Applied Psycholinguistics **20** (1999), 119–148
Printed in the United States of America

Gender processing in native and nonnative Russian speakers

ROMAN TARABAN
Texas Tech University

VERA KEMPE
State University of New York, Oswego

ADDRESS FOR CORRESPONDENCE
Roman Taraban, Department of Psychology, Texas Tech University, Lubbock, TX 79409-2051.
Email: tirmt@ttacs.ttu.edu

ABSTRACT
Traditional computational accounts of gender representation and learning (e.g., Carroll, 1989, 1995) differ radically from cue-based and connectionist accounts. The latter but not the former predicts that access to noun gender will vary depending on the reliability of noun endings (and other sublexical elements and morphological constituents) in marking gender, and that agreement markers can be used strategically to constrain the genders of ambiguously marked nouns. Adult native (L1) speakers of Russian (Experiment 1) and advanced nonnative (L2) speakers (Experiment 2) read Russian sentences on a computer and were asked to choose one of two inflected past tense verbs in a forced choice task. The verbs either matched or mismatched the gender of the subject NP. Half of the target trials used opaque (end-palatalized) subject nouns, which were ambiguously marked for gender, and the other half used transparent (regularly marked) subject nouns. Noun type was crossed with the presence or absence of a gender-marked adjective in the subject NP. When an adjective was present in the subject NP, both L1 and L2 speakers were significantly faster at reading and selecting the correct verb form. L2 but not L1 speakers showed longer reading and choice latencies and made more errors when the subject noun was opaque. The data showed that L2 speakers used adjective inflections to disambiguate the gender of opaque subject nouns and to select gender appropriate verb inflections. The accuracy data for L1 and L2 speakers were tested against several connectionist models. The models' success in accounting for the data suggested that L1 and L2 speakers may depend on a common learning mechanism and thus resemble one another at the computational level, contrary to traditional computational accounts (Carroll, 1989, 1995).

From a linguist's perspective, gender is a morphosyntactic property of nouns. Nouns have an intrinsic gender. Adjectives, pronouns, numerals, adverbs, verbs, and other classes do not. Nonnominal constituents reflect the gender of associated nouns through inflectional morphology or form changes. For purposes of grammatical analysis, gender is most transparently reflected through the morphological association of words in a sentence: "The determining criterion of

© 1999 Cambridge University Press 0142-7164/99 $9.50

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

gender is agreement. . . . Saying that a language has three genders implies that there are three classes of nouns which can be distinguished syntactically by the agreements they take" (Corbett, 1991, p. 4). For native (L1) and nonnative (L2) speakers, though, it is still an open question as to what role agreement plays, if any, on gender assignment to nouns (Carroll, 1995; Sokolik & Smith, 1992). The broader issue, which is the topic of this article, concerns the interactions between gender attribution (knowing the gender of nouns) and gender assignment (knowing the co-occurrence requirements for constituents related to the noun) (see Carroll, 1995, for further elaboration of these terms). These interactions are examined in terms of the immediate on-line processing of gender as well as in terms of the longer term processes associated with learning gender. Standard experimental methods and connectionist modeling are brought to bear on these issues.

Traditional models of gender representation and processing are based on well-structured symbolic representations (Carroll, 1989; Corbett & Fraser, 1993; Neidle, 1982). Gender is a feature of individual nouns or other morphological units in a morphosyntactic lexicon (e.g., in French [+masc]; Carroll, 1989, p. 558). Representations are morphological, as opposed to those based on phonetic or phonological codes (Carroll, 1989): "Because only psycholinguistic processes can construct such representations, we must hypothesize ones that will, for example, construct and label word trees, define a particular constituent as the head of a word, and elaborate percolation principles which will ensure that the morphosyntactic features of the head are reassigned to the word as a whole" (p. 567). Gender agreement depends on more than the mere co-occurrence of potentially relevant constituents (Carroll, 1995): "gender agreement with a given noun takes place only when both the noun and the agreeing constituent occur in highly constrained syntactic positions (namely, under government and c-command)" (p. 200). Related examples for Russian can be found in Neidle (1982), where lexical entries include a definitive gender feature (e.g., [+masc]) and where lexical representations are processed within the well defined syntactic requirements of lexical functional grammar (Bresnan, 1982).

According to standard generative views (e.g., Chomsky, 1965), word classes are defined by sets of morphosyntactic features. Gender could be a universally available morphosyntactic feature associated with nouns. Given that the gender feature is universally available in L1 acquisition, the feature needs to be parametrically set, not learned. Once a child figures out that the language has gender, "*no* nominal expression can be lexically encoded without a gender feature. This follows because the gender feature is part of the specification of what a noun is" (Carroll, 1989, p. 572). If gender is acquired and encoded in the manner outlined by Carroll, then it follows that L1 speakers should show little difficulty in accessing and using gender. This is precisely what Carroll (1989) claims, with respect to native French speakers: "Francophones, adult and child alike, have consistent and virtually instantaneous access to the gender of known words because every nominal element bears a gender feature in its lexical entry and that information is made available during production and speech decoding processes" (p. 572). In support of Carroll, developmental data suggest that learning

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

about gender proceeds quickly. By the age of 3, German, Polish, and Russian children use gender reliably in their native language (Maratsos, 1988).

On the other hand, even native Italian adults find processing the gender of some nouns to be more difficult than others. Bates, Devescovi, Pizzamiglio, D'Amico, and Hernandez (1995) asked adult L1 speakers of Italian to indicate the gender of Italian nouns that were presented auditorily: some ended in a reliable cue to gender (i.e., the vowels *o* for masculine and *a* for feminine) and others ended in an ambiguous cue (i.e., the vowel *e,* which occurs with masculine and feminine nouns). Participants made more errors and took longer to respond when gender was ambiguously marked with the *-e* ending. This finding suggests that even adult L1 speakers may still have some difficulty processing gender. Findings like these are difficult to reconcile with the traditional computational model outlined by Carroll.

Researchers have also found that L1 speakers do not always fit the predictions of the traditional computational model. These studies examined the roles of phonological, morphological, and orthographic features of nouns in gender classification. Tucker, Lambert, and Rigault (1977) tabulated the within- and between-category distributions of phonological endings of nouns in the French dictionary *Petit Larousse* and were able to identify a set of noun endings that could be found in 85% of the 31,619 nouns in the dictionary. The endings varied in how strongly they were correlated with the masculine and feminine genders. When Tucker et al. tested the endings using rare and novel words, participants aged 7 to 18 years chose a masculine or feminine determiner (*un* or *une*) in close proportion to the distribution of these endings on masculine and feminine nouns in the *Petit Larousse.* The Tucker et al. data indicated that phonological markers that vary in reliability are nevertheless useful for classifying gender, and that L1 speakers rely on these markers, especially when words are unfamiliar or new. The traditional computational model that Carroll (1989) outlined does not provide for the type of inferencing evidenced in Tucker et al. This criticism, though, needs to be tempered with the observation that Carroll allowed for the possibility that native French speakers would resort to heuristic rules based on sublexical attributes of nouns when the gender of nouns was not known and the nouns were not morphologically analyzable. Some of these rules could be like those discovered by Tucker et al. Others could be based on the semantics of the nouns (cf. Zubin & Koepcke, 1986). However, this leads to several dilemmas. Why do individuals need two methods for processing gender – one that uses gender-marked lexical representations and one based on heuristic rules? How would an individual determine which method to use, particularly for lower frequency words which have weak lexical entries and uncertain gender marking?

Research with L2 speakers indicates that they behave in a manner similar to the L1 speakers in Tucker et al. (1977), relying on phonological and orthographic markers that vary in reliability for classifying gender. Taraban and Roark (1996) found that English speakers who learned gender-like classifications for French nouns made fewer errors when orthographic cues more reliably indicated the correct gender. In an artificial language training study, Brooks,

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

Braine, Catalano, and Brody (1993) showed that the presence of nouns that were reliably marked for gender made it possible to learn the genders of ambiguously marked nouns. Brooks et al. and Ellis and Schmidt (1997) showed that the frequency of exposure to nouns played a role in learning gender. Findings like these have led some researchers to question the need to postulate a universally available gender feature and traditional lexical representations (Brooks et al., 1993; Ellis & Schmidt, 1997; MacWhinney, Leinbach, Taraban, & McDonald, 1989; Sokolik & Smith, 1992; Taraban, McDonald, & MacWhinney, 1989; Taraban & Roark, 1996). Several tests of the traditional view have been carried out using connectionist network simulations. These simulations have shown that noun genders can be learned using phonological or orthographic noun representations, and that the patterns of learning capture the frequency and regularity effects found in the human data.

The issues to this point generally relate to what Carroll (1995) called gender attribution, or knowing the gender of a given noun. A separate set of issues relate to gender agreement, or knowing how nouns govern the forms of related constituents. In the traditional computational model, independent lexical entries encode gender features. On the lexical models of Carroll (1989, 1995) and Neidle (1982), nominal gender coding is definitive and access is "instantaneous." Gender is recovered from an utterance according to the structural principles (e.g., c-command) and processes (e.g., percolation) mentioned earlier. For these reasons, it is difficult to understand how agreement markers could prime the gender of related nominals. Relatedly, Corbett (1991) asserted that gender classification must simply be known. He rejected the possibility that gender attribution could be based on agreement, for example, in the form of a rule like "Noun x is feminine because it takes agreement y" (p. 33). However, if native speakers have probabilistic knowledge of noun gender (Bates et al., 1995; Tucker et al., 1977), there would be a strategic advantage in using the reliable gender markers of agreeing constituents to constrain nominal gender. This could be carried out in several ways computationally. One way would be through associative links between the relevant constituents, as implemented through connectionist models.

