency already less than 1 per mln. Based on the comparison of the estimated whole-word frequencies of the Russian forms in the paradigm chosen for the LDT with the 6 per million threshold obtained by Alegre and Gordon (1999), it appears that all the word-forms included in the experiment are below the threshold for English. It is important to remember that the threshold found for English, with 7 per million being the lowest frequency showing the whole-word effects is relevant for Russian. And indeed, Russian has so many more forms in the conjugational paradigm than English that a threshold on storage similar to English would imply that there are several times more verb forms stored in the Russian mental lexicon. While in principle this is not impossible, it still needs empirical proof, which is lacking at present. In the absence of the data indicating that Russian has a larger storage space for the inflected forms, it is safe to assume that all the verb forms in our LDT are stored decomposed. This would mean that in order to perform the LDT, the subjects would need to analyze and decompose the stimuli. This operation took less time with the higher-frequency conjugational pattern, with the differences most visible in the higher token frequency range. However, the infinitive produced shorter response latencies, than the 3rd person singular, which indicates that it may be stored and accessed differently, possibly, as a whole word, given its special status of a citation form.

5.2.4 Conclusion
The results of the LDT experiment confirmed the hypothesis that the frequency of the form in the paradigm affects the response latency to a visually presented verb. While the experiment was not designed to determine whether the verb forms are stored decomposed or as a whole word and which frequency ranges correspond to these two different kinds of storage, it clearly demonstrated that 2nd person singular non-past tense generate longer RTs than 3rd person singular forms, and the RTs to the infinitive are the shortest. The fact that 2nd person singular forms are much more rare and at the same time take more time to process, and this effect remains across the verb classes, supports the idea that the frequency of the pattern or rule plays a role in morphological processing. Combined with a claim that Russian verb conjugation is rule-based rather than phonological association-based advocated elsewhere (see Gor 2004), these results support the hypothesis that rule application is affected by the probability of the verb form, which involves the use of this rule. The obtained results raise the issue of the unit of storage – whether it is the basic stem, or the infinitive, or both, or something else, and whether this depends on the regularity of the morphological processes. The results of the pilot LDT1 experiment indicate that the infinitive is recognized faster than the 3rd person singular form, despite the fact that these two forms have very similar estimated whole-word frequencies. This difference is smaller in the low-frequency verbs, but it does not disappear. At the same time, this tendency is absent in nonce verbs. These facts support the idea that the infinitive has a special status...
with regards to storage, hence no shorter latency in nonce verbs, in which none of the forms can be potentially stored.38

6. General conclusions

Two experimental methods, a real and nonce verb generation task and a lexical decision task were reviewed in this paper, made significant contributions to the understanding of the mechanisms underlying the processing of inflectional morphology, and more broadly, linguistic processing in general. The use of these experimental methods is enriched by targeting different populations of speakers, native adults and children, as well as second language learners, and also by focusing on languages with complex inflectional morphology and numerous verb classes varying in the degree of regularity, in addition to English. The reviewed studies with Italian and Russian verbs demonstrated that phonological similarity and type frequency influenced generalization rates in verb generation experiments. The results of LDT experiments with Russian stimuli showed that both type and token frequencies affect RTs. Also, a LDT experiment with English word stimuli provided evidence that the whole-word representations of morphologically complex inflected words, including inflected verbs, with the frequency above 6 per million are available in linguistic processing. The latter finding challenges the dual-system account of morphological processing. The observed influence of the input frequencies suggests that the speakers brought to those two tasks, the verb generation task and the LDT, their knowledge of linguistic probabilities. The Russian data demonstrated that, while token frequency influenced the speed of access to or retrieval of the lexical item, type frequency, or the frequency of the pattern, influenced generalization rates in novel verb generation and RTs in a LDT with visually presented verbal stimuli. In the verb generation task, it was the frequency of the verb class that affected the generalization rates. In the LDT, the frequency of the particular form in the paradigm was one of the significant factors for the high-frequency verbs. The presence of these type frequency effects across all the verb classes, including the regular default class, implies that the kind of processing, which would qualify as regular according to the dual-system approach standards, is influenced by linguistic probabilities. In this sense, the processing of Russian inflectional morphology exhibits the influence of probabilities on rule application. At the same time, the two sets of studies showed the differences in verbal processing in three populations of speakers – native Russian adults and children, and American learners of Russian. Only adult L1 speakers could fully access the frequencies observed in native Russian and make efficient use of the native probabilities. Adult beginning L2 learners’ performance on the verb generation task reflected the fact that the frequencies they relied on were non-native, as they were derived from a reduced verbal system available to them. However, adult L2 stem recognition rates in nonce verb generation were in most cases higher than those of L1 children. This last population of

38. Unfortunately, no post hoc paired comparisons tests were reported in the Sliusar paper.
speakers, L1 children, demonstrated a clear developmental tendency from age 4 to 6 in a verb generation task.

And finally, experimental study of the processing of complex verbal morphology in several languages challenges the original strong claims made by the proponents of two main theoretical positions, the dual-system and the single-system approaches. More precisely, it shows that the “litmus test” used by both parties, which relies on the assumption that input frequencies and probabilities, as well as phonological similarity play a role only in irregular processing based on associative patterning, and not in regular processing based on symbolic rule computation, turns out to be false. In fact, the results of the reported studies suggest that frequencies and phonological similarity influence both regular and irregular processing, and therefore the probability-based argument cannot be used to qualify the type of processing as associative patterning or symbolic rule application. At the same time, in addition to this negative result, the research data obtained by different experimental methods suggest another possibility, namely, that symbolic rule computation is not completely immune to the influence of linguistic probabilities, and that there is no categorical division between the lexicon and grammar, on the opposite, these two entities are interrelated. In a sense, this development would illustrate the position expressed in the Minimalist Program that words are inflected in the lexicon (Chomsky 1995).

References