A number of studies involving lexical access have demonstrated priming effects between modifiers and nouns. These types of findings are more readily explained within connectionist networks or other associative networks than they are through the model described by Carroll (1989, 1995). Agreement markers have been shown to facilitate processing in some cases and to inhibit processing in others, depending on whether they were congruent or incongruent with the nouns that followed. In a study by Grosjean, Dommergues, Cornu, Guillelmon, and Besson (1994) using native French speakers, auditorily presented nouns were preceded by an article that carried overt phonetic information about the gender of the noun that followed (e.g., *une plage* 'a beach') or a word with no phonetic information about gender (e.g., *jolie plage* 'nice beach'). In a gating paradigm in which increasingly longer portions of a noun were presented, participants needed a smaller fragment to recognize a noun and responded with more confidence when the preceding word carried gender marking. This finding suggests that the gender cues of agreeing constituents may directly assist in lexical

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

access for related nouns. Pilot work in Italian reported in Bates et al. (1995) showed that gender marking on an adjective prime can facilitate or inhibit lexical access, as measured by the repetition of auditorily presented nouns (e.g., BRUTTO–TORRE [masc.–masc.] vs. BRUTTA–TORRE [fem.–masc.]). Schriefers (1992) presented related findings for noun phrase production. In several studies using a lexical decision task, lexical decisions for nouns were faster when the nouns were preceded by words with congruent gender markers than by those with incongruent gender markers (Colé & Segui, 1994; Grosjean et al., 1994; Gurjanov, Lukatela, Lukatela, Savić, & Turvey, 1985; Gurjanov, Lukatela, Moskovljević, Savić, & Turvey, 1985). For instance, in Gurjanov, Lukatela, Lukatela et al., singular, nominative nouns were preceded by possessive adjectives (e.g., *moj* 'my') that always agreed in case and number but only half the time in gender. Lexical decisions were faster when the gender of the prime was congruent with the gender of the noun (*moj lonac*) than when it was incongruent (*moja lonac*). None of these tasks was explicitly geared towards processing gender. Nonetheless, participants were sensitive to the gender information carried by the related constituents that were used as primes.

One interpretation of the results from the lexical decision studies (Colé & Segui, 1994; Grosjean et al., 1994; Gurjanov, Lukatela, Lukatela et al., 1985; Gurjanov, Lukatela, Moskovljević et al., 1985; Lukatela, Kostić, Todorović, Carello, & Turvey, 1987) that has been proposed is that inflectional morphemes or other inflected forms that occur earlier in a sequence may either facilitate or inhibit later processing, due to the activity of an independent syntactic module, like the one proposed by Forster (1979). The activity of this module is not "an automatic spreading of influences among connected representations in the internal lexicon" (Gurjanov, Lukatela, Moskovljević et al., 1985, p. 69) but is closer to obligatory checking for grammatical coherence. This interpretation is consistent with the traditional computational model proposed by Carroll (1989, 1995). Essentially gender could be computed as Carroll described and the outcomes checked for grammatical coherence in a syntactic module. Several potential problems with this explanation have already been presented. In sum, it fails to explain what individuals do when a noun is rare or new, and it has difficulty explaining findings like those of Tucker et al. (1977). The activation of weighted cues in a connectionist network, as opposed to gender checking, provides an alternative explanation that is better able to account for those effects.

The findings of Bates et al. (1995), Tucker et al. (1977), and others suggest that ambiguous gender marking would result in slower and more errorful processing relative to those instances when gender marking was less ambiguous. On the other hand, approaches that assume the lexical representation of gender, such as Carroll (1989, 1995) or Corbett (1991), predict no differences in the processing of regularly and ambiguously marked nouns. Experiments 1 and 2 explored these contrasting hypotheses by examining the processing of sentences requiring gender agreement by L1 speakers and advanced L2 speakers of Russian.

The priming studies presented earlier suggest that gender cues or agreement markers occurring early in a sentence could facilitate gender processing later in the sentence. However, those studies did not examine the effects of gender markers in the subject NP on gender processing in the verb phrase. Experiments

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

1 and 2 examined these longer distance dependencies using intact sentences. Furthermore, previous studies have compared congruent and incongruent marking rather than congruent marking and a more natural neutral baseline. Experiments 1 and 2 used a neutral baseline in order to focus more directly on the benefits of agreement marking to the nominal constituent the adjective is part of and to the syntactic processing of the main verb.

The patterns of effects that were observed in the two experiments were examined further using connectionist models. A major goal was to test the adequacy of the traditional computational account and that of cue based and connectionist accounts.

EXPERIMENT 1

The distributional analyses of Tucker et al. (1977) were based on the cue reliability of word features (or combinations of features). Cue reliability is the conditional probability of a category given a particular cue. In French, for instance, the cue reliability of an *-eur* ending is .95 for masculine gender (Tucker et al., 1977). This means that the probability of masculine gender is about .95 given that a noun ends in an *-eur* noun ending. When a representative language sample is used for computing cue reliabilities, the statistic is fairly predictive of learning and processing difficulty. The notion of cue reliability has been elaborated through several other cue statistics in the Competition Model (Bates & MacWhinney, 1989; MacWhinney, 1987a, 1987b, 1989; McDonald & Heilenman, 1991).[1]

In Russian, noun gender cues are not always reliable. In particular, there is a class of feminine nouns (e.g., *sol'*) and masculine nouns (e.g., *lokot'*) whose nominative forms end in palatalized consonants.[2] The distribution of these end-palatalized nouns to the feminine and masculine genders is roughly even (based on Pulkina, 1978). Thus, end-palatalization does not favor either gender, and the distribution of this particular cue cannot be used as a gender marker in the same way as other cues (e.g., the *-a/-ja* marker for feminine).[3] Using the designations in Bates et al. (1995), nouns with end-palatalized endings will be referred to as "opaque," and nouns with endings that reliably signal gender will be referred to as "transparent." If phonological and morphological cues are used for noun gender classification, one prediction is that L1 speakers would find opaque nouns more difficult to process than transparent nouns. A second related prediction is that difficulty in accessing the gender of opaque nouns would lead to difficulty in selecting correct agreement markers.

The stimuli used in both experiments were intact, ordinary Russian sentences that were constructed around two types of nouns (transparent and opaque) and two genders (masculine and feminine). Sentences with subject nouns like *muka* (example 1), which are transparently marked for feminine gender, were compared with sentences with opaque feminine nouns like *sol'* (example 2), which are ambiguous with respect to morphophonological gender marking. Similarly, sentences with transparently marked masculine nouns like *palec* (example 3) were compared with sentences with opaque masculine nouns like *lokot'* (exam-

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

ple 4). Gender marking on the transparent nouns was quite reliable. The opaque nouns were phonologically ambiguous with respect to gender.

(1) Dazhe (obychnaja) muka teper' izchezla iz magazinov. (fem., transparent noun)
'Even (ordinary) flour now vanished from the stores.'
(2) Dazhe (obychnaja) sol' teper' izchezla iz magazinov. (fem., opaque noun)
'Even (ordinary) salt now vanished from the stores.'
(3) Nakanune (otjokshiy) palec sil'no bolel ot udara. (masc., transparent noun)
'The day before the (swollen) finger strongly hurt from the injury.'
(4) Nakanune (otjokshiy) lokot' sil'no bolel ot udara. (masc., opaque noun)
'The day before the (swollen) elbow strongly hurt from the injury.'

The stimuli contained an optional adjective before the subject noun. Adjectives were used to test the role of agreement markers. Previous studies have indicated that gender marking on the adjectives could speed up access to the following noun. Two other effects of adjectives are consistent with earlier studies but have not been directly tested prior to this experiment: that the unambiguous gender markers on the adjectives could reduce or eliminate any ambiguity associated with opaque nouns, and that gender marking on the adjective could facilitate acceptance of a gender marked verb that occurred later in the sentence.

The gender of the nouns was included as a factor because of the possibility that the two types of nouns would behave differently. This possibility is based on linguistic classifications that put masculine transparent and opaque nouns into the same declension but feminine transparent and opaque nouns into different declensions. Feminine nouns also have more phonological and visual similarity across the adjective, noun, and verb inflections (e.g., *-aya, -a, -ala*) compared with masculine nouns (e.g., *-iy*, null, *-al*). If feminine and masculine nouns have similar effects in the experiment, then that would suggest that the differences in linguistic classification may not be important in processing, and it would help to rule out the influence of visual and phonological similarity.

METHOD

Participants

The participants in this study were 26 adult L1 speakers of Russian. Twenty-three participants had lived less than two years in the United States; the remaining three participants had lived three, seven, and eight years in the United States but reported predominantly using Russian in their everyday lives. All participants had been exposed to English or another second language after age 7.

Materials

The materials were based on 24 sentence frames that had the structure *initial segment with optional adjective/subject noun/adverb/verb segment with prepositional phrase,* as illustrated in examples 1 through 4. Of the sentence frames, 12 were used with masculine nouns and 12 with feminine nouns. Each sentence

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

Table 1. *Mean log frequencies per million* (*standard deviations*) *for the transparent and opaque nouns, by gender*

Masculine		Feminine	
Transparent	Opaque	Transparent	Opaque
1.47 (.69)	1.55 (.60)	1.60 (.55)	1.51 (.84)

frame was used twice: once with a subject noun that was transparently marked for gender and once with its opaque counterpart. Each of the sentences could appear with or without an adjective that modified the subject noun. This made a total of 96 possible experimental sentences, which are listed in the Appendix. The subject nouns across the four sets of items were closely matched for word frequency, using Zasorina (1977) as a source, as shown in Table 1. A one-way analysis of variance confirmed that the mean log frequencies of the four types of nouns were not significantly different, $F(3, 44) = 0.75$, n.s.[4]

Additionally, each sentence frame was used to construct a filler sentence with a transparent noun whose gender was the opposite of the experimental sentences. These also appeared with or without an adjective (total of 48 sentences). The opposite-gender fillers were included in order to prevent a strategic association of the sentence frames with a particular gender. In addition, there were 36 fillers involving neuter gender with a syntactic structure that was similar to that of the target sentences. For practice, 12 sentences like the experimental and filler sentences were used.

Sentence sets for use in the experiment were created in pairs. Both sets included the 24 sentence frames using opaque subject nouns, the same 24 sentence frames using transparent subject nouns, and the 60 filler sentences. For one set, half of the sentences with transparent gender marked nouns and half with opaque nouns were randomly selected to include the optional adjectives. In the complementary set, the other half of the nouns appeared without the optional adjectives.

It is possible that the inclusion of adjectives leads to greater specification of sentence contexts and an increase in the semantic plausibility of the verbs. This, in turn, could facilitate performance in the verb segment so that any benefits for sentences with adjectives could reflect greater verb plausibility rather than priming of gender. In order to rule out this possibility, ratings of the semantic plausibility of the experimental sentences were obtained. The sentences were presented with correct gender agreement between subject noun and verb. A total of 36 native Russian speakers who did not participate in the main experiment were asked "How plausibly does the action described by the verb fit into the sentence context?" and were told to rate the sentences using a scale of 0 (not plausible) to 5 (very plausible). Each participant saw each sentence frame in only one condition. In addition to the experimental sentences, participants rated 24 filler sentences that represented a wide range of plausibility. The mean plausibility ratings for the experimental sentences with and without adjectives were

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

virtually identical (3.78 vs. 3.73) and thus were not significantly different. This ruled out the possibility that any effects that were found would be due to differences in semantic plausibility arising from the presence of adjectives.

Near the conclusion of the study we realized that the masculine opaque noun *tufel'*, which was used in the experiment, has a transparent feminine counterpart, *tuflja*, in Russian, and that there are dialectological preferences for one or the other form. Thus, the sentences with *tufel'* and the associated transparent noun were not included in the analyses.

Procedure

Each participant read and selected verbs for the 12 practice sentences and the 108 sentences in one of the paired sets described in the Materials section. Sentences were presented in random order. The target sentences included (1) 12 sentences with masculine opaque subject nouns and 12 sentences using the same sentence frames but with transparent masculine nouns and (2) 12 sentences with feminine opaque subject nouns and 12 sentences using the same sentence frames but with transparent feminine nouns. Six of the sentences in each group of 12 appeared with the optional adjective. The 60 filler sentences included 24 opposite-gender sentences and 36 sentences involving neuter gender. For each participant, half of the filler sentences appeared with an adjective modifier and the other half without the modifier.

The sentences were presented one at a time on a computer screen using a moving-window method. Each trial began with the presentation of a fixation mark. Participants read through each sentence by pressing the spacebar on the computer keyboard. Each sentence was presented in segments in a noncumulative fashion, consisting of an introductory word or phrase and the adjective (if present), the noun, the adverb, and the rest of the sentence (S = segment):

Nakanune (otjokshiy)	palec	sil'no	BOLELA BOLEL	ot udara.
S1	S2	S3	S4	

The final segment contained the main verb choices, which were in capital letters and were varied randomly in terms of which gender version appeared first and second. The main verb was selected by pressing a keyboard character marked in Russian to indicate the verb on the left or a keyboard character indicating the verb on the right.

Design

The proportions of errors and reading and choice latencies for correct sentences were used as dependent variables in analyses of variance. Latencies for the initial segment, subject noun, and adverb reflect reading latencies. In the verb segment, the latencies reflect the combination of reading and verb choice latencies – these are referred to as reading/choice latencies. The data from sentences for which the participants had provided a correct verb choice were used in the analysis of latencies. Effects were examined separately for each sentence position (the initial segment, subject noun, adverb, and the verb segment).

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

Table 2. *Mean accuracy rates for Experiment 1 for L1 Russian speakers and mean* (*median*) *Luce ratios for Model 3 based on five simulations*

Condition	L1 Participants	Model 3
Transparent nouns, no adjectives	.99	.99 (.99)
Opaque nouns, no adjectives	.98	.96 (.97)
Transparent nouns, with adjectives	.99	.99 (1.00)
Opaque nouns, with adjectives	.98	.99 (.98)

Each dependent measure was analyzed separately using participants and items (i.e., the 24 sentence frames) as random factors. When possible, the *F* statistics using participants (*F1*) and items (*F2*) were combined into a *min F′* statistic (Clark, 1973).[5] In the *F1* analyses, the type of subject noun (transparent or opaque), the gender of the subject noun (masculine or feminine), and the presence versus absence of an adjective were within-subjects factors. In the *F2* analyses, gender was a between-items factor and the remaining factors were within-items. A main effect or interaction was considered statistically significant only if a *p* value less than 0.05 was achieved in both the *F1* and *F2* analyses. Interactions were analyzed by decomposing the error term for the interaction and determining which components of the effect accounted for a significant measure of variance.

Homogeneity of variance assumptions were tested using Cochran-C and Bartlett-Box tests (SPSS Inc., 1991). When there was a failure to meet these assumptions for the error data, the data were transformed using the arcsin transformation recommended by Winer (1962, p. 221).

The latency data were transformed using a *z* transformation recommended by Bush, Hess, and Wolford (1993). The transformation improves power in within-subject designs by scaling the effect of an independent variable for a given individual (or item) in terms of that individual's (or item's) variability or reactivity. The transformations were carried out separately for each sentence position by first finding the mean and standard deviation of all the latencies for a given individual (item) across all the experimental conditions. The raw latencies were transformed to *z* scores by subtracting the mean and dividing by the individual's (item's) standard deviation. Finally, the means of the *z* scores (instead of the raw scores) within each condition were computed to represent the participant (item) in that condition. For clarity, untransformed means are reported in the text.

RESULTS

Table 2 summarizes the accuracy rates for verb choices and Figure 1(A) summarizes the reading/choice latencies for each sentence segment by experimental condition (collapsed across gender). An examination of the accuracy data using analyses of variance and the factors of noun gender (masculine vs. feminine),

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

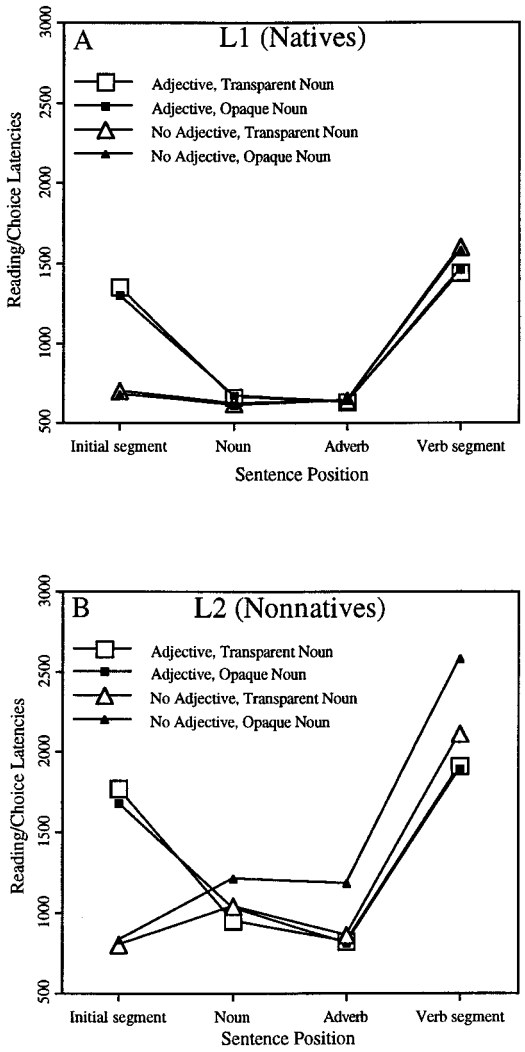


Figure 1. Mean reading times, by segment, for L1 (A) and L2 (B) participants.

noun endings (transparent vs. opaque), and adjectives (present vs. absent) revealed no significant main effects or interactions.

The analyses of the transformed reading/choice times were carried out using gender (masculine vs. feminine) and adjective (present vs. absent) for the initial segment and gender, adjective, and noun ending (transparent vs. opaque) for the remaining segments. In the initial segment, native speakers showed a significant effect for the presence of an adjective, *min F′* (1, 34) = 314.05, $p < .001$, because

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

of the reading time required for the additional word. There were no significant effects at the subject noun nor at the adverb. In the verb segment, when sentences had an adjective in the subject NP, reading/choice latencies were shorter compared to sentences without an adjective, $min\ F'\ (1, 33) = 5.35$, $p < .05$. This effect did not depend on whether or not the subject noun was transparent or opaque.

DISCUSSION

The first question in this experiment concerned the processing effects of gender cue marking on nouns. L1 speakers did not find sentences with opaque subject nouns any more difficult to process than those with transparently marked nouns. Reading latencies on the noun and adverb segments and the reading/choice latencies and error rates in the verb segment were similar for the two types of nouns. These results contrast with those of Bates et al. (1995), who found that native adult speakers of Italian had more difficulty indicating the gender of opaque nouns compared to transparent nouns. They also failed to confirm the findings of MacWhinney et al. (1989) and Taraban and Roark (1996), who showed that low cue reliability increased error rates for gender classification.

The second question was whether reliably marked adjectives would reduce the ambiguity associated with opaque nouns. Given that L1 speakers showed no evidence that the gender of the opaque nouns was ambiguous, either at the subject noun or main verb, it was not possible to test this question.

The priming studies that were reviewed supported a prediction that gender agreement marking would facilitate processing of the subject noun and/or the main verb. In the present data, there was no evidence of an effect in the reading times for the subject noun, but there was a reliable effect in the reading/choice times in the verb segment. L1 speakers were faster in selecting the gender appropriate form of the past tense verb when there was an adjective modifying the subject noun. This is the first study to show that gender agreement in the subject noun could facilitate the selection of the gender appropriate form of the main verb. It is also the first study to show gender-related lexical syntactic facilitation relative to a neutral baseline that corresponds closely to natural Russian sentences. Facilitation was similar for transparent and opaque nouns. Given that the L1 speakers knew the opaque nouns well, it is not totally surprising that the facilitation was independent of the type of noun.

There are several possible reasons why facilitation did not occur on the subject noun when preceded by a gender marked adjective in this study. One possibility is that the task of reading the subject noun differed either qualitatively or simply in difficulty from the verb choice task (and from the lexical decision, word recognition, and gender monitoring tasks in the other experiments). Verb choice required participants to make a decision, whereas reading the subject NP did not. In support of this possibility, Bates et al. (1995) found an effect for ambiguous gender marking when the task required gender identification but not when it required word repetition.

The null effect for the ambiguously marked subject nouns is consistent with

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

the claims of Carroll (1989) and Corbett (1991) that native speakers simply know the genders of nouns and have instantaneous access to noun gender. This may indeed be true, but it is important to consider the several ways in which processing automaticity of L1 speakers could come about. One possible way, attributed to Carroll (1989), has already been described. An alternative explanation may be offered when considering learning processes. L1 speakers' performance may simply reflect asymptotic learning levels. Although cue reliability is a useful metric for predicting learning and processing difficulties, its predictive utility diminishes as learning approaches asymptotic levels (Taraban & Roark, 1996, Figure 2).

The learning curves of connectionist networks provide a precise specification of learning and its eventual culmination in asymptotic performance. Networks that use the backpropagation learning algorithm eventually end at a high level of performance even when individual cues have no predictive value (e.g., see the XOR problem in McClelland & Rumelhart, 1988). Ellis and Schmidt (1997) showed that, when the frequency of exposure to irregular French nouns is high, performance in a connectionist network approaches that of regular nouns. Plaut, McClelland, Seidenberg, and Patterson (1996) showed similar frequency-by-regularity interactions for reading English words. Reasoning from the learning curves of networks like these, it was plausible that the L1 participants failed to show an effect for cue reliability because they had overlearned the gender of opaque nouns. It therefore followed that Russian speakers with less fluency than those in this experiment could be used to test whether the interaction of regularity and frequency observed in other domains applies to the learning of noun gender as well.

EXPERIMENT 2

Several findings in the literature suggest that it may be easier to detect an effect of cue reliability among speakers with lower levels of proficiency. The learning curves in connectionist models, as well as related empirical findings, suggest that Russian speakers with less experience with the language could show cue reliability effects. Furthermore, if less able speakers were to show evidence that the gender of opaque nouns was relatively more ambiguous than the gender of transparent nouns, it would be possible to test for an interaction between the presence/absence of adjectives and noun type. Thus, Experiment 1 was replicated using less fluent Russian speakers.

Participants

The participants were 18 L2 speakers who had studied Russian from 2.5 to 8 years and were mostly graduate students and faculty in the Slavic Languages Department of the University of Texas. Using a scale of 1 (no knowledge of the language) to 6 (native speaker competence), their self-estimated language proficiency in Russian ranged from 3.25 to 5.5.

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

Table 3. *Mean* (*median*) *accuracy rates for Experiment 2 for L2 participants and mean Luce ratios for Model 3 based on five simulations*

Condition	L2 Participants	Model 3
Transparent nouns, no adjectives	.98	.97 (.97)
Opaque nouns, no adjectives	.72	.73 (.73)
Transparent nouns, with adjectives	.98	.98 (.99)
Opaque nouns, with adjectives	.93	.92 (.93)

RESULTS

The error and latency data are summarized in Table 3 and Figure 1(B), respectively. In the error data, there was a main effect for adjective, *min F′* (1, 37) = 10.26, $p < .005$, for noun ending, *min F′* (1, 35) = 31.97, $p < .001$, and for the Adjective × Noun Ending interaction, *min F′* (1, 36) = 26.88, $p < .001$. A decomposition of the interaction indicated that end-palatalization had a significant effect in sentences without an adjective modifier, *min F′* (1, 38) = 22.16, $p < .001$, as well as in sentences with an adjective modifier, *min F′* (1, 36) = 4.97, $p < .05$.

Turning to the latency data, in the initial segment, there was a significant effect for the presence of an adjective, *min F′* (1, 33) = 110.12, $p < .001$, due to the time required to process an additional word. There was no significant effect at the subject noun; however, there were significant effects at the adverb. There was a main effect for adjective, *F1* (1, 17) = 6.06, $p < .03$, *F2* (1, 21) = 7.10, $p < .02$, for noun ending, *min F′* (1, 33) = 5.81, $p < .05$, and for the Adjective × Noun Ending interaction, *F1* (1, 17) = 7.89, $p < .02$, *F2* (1, 21) = 6.50, $p < .02$. The decomposition of this interaction showed that end-palatalization increased reading latencies when the subject noun was not modified by an adjective, *min F′* (1, 37) = 7.06, $p < .03$. When an adjective was present, the differences in latencies at the adverb were not significant. Thus, it appears that the opaque nouns were relatively more difficult than transparent nouns, but that the difference was eliminated when the adjective provided reliable information about gender. Note that the effects of an opaque subject noun first became significant on the adverb; this could be due to the moving-window method in which cognitive processing may lag somewhat behind an individual's key presses. An examination of Figure 1(B) shows that latencies were longer for opaque nouns.

Latencies in the verb segment were longer when sentences had opaque subject nouns, *min F′* (1, 37) = 7.05, $p < .03$. Latencies were shorter when there was an adjective in the subject NP, *min F′* (1, 37) = 17.84, $p < .001$. Both of these effects were qualified by an Adjective × Noun Ending interaction, *min F′* (1, 37) = 6.28, $p < .03$. The decomposition of this interaction indicated that the longer latencies occurred when sentences had opaque subject nouns and no adjective in the subject NP, *min F′* (1, 38) = 8.77, $p < .01$. When there was an adjective, the differences between sentences with transparent and opaque nouns were not significant.

Using the Bush et al. (1993) transformation, the overall mean for an individ-

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

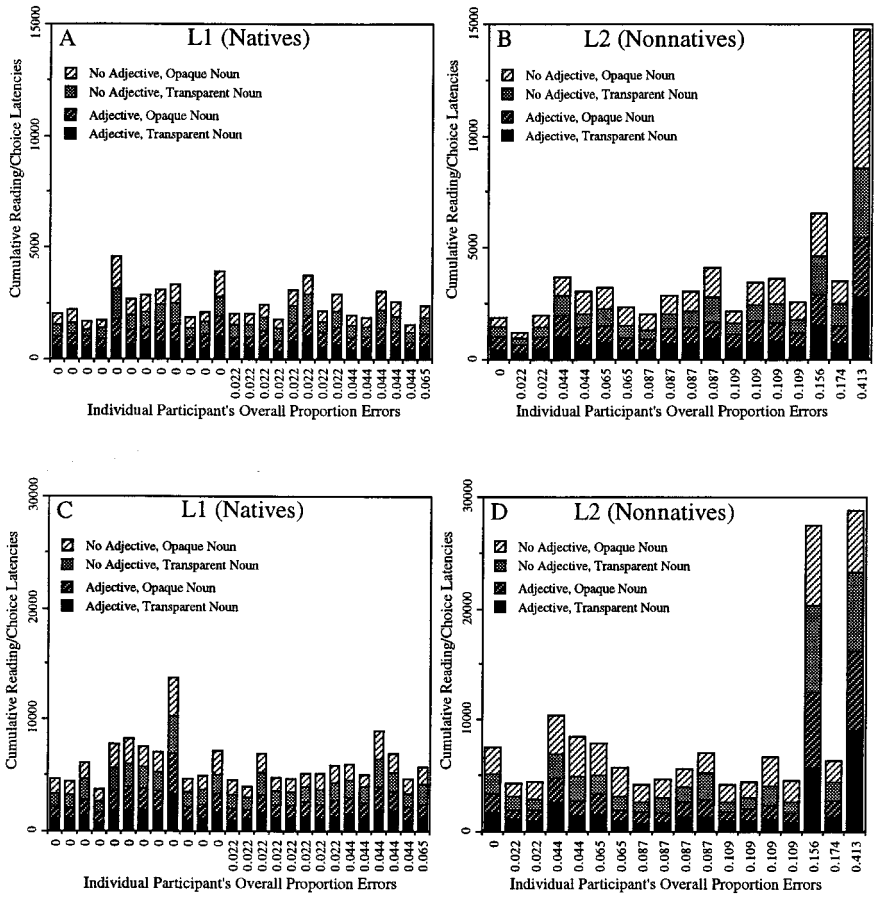


Figure 2. Mean reading times for the adverb versus mean error rate on the verb choice, by participant, for L1 (A) and L2 (B) participants and mean reading/choice times for the verb segment versus mean error rate on the verb choice for L1 (C) and L2 (D) participants.

ual or item was 0, making it impossible to examine between-subject or between-item effects using the transformed data. This was potentially a problem because gender was a between-item factor. In follow-up analyses, the raw data were checked for significant main effects or interactions involving gender. None were found, suggesting that gender was not a significant factor.

How comparable were the L1 and L2 speakers? In order to visualize the similarities and differences between the two groups, reading latencies for the adverb (Figure 2, A and B) and reading/choice latencies for the verb segment (Figure 2, C and D) were plotted against the total proportion of errors for each participant. The height of the bands for any given bar in the figures indicates the mean latency for that condition for a particular individual. Ignoring for a

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

moment the two L2 participants with long latencies, reading times for the L1 and L2 participants were similar. The graphs also allow one to examine visually whether participants with lower error rates also had lower latencies, which could be taken as a trend associated with emerging expertise and automaticity. This did not appear to be true for either the L1 or L2 participants. Finally, perhaps the L1 or L2 participants traded speed for accuracy. Visual inspection of the latencies relative to proportion errors did not suggest a speed–accuracy tradeoff. Because the two L2 participants with long latencies appeared to be outliers, the L2 data were reanalyzed without these two participants. For brevity, the specific *F* values will not be reported here. The pattern of significant main effects, interaction effects, and the decompositions of the interactions for the error and latency data were identical for the noun, adverb, and verb segments, whether or not the two outliers were included, with one exception: the effect of adjective was only marginally significant in the adverb region in the *F1* analysis, $F1$ (1, 15) = 3.91, $p < .07$. In general, the latency effects in the L2 data were not due to these two outliers. Given the close similarities between (A) and (B) and between (C) and (D) in Figure 2 when the outliers are eliminated, it is reasonable to conclude that the difficulty that L2 participants experienced with opaque nouns and the benefit they gained from adjective modifiers were not due to gross differences between the two groups but rather to more subtle differences in knowledge and experience.

DISCUSSION

Advanced L2 speakers exhibited difficulty in processing sentences with ambiguously marked subject nouns. There was clear evidence that the cue reliability of the subject nouns affected gender classification for these nouns and gender agreement for associated verbs. When unambiguous agreement markers on the adjective were available in the subject NP, the error rates for verb choice were reduced (.07 for opaque nouns vs. .02 for transparent nouns) compared to the subject NPs without an adjective modifier (.28 for opaque nouns vs. .02 for transparent nouns) but were still significantly different. Ambiguous gender marking on the subject noun produced longer reading times on the adverb and longer reading/choice times in the verb segment, but these differences were eliminated when a gender-marked adjective modified the subject noun. Thus, the L2 speakers here, as well as the L1 speakers in Experiment 1, showed a main effect in the verb segment indicating the facilitation provided by adjective marking in the subject NP. However, this effect was qualified by several interactions that showed that gender agreement markers became particularly important when the related nouns were ambiguously marked for gender. The predictions made at the beginning of Experiment 1 for cue reliability effects and facilitation due to priming were upheld in this experiment.

The absence of main effects or interactions with gender (masculine vs. feminine) ruled out the possibility that the visual or phonological similarity of feminine markers across linguistic categories (e.g., *-aya, -a, -ala*) was a factor. There was also no indication that the linguistic classifications that place transparent

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

and opaque masculine nouns in the same declension but transparent and opaque feminine nouns in different declensions made a difference.

These results suggest that cue reliability effects, as they relate to noun classification and gender agreement, are conditioned by the state of knowledge of the respondent. This prediction was derived from the general learning properties of connectionist networks and was confirmed by the marked difference in the behavior of the L2 speakers compared to L1 speakers. As anticipated at the outset of Experiment 2, cue reliability effects point to a potential flaw in the syntactic gender checking explanation for the priming effects observed in other experiments (Colé & Segui, 1994; Grosjean et al., 1994; Gurjanov, Lukatela, Lukatela et al., 1985; Gurjanov, Lukatela, Moskovljević et al., 1985; Lukatela et al., 1987). Forster's (1979) model, as well as other models that encode unambiguous entries for words and inflections (Corbett & Fraser, 1993; Levelt, 1989; Neidle, 1982), may capture characteristics of asymptotic performance, but the models have difficulty generalizing to earlier stages of learning. It is not clear how one could check for grammatical coherence when one is not sure of what the initial classifications are or how to relate these to agreement marking. For less able speakers, like the L2 participants in this study, a gender checking explanation would miss the important contributions that learning factors make to an explanation of processing.

Nouns determine the gender markings on related words like adjectives, determiners, pronouns, and verbs, which are marked to agree with the genders of the nouns. How do noun genders relate to these other word classes? Maratsos (1988) described how word representations could be organized in order to bring about noun genders and gender agreement. Maratsos pointed out that members of grammatical categories mark syntactic functions in similar ways. These shared methods of syntactic marking define the category. A noun is "masculine," "feminine," or "neuter" because it belongs to a set of nouns that use a particular set of agreement markers. The present experiment and related priming studies suggest that individuals do indeed have interconnections between noun genders and gender markers on related words, and that these connections are evoked readily. In order to compensate for incomplete learning, less able speakers may rely on a type of implicit inferencing whereby the known genders of constituents are used to disambiguate the genders of related constituents.

Experiments 1 and 2 provide snapshots of Russian speakers with different levels of gender processing skill. The L2 participants had less experience with the language and therefore, quite plausibly, their performance would represent an earlier point in the L1 participants' competency. One way to examine this possibility would be through the collection of far more extensive developmental data than are available here. Another option, which is pursued next, is to use connectionist simulations to track the emergence of gender competence with experience. The simulations are also used to examine the interconnections between gender marking for noun modifiers and subject-nouns and verbs.

Each of three connectionist models was tested using multiple simulations. Model 1 and Model 2 used a representative language sample and examined cue reliability effects on the acquisition of noun gender labels. The primary goal was to show that, early in learning, opaque noun gender marking results in

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

lower accuracy in gender selection. With sufficient exposure, accuracy for opaque nouns is similar to that of transparent nouns. This is the frequency by regularity interaction described earlier. Model 1 used orthographic input representations and Model 2 used phonological input representations. It was important to conduct simulations with both types of representations because the performance of the models and the conclusions about learning (cf. Carroll, 1995) that could be drawn could differ with different types of input representations. L1 learners (and the participants in Experiment 1) generally depend on auditory inputs for gender learning, whereas L2 learners (and the participants in Experiment 2) depend in greater measure on printed texts.

Model 3 was used to mimic the pattern of error data found for the L2 participants in Experiment 2 using a connectionist network in conjunction with a small artificial language. One goal was to observe how well the network could model the L2 speakers. Another goal was to determine whether the network would asymptote in a manner that mirrored the performance of the L1 speakers.

Carroll (1995) suggested that several manipulations were critical to evaluating the capacity of a connectionist network to model human gender learning. One was to test whether the network could use agreement marking to learn gender. Another was to test gender learning in the context of sentences in order to show that gender learning depended on linguistically related constituents and not simply co-occurring constituents. Both tests were implemented using Model 3. Furthermore, in Model 3, gender categories were implicit in the network's encoding of agreement markers and their association to sets of nouns, as suggested by Maratsos (1988).

CONNECTIONIST MODELS WITH EXPLICIT GENDER UNITS

Two sets of simulations were conducted in order to model the learning of Russian noun gender. The simulations used the backpropagation learning rule, as described in Rumelhart, Hinton, and Williams (1986) and implemented by McClelland and Rumelhart (1988). Simulations for Model 1 used orthographic representations and simulations for Model 2 used phonological representations. The 200 most frequent Russian nouns, as listed in the Zasorina (1977) frequency count, were used for learning. By assumption, the frequency of opaque nouns among the 200 most frequent Russian nouns should be similar to the frequency of opaque nouns in the input of Russian children. The 48 nouns used in Experiments 1 and 2 were also included. The normative frequency of each noun was scaled down in order to keep the training tractable, using the formula $(F/100) + 1$, where F is the normative frequency listed in Zasorina (1977). Each noun was represented with its scaled frequency. In coding the words for inputting into the connectionist networks, different word lengths had to be accommodated while still allowing the network to discover regularities in the word structure. Because the information relevant to gender marking is present at word ends, all words were right aligned. Furthermore, because Russian syllables can contain clusters with up to four consonants, the words were also syllable right aligned. The template and two examples of the syllable right-aligned coding are given here:

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

C C V C C C C V C C C C V C C C C V

P A L K A

T R O S T'

Orthographic and phonological representations also coded a very basic semantic feature of nouns – that is, whether the nouns were abstract or not. The use of the concrete/abstract feature was motivated as follows. Russian has a subclass of feminine opaque nouns that end in the derivational morpheme *-ost'* and refer to abstractions like the well-known *glasnost'*. Because the Zasorina corpus is based on written texts, the list of the 200 most frequent nouns contains a disproportionally large number of these words. In order to make it easier for the simulation model to identify these words, the semantic input unit +/– abstract was included. Note that the additional feature makes it easier, not more difficult, to learn the opaque nouns.

For the phonological representations, each phoneme at each position was coded by eleven phonological features that were suggested for the phonological description of Russian by Kucera and Monroe (1968). These features were vocalic, consonantal, compact, diffuse, low, stressed, nasal, continuous, voiced, sharp, and strident. If a letter was present at a certain position in the template, that letter's code was represented by a sequence of eleven integers consisting of zeros, negative ones, and ones. If a position in the template was not occupied by a letter, the empty position was coded by eleven zeros. Note that consonant palatalization is represented by +consonantal, +sharp. The coding of the phonological features at each position and the semantic features resulted in a total of 199 representational units.

For the orthographic representations, each letter at each position was coded using a fixed-length Huffman code. Huffman codes are often used for file compression. For the Cyrillic alphabet, the Huffman code requires 5 bits in order to assign a unique string of bits to each letter. If a position in the template was occupied by a letter, that letter was represented by its Huffman code as a sequence of five integers consisting of positive and negative ones. If a position in the template was not occupied by a letter, the empty position was coded by five zeros. Consonant palatalization was represented by a separate orthographic symbol, the soft sign. For two words in the list the addition of the soft sign resulted in a cluster of five consonants. In order to preserve the length of the syllable template, these two words were dropped from the orthographic training corpus. Because these words were the only four-syllable words with a two-consonant cluster at the word beginning, the first consonant position became redundant and was also eliminated, resulting in a total of 17 letter positions. The orthographic coding together with the semantic feature required a total of 86 input units.

The network architecture for Models 1 and 2 consisted of an input layer, a layer of hidden units, and three output units that corresponded to the masculine, feminine, and neuter genders. The number of hidden units was kept roughly proportional to the number of input units in both networks. Thus, the phonologi-

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

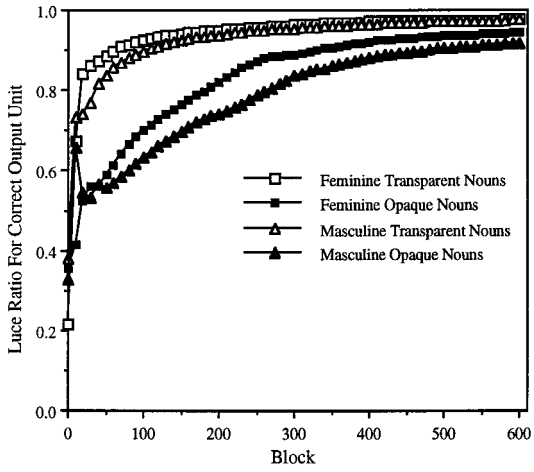


Figure 3. Luce ratios for the orthographic network for four types of nouns: transparent feminine and masculine nouns and opaque feminine and masculine nouns.

cal network had 75 hidden units and the orthographic network had 30 hidden units.

In a series of pilot simulations, the optimal parameter settings of the networks were determined. Values were used that prevented the networks from settling into local minima and allowed them to learn the whole set of training nouns. This was achieved at a learning rate of .05 and a momentum of .85 (see Rumelhart et al., 1986, for a discussion of the learning rate and momentum parameters). Each simulation was initiated by setting the connections in the network to small random weights. The network was trained for 600 blocks of trials, with each input pattern appearing once in random order in each block. After 600 blocks, learning was complete. The strength of the output units was computed using Luce (1959) ratios. For example, the strength of the feminine output unit was the ratio of the activation of the feminine output unit divided by the sum of the activation of all output units. The Luce ratio can be taken as a measure of strength of access of the correct noun. Other statistics are available for comparing a network's performance to human data. The Luce ratio was chosen because of its ease of computability and its use in other simulations.

The results of the simulations using orthographic representations are presented in Figure 3. The plotted values represent the mean values derived from three independent simulations. As can be observed in Figure 3, the correct gender classification is learned faster for the transparent nouns compared to opaque nouns (i.e., those that end in a soft sign). Over the course of learning, gender classification for opaque nouns becomes almost as fast and reliable as for transparent nouns, and the difference between both types of nouns diminishes greatly. The initial stage of learning, where performance on opaque nouns is significantly worse than performance on transparent nouns, corresponds to the results observed with L2 Russian speakers in Experiment 2. The later stages of

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

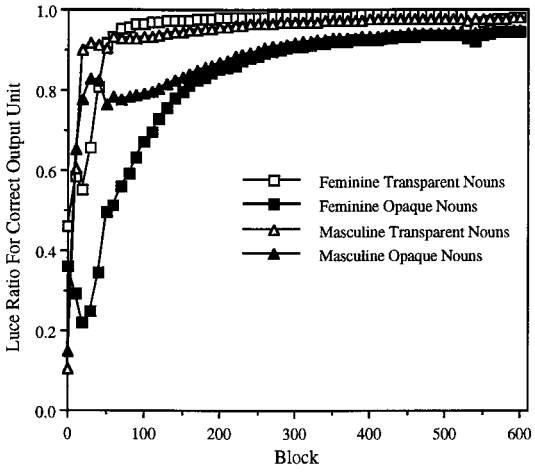


Figure 4. Luce ratios for the phonological network for four types of nouns: transparent feminine and masculine nouns and opaque feminine and masculine nouns.

learning, where performance for all nouns is almost equivalent, corresponds to the results found for the L1 participants in Experiment 1. Thus, the simulations support the plausibility of the suggestion that L2 learners and L1 speakers are at different points on the same learning curve for gender.

There is, however, one important difference between these simulations and the human data. L1 speakers acquire gender at an age when they receive predominantly auditory input. It is reasonable to assume that gender acquisition is basically completed by the age when children first encounter written language. Adult L2 learners of Russian, on the other hand, are more likely to receive significant parts of language input through written text. Thus, the orthographic coding provides a closer characterization of the input of adult L2 learners, whereas phonological coding resembles the situation of L1 speakers in that it provides a representation of the auditory input of Russian children. A connectionist network with phonological inputs (Model 2) was examined in order to see whether the input modality changes the learning trajectory.

As can be seen from Figure 4, again representing the mean values for three independent simulations, the final outcome of learning is the same as when orthographic inputs are used. Ultimately, the gender of all nouns is learned, and the differences between transparent nouns and opaque nouns become negligible. Differences between input modalities can be observed in the early stages of learning. With phonological inputs, there is a much longer delay for learning feminine opaque nouns than for learning masculine opaque nouns. A closer inspection of network performance revealed that, for opaque feminine nouns at the initial stages of learning, the activation of the masculine output unit is stronger than the activation of the feminine output unit. Thus, the network initially interprets these nouns as masculine. This is because the similarity between masculine nouns and opaque feminine nouns is much greater when using phono-

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

logical representations than it is when using orthographic representations. Essentially, the final soft sign in the orthographic representation of opaque nouns makes them equally dissimilar to transparent masculine and feminine nouns, while the phonological representation of the opaque nouns makes them more similar to masculine nouns because of the features +consonantal, –vocalic in the final phoneme. Observations reported in the diary of A. N. Gvozdev (1961), a well-documented longitudinal study of language acquisition in a Russian child, confirm that opaque feminine nouns are more difficult to acquire than opaque masculine nouns.

These simulations of Russian gender learning suggest that L2 learners of Russian should exhibit problems with all opaque Russian nouns, while adult L1 speakers should not exhibit any difficulties in processing these nouns. This is what the data from Experiments 1 and 2 show. Models 1 and 2 also suggest that the input modality can lead to somewhat different learning trajectories. With phonological inputs, the network exhibits the same difficulties in learning the gender of opaque feminine nouns that have also been observed in Russian children.

A CONNECTIONIST MODEL WITH IMPLICIT GENDER

In order to model more fully the accuracy data from Experiments 1 and 2, simulations were undertaken that encoded primitive sentence strings as inputs and word nodes and inflections as outputs. The purpose of these simulations was twofold: to examine whether gender marking effects for nouns could be attained without an explicit representation of gender and to determine whether agreement markers would show the same facilitation in this artificial learning environment as they did in the human data. These simulations used the backpropagation learning rule (Rumelhart et al., 1986).

The inputs to the network consisted of five word positions: an introductory word, an optional adjective, the subject noun, an adverb, and the main verb. Each of these words was represented by a 6-bit binary string. There were also two 1-bit inflections that were coded for adjective inflections (*-iy, -aja*), three 1-bit inflections for noun inflections (null, *-a,* opaque), and two 1-bit inflections for verb inflections (*-al, -ala*). Each word that was used as an input and each inflection were associated with a single output unit.

A total of 46 words was used for each simulation. At the beginning of a simulation, a unique binary number between 1 and 46 was randomly assigned to each word as its representation. There were 9 transparent masculine nouns, 9 transparent feminine nouns, 3 opaque masculine nouns, 3 opaque feminine nouns, 8 adjectives, and 12 verbs, as well as 1 introductory word and 1 adverb. Each transparent and opaque noun appeared once in each block of trials. Therefore, the frequency of any opaque noun compared to any transparent noun was identical at any point in learning. Recall that transparent and opaque nouns in Experiments 1 and 2 were also matched for frequency of occurrence. The proportion of opaque nouns to transparent nouns was lower in order to reflect the lower distribution of opaque nouns in Russian. Because there is currently no

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

computerized version of a Russian frequency count, it was impossible to determine precisely the frequency of opaque nouns in the language. In order to obtain a rough estimate, 700 nouns in blocks of Russian text from different sources (novels, newspaper editorials, children's books) were examined. About 11% of the nouns were opaque; therefore, the simulation may have made the learning of the opaque nouns somewhat easier than it is in Russian.

The network architecture for Model 3 consisted of an input layer consisting of 37 units (5 words, each 6 units long, and 7 inflection units), a hidden layer with 30 units, and an output layer with 53 units (46 words and 7 inflections). After several pilot tests, the learning rate parameter was fixed at .0125, the range of values for the random startup weights was limited to the range of ±0.1, and the range of the outputs was 0–1. (The remaining parameters were the default parameters used in McClelland & Rumelhart, 1988.)

Each simulation consisted of 6,000 blocks of trials; 24 sentences appeared in each block in random order. Throughout the simulation each noun retained its gender, but during each block the adjectives and verbs were randomly reassigned to nouns, with several constraints. The 12 verbs were randomly assigned to the 12 masculine nouns and randomly reassigned to the 12 feminine nouns. Of the adjectives, 3 were randomly assigned to transparent masculine nouns, 3 to transparent feminine nouns, 1 to the opaque masculine nouns, and 1 to the opaque feminine nouns. Thus, in each block of trials, one third of the sentences appeared with an optional adjective. This proportion was based on an analysis of a representative sample of Russian texts. The genders of the adjectives and verbs changed from block to block in order to match the genders of the subject nouns. Note that each block of trials consisted of a new set of sentences. Given the large number of possible combinations of adjectives, nouns, and verbs into sentences and sentences into sets, the inputs were relatively productive in a linguistic sense.

Learning took place in 20-block increments. During a learning trial, an entire sentence was presented on the input layer of the network, and feedback was provided for the entire sentence. This is comparable to an episode in which a child uses a meaningful physical context in order to make sense out of a sentence string (Braine, 1989). For each block, half of the trials with adjectives and half of trials without adjectives were comprehension-like learning trials. On these trials, inflections for the adjective (if present), noun, and verb were coded on the input layer. On production-like learning trials, inflections for the adjective (if present) and verb were withheld. Nouns were left intact as noun endings are not considered to be productive.

After every 20 blocks of learning, the network was tested using a new set of sentences in which the verb inflections were withheld on the input layer. Activations for the correct verb inflections on the output layer were calculated, using the Luce (1959) ratio (i.e., the activation of the correct verb inflection divided by the activation of both inflections). For each of five independent simulations, the blocks with the closest fits to the L1 (Experiment 1) and L2 (Experiment 2) participants' means were identified by calculating sums of squared error (SSE) and finding the blocks with the smallest sums. Because of differences in the random weights at the beginning of each simulation, the temporal positions of

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

the best blocks varied. Nevertheless, when matching L2 participants' means the best-fitting block always came earlier (mean = block 324) than when matching L1 participants' means (mean = block 2,808).

Tables 2 and 3 list the accuracy means for each condition, averaged across five independent simulations. The fit of the simulations to the human data is quite good. In Table 3, the simulation means are low for opaque nouns when these appeared without an adjective but improve greatly when an adjective modified the subject noun. L2 participants showed a similar pattern. In Table 2, there are very small differences between simulation means, as was true for the L1 participants.

The success of Model 3 in fitting the error data confirmed that processing difficulties associated with the gender classification of opaque nouns can be modeled without recourse to explicit gender labels. Model 3 achieved this by finding reliable cues and then associating them into the patterns of inflections occurring on related adjectives, nouns, and verbs. The success of the model provided positive evidence for the suggestion by Maratsos (1988) and Corbett (1991) that gender categories are implicit in the patterns of agreement markers that are required by different sets of nouns. It does not disprove the need for the explicit marking of gender features in the lexicon (Carroll, 1989, 1995), but it does show that there may be other alternatives.

In the L2 data in Experiment 2, there were significant effects of facilitation brought about by reliable agreement markers on adjectives. The present simulations demonstrated those effects early in learning. These effects of agreement markers in the data from Experiment 2 and in Model 3 are consistent with the role of agreement markers in learning noun gender that has been readily acknowledged by researchers interested in gender (Carroll, 1989, 1995; Maratsos, 1988; Sokolik & Smith, 1992). After sufficient experience, Model 3 reached a high level of accuracy. Opaque noun effects disappeared as did facilitation for agreement markers. This pattern matched the accuracy data of L1 participants in Experiment 1.

Models 1 and 2 are susceptible to several of the criticisms that Carroll (1995) leveled against Sokolik and Smith (1992), who modeled learning French gender using a connectionist network. Sokolik and Smith presented single nouns on an input layer and provided feedback for correct gender labels on the output layer. By presenting the gender learning problem in the context of sentences that included co-occurring and relevant constituents (e.g., adjective modifiers) as well as co-occurring but unrelated constituents (e.g., adverb modifiers) and by constantly testing the network with new sets of sentences, Model 3 begins to address a number of the requirements that Carroll laid out for the proper testing of connectionist models that learn gender.

GENERAL DISCUSSION

In order to access the gender of isolated nouns, a person must ultimately learn each specific noun and its gender. Corbett (1991) suggested that native speakers must simply know the genders of nouns. Pulkina (1978) emphasized direct knowledge for opaque nouns. An examination of the data for L2 speakers tested

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

in this study suggests that individuals may use gender agreement markers in order to bolster gender classification when there is some uncertainty about the inherent gender of nouns, such as opaque nouns in Russian. The use of gender agreement markers to facilitate processing may extend far into development, as evidenced by the small but reliable advantage in L1 participants' response latencies in sentences containing reliably marked adjectives.

The experimental findings and simulation results suggest that facilitation and inhibition in priming studies on gender effects may be due to more than simply checking for grammatical coherence (Colé & Segui, 1994; Grosjean et al., 1994; Gurjanov, Lukatela, Lukatela et al., 1985; Gurjanov, Lukatela, Moskovljević et al., 1985; Lukatela et al., 1987). There appears to be a constructive character to processing that is modulated by the particular state of learning of the speaker. This was particularly evident for the L2 participants.

L2 participants differed significantly from L1 participants in error rates, but their overall reading and choice latencies were comparable (see Figure 2). This may seem surprising, and it differs from results reported in Kempe and MacWhinney (1998), which showed slower processing speed for adult L2 learners than for L1 speakers in a lexical decision task. The discrepancy is most likely due to the fact that the overall exposure to Russian was higher for the L2 participants tested in Experiment 2 than it was for the L2 learners tested in Kempe and MacWhinney. Simply put, the L2 participants in this experiment may have been further along the learning curve.

The results from these experiments and simulations are preliminary and need to be replicated and extended through additional experimentation and modeling. Also, some care needs to be exercised when generalizing from the experimental results in a particular language as the specific interactions of gender cues in other languages might lead to different effects in processing.

A major cautionary note relates to the qualitative differences that may have existed between the L1 and L2 participants in Experiments 1 and 2. Carroll (1989) suggested that, early in development, the universally available gender feature allows for the coding of gender as a required morphosyntactic feature of all lexical entries. At some point in development this feature "atrophies," and L2 learners need to resort to heuristic strategies based on general purpose learning mechanisms. It is possible to explain the results of Experiments 1 and 2 using this type of reasoning. L1 participants failed to show difficulty with opaque nouns because of definitive morphosyntactic lexical coding of all nouns. L2 participants relied on heuristic strategies based in part on the sublexical properties of nouns and therefore showed standard frequency by regularity interactions associated with learning.

The three connectionist models that were tested provide an alternative account to Carroll (1989) and others who subscribe to traditional computational models (e.g., Neidle, 1982). Models 1 and 2 provided a preliminary computational account of the difficulties inherent in learning noun gender classifications when morphophonological gender marking is ambiguous. The third set of simulations examined the possibility that gender is an implicit category that is encoded in networks of correlated inflectional patterns, as suggested by Maratsos (1988; see also MacWhinney et al., 1989; Taraban et al., 1989). The learning trajectories

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

evidenced in these rudimentary simulations justify further research into the possibility that L2 learners and L1 speakers are on the same learning continuum with respect to many basic aspects of correct grammatical usage.

Language learning does not typically take place in one trial. Rather, internal representations for lexical items, inflectional morphemes, pronominal forms, and so on (and the paradigms into which they fit) have to be acquired over time. Connectionist learning provides one computational account of how this might take place. In the present study, connectionist learning predicted that with sufficient learning the processing differences between transparent and opaque nouns should disappear, which was confirmed through the connectionist simulations.

One may wonder whether the results of the experiments and simulations simply provide another example of lexical frequency by regularity effects (e.g., Plaut et al., 1996). Certainly, the findings reported here extend that effect in new ways. First, the assumed noun frequency differences for L1 and L2 participants interacted with gender processes for other constituents. For L2 learners, disambiguating cues reduced the ambiguity of the gender of opaque nouns. Second, nonnominal gender markers also influenced gender marking on the main verb. In sum, frequency by regularity effects for nouns (e.g., Models 1 and 2) are found to be more intricately interwoven into other gender processes when examined in the context of sentences (e.g., Model 3).

The connectionist simulations also raise the possibility that length of experience may be the determining factor in learning a gender system. Specifically, L1 and L2 speakers may rely on similar learning mechanisms for mastering gender and simply may be at different points on the learning curve. This suggests that L1 and L2 learners may resemble one another on a computational level, contrary to traditional computational accounts (Carroll, 1989, 1995). Speculations like these deserve more attention, as well as the computational means for testing them, in order to begin making more connections between fluent native-like language competence and a long developmental history that leads to that fluency.

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

APPENDIX

Feminine opaque subject NP/Feminine transparent subject NP/Masculine filler (optional adjectives are in parentheses)

Только её (левая) голень немного выглядывала из-под одеяла.
Только её (левая) ступня немного выглядывала из-под одеяла.
Только её (левый) мизинец немного выглядывал из-под одеяла.

Конечно, (рассыпанная) фасоль повсюду лежала на полу.
Конечно, (рассыпанная) крупа повсюду лежала на полу.
Конечно, (рассыпанный) сахар повсюду лежал на полу.

И тогда (снежная) метель сильно разбушевалась над степью.
И тогда (снежная) вьюга сильно разбушевалась над степью.
И тогда (снежный) буран сильно разбушевался над степью.

Его (старая) трость обычно стояла в корридоре.
Его (старая) палка обычно стояла в корридоре.
Его (старый) топор обычно стоял в корридоре.

Наверное, (маленькая) щель незаметно образовалась во льду.
Наверное, (маленькая) дыра незаметно образовалась во льду.
Наверное, (маленький) проём незаметно образовался во льду.

Действительно, (новая) тетрадь открыто лежала на столе.
Действительно, (новая) брошюра открыто лежала на столе.
Действительно, (новый) сборник открыто лежал на столе.

Его (правая) бровь заметно распухла после драки.
Его (правая) щека заметно распухла после драки.
Его (правый) глаз заметно распух после драки.

Даже (обычная) соль теперь исчезла из магазинов.
Даже (обычная) мука теперь исчезла из магазинов.
Даже (обычный) рис теперь исчез из магазинов.

Его (единственная) лошадь тоже сдохла от голода.
Его (единственная) корова тоже сдохла от голода.
Его (единственный) кролик тоже сдох от голода.

И вот (ранняя) осень нежданно пришла в тундру.
И вот (ранняя) весна нежданно пришла в тундру.
И вот (ранний) мороз нежданно пришёл в тундру.

Его (раненная) грудь прекрасно зажила после операции.
Его (раненная) спина прекрасно зажила после операции.
Его (раненная) живот прекрасно зажил после операции.

Кажется, (деревянная) дверь недолго устояла под напором.
Кажется, (деревянная) стена недолго устояла под напором.
Кажется, (деревянный) забор недолго устоял под напором.

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

APPENDIX (*continued*)

Masculine opaque subject NP/Masculine transparent subject NP/Feminine filler (optional adjectives are in parentheses)

Вчера (последний) жёлудь наконец-то упал с ветки.
Вчера (последний) каштан наконец-то упал с ветки.
Вчера (последняя) вишня наконец-то упала с ветки.

Его (левый) туфель забыто валялся в углу.
Его (левый) башмак забыто валялся в углу.
Его (левая) варежка забыто валялась в углу.

И тогда (громкий) вопль вдруг донёсся из гостиной.
И тогда (громкий) плач вдруг донёсся из гостиной.
И тогда (громкая) песня вдруг донеслась из гостиной.

Слава Богу, (сильный) ливень вовремя прекратился к вечеру.
Слава Богу, (сильный) ураган вовремя прекратился к вечеру.
Слава Богу, (сильная) гроза вовремя прекратилась к вечеру.

Утром (одинокий) голубь внезапно появился над домом.
Утром (одинокий) ястреб внезапно появился над домом.
Утром (одинокая) синица внезапно появилась над домом.

Как всегда, (жареный) гусь ароматно дымился на блюде.
Как всегда, (жареный) заяц ароматно дымился на блюде.
Как всегда, (жареная) утка ароматно дымилась на блюде.

Накануне (отёкший) локоть сильно болел от удара.
Накануне (отёкший) палец сильно болел от удара.
Накануне (отёкшая) голова сильно болела от удара.

Его (старый) конь понуро волочился по дороге.
Его (старый) осёл понуро волочился по дороге.
Его (старый) овца понуро волочилась по дороге.

К счастью, (коричневый) портфель по-прежднему стоял под сиденьем.
К счастью, (коричневый) чемодан по-прежднему стоял под сиденьем.
К счастью, (коричневая) сумочка по-прежднему стояла под сиденьем.

Даже (большой) огонь окончательно погас к утру.
Даже (большой) костёр окончательно погас к утру.
Даже (большая) печка окончательно погасла к утру.

Наконец (нагруженний) корабль медленно ушёл из гавани.
Наконец (нагруженний) пароход медленно ушёл из гавани.
Наконец (нагруженная) шлюпка медленно ушла из гавани.

Кстати, (ожидаемий) дождь действительно начался после обеда.
Кстати, (ожидаемий) ветер действительно начался после обеда.
Кстати, (ожидаемая) буря действительно началась после обеда.

ACKNOWLEDGMENT

We acknowledge the contributions of Lisa Lima in a pilot study. Thanks are due to Michael Katz for assistance in recruiting participants for Experiment 2 and to Brian MacWhinney, Catherine Snow, and an anonymous reviewer for their helpful comments on an earlier draft of this article.

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

NOTES

1. Other statistics include cue availability, cue validity, and conflict validity. These are not important to the present discussion.
2. The spellings of end-palatalized consonants end in a **-ь**, which is referred to as a soft sign and is indicated with an apostrophe in the transcriptions.
3. There may be some exceptions to this generalization – for example, a noun ending in a palatalized (soft) sibilant is reliably feminine.
4. A frequency manipulation would have been very informative with respect to the issues at hand. Unfortunately, the limited number of end-palatalized nouns in Russian made it impossible to vary frequency systematically while at the same time maintaining semantic consistency across the versions of the sentence frames.
5. Note that the degrees of freedom for the denominator in the *min F'* statistic is sensitive to the *F1* and *F2* values.

REFERENCES

Bates, E., Devescovi, A., Pizzamiglio, L., D'Amico, S., & Hernandez, A. (1995). Gender and lexical access in Italian. *Perception & Psychophysics, 57,* 847–862.

Bates, E., & MacWhinney, B. (1989). *The crosslinguistic study of sentence processing.* New York: Cambridge University Press.

Braine, M. (1989). Modeling the acquisition of linguistic structure. In Y. Levy, I. Schlesinger, & M. Braine (Eds.), *Categories and processes in language acquisition.* Hillsdale, NJ: Erlbaum.

Bresnan, J. (Ed.). (1982). *The mental representation of grammatical relations.* Cambridge, MA: MIT Press.

Brooks, P., Braine, M., Catalano, L., & Brody, R. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32,* 76–95.

Bush, L., Hess, U., & Wolford, G. (1993). Transformations for within-subject designs: A Monte Carlo investigation. *Psychological Bulletin, 113,* 566–579.

Carroll, S. (1989). Second-language acquisition and the computational paradigm. *Language Learning, 39,* 535–594.

(1995). The hidden dangers of computer modelling: Remarks on Sokolik and Smith's connectionist learning model of French gender. *Second Language Research, 11,* 193–205.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Clark, H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335–359.

Colé, P., & Segui, J. (1994). Grammatical incongruency and vocabulary types. *Memory and Cognition, 22,* 387–394.

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

Corbett, G., & Fraser, N. (1993). Network morphology: A DATR account of Russian nominal inflection. *Journal of Linguistics, 29,* 113–142.

Ellis, N., & Schmidt, R. (1997). Morphology and longer distance dependencies. *Studies in Second Language Acquisition, 19,* 145–171.

Forster, K. (1979). Levels of processing and the structure of the language processor. In W. Cooper & E. Walker (Eds.), *Sentence processing* (pp. 27–85). Hillsdale, NJ: Erlbaum.

Grosjean, F., Dommergues, J., Cornu, E., Guillelmon, D., & Besson, C. (1994). The gender-marking effect in spoken word recognition. *Perception & Psychophysics, 56,* 590–598.

Gurjanov, M., Lukatela, G., Lukatela, K., Savić, M., & Turvey, M. (1985). Grammatical priming of inflected nouns by the gender of possessive adjectives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 692–701.

Gurjanov, M., Lukatela, G., Moskovljević, J., Savić, M., & Turvey, M. (1985). Grammatical priming of inflected nouns by inflected adjectives. *Cognition, 19,* 55–71.

Gvozdev, A. N. (1961). *Voprosy izuchenija detskoy rechi.* Moscow: Izdatelstvo Pedagogicheskikh Nauk.

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press

Kempe, V., & MacWhinney, B. (1998). The acquisition of case marking by adult learners of Russian and German. *Studies in Second Language Acquisition, 20,* 543–588.

Kucera, H., & Monroe, G. (1968). *A comparative quantitative phonology of Russian, Czech, and German.* New York: Elsevier.

Levelt, W. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Luce, R. (1959). *Individual choice behavior.* New York: Wiley.

Lukatela, G., Kostić, A., Todorović, D., Carello, C., & Turvey, M. (1987). Type and number of violations and the grammatical congruency effect in lexical decision. *Psychological Research, 49,* 37–43.

MacWhinney, B. (1987a). Applying the competition model to bilingualism. *Applied Psycholinguistics, 8,* 315–327.

(1987b). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Erlbaum.

(1989). Competition and lexical categorization. In R. Corrigan, F. Eckman, & M. Noonan (Eds.), *Linguistic categorization* (pp. 195–242). Philadelphia: Benjamins.

MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language, 28,* 255–277.

Maratsos, M. (1988). The acquisition of formal word classes. In Y. Levy, I. Schlesinger, & M. Braine (Eds.), *Categories and processes in language acquisition* (pp. 31–44). Hillsdale, NJ: Erlbaum.

McClelland, J., & Rumelhart, D. (1988). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises.* Cambridge, MA: MIT Press.

McDonald, J., & Heilenman, K. (1991). Determinants of cue strength in adult first and second language speakers of French. *Applied Psycholinguistics, 12,* 313–348.

Neidle, C. (1982). Case agreement in Russian. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 391–426). Cambridge, MA: MIT Press.

Plaut, D., McClelland, J., Seidenberg, M., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

Pulkina, I. (1978). *A short Russian reference grammar.* Moscow: Russian Language Publishers.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D. Rumelhart, J. McClelland, & The PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.

Schriefers, H. (1992). Lexical access in the production of noun phrases. *Cognition, 45,* 33–54.

Sokolik, M., & Smith, M. (1992). Assignment of gender to French nouns in primary and secondary language: A connectionist model. *Second Language Research, 8,* 39–58.

SPSS Inc. (1991). *Statistical algorithms.* Chicago: Author.

Taraban, R., McDonald, J., & MacWhinney, B. (1989). Category learning in a connectionist model: Learning to decline the German definite article. In R. Corrigan, F. Eckman, & M. Noonan (Eds.), *Linguistic categorization* (pp. 163–194). Philadelphia: Benjamins.

Taraban, R., & Roark, B. (1996). Competition in language-based categories. *Applied Psycholinguistics, 17,* 125–148.

Tucker, G., Lambert, W., & Rigault, A. (1977). *The French speaker's skill with grammatical gender: An example of rule-governed behavior.* The Hague: Mouton.

Winer, B. (1962). *Statistical principles in experimental design.* New York: McGraw-Hill.

Zasorina, L. (Ed.). (1977). *Chastotnyj slovar russkogo jazyka.* Moscow: Russkij Jazyk Publishers.

Zubin, D., & Koepcke, K. (1986). Gender and folk taxonomy: The indexical relation between grammatical and lexical categorization. In C. Craig (Ed.), *Noun classes and categorization.* Amsterdam: Benjamins.

https://doi.org/10.1017/S0142716499001046 Published online by Cambridge University Press