SSLA, **20**, 543–587. Printed in the United States of America.

THE ACQUISITION OF CASE MARKING BY ADULT LEARNERS OF RUSSIAN AND GERMAN

Vera Kempe
University of Toledo

Brian MacWhinney
Carnegie Mellon University

This study investigated the acquisition of the comprehension of overt morphological case marking by adult native speakers of English who were learning Russian or German as a second language (L2). The Russian case-marking system is more complex than the German system, but it almost always provides the listener with case inflections that are reliable cues to sentence interpretation. Two approaches to learning of inflectional morphology were contrasted: the rule-based approach, which predicts that learning is determined by paradigm complexity; and the associative approach, which predicts that learning is determined by the cue validity of individual inflections. A computerized picture-choice task probed the comprehension of L2 learners by varying the cues of case marking, noun configuration, and noun animacy. The results demonstrated that learners of Russian use case marking much earlier than learners of German and that learners of German rely more on animacy to supplement the weaker case-marking cue. In order to further explore the underlying mechanisms of learning, a connectionist model was developed that correctly simulated the obtained results. Together, these findings support the view that adult L2 learning is associative and driven by the validity of cues in the input.

This research was supported by a fellowship from the German Academic Exchange Council (DAAD) to the first author. We wish to thank Patricia J. Brooks and Roman Taraban for helpful comments on an earlier version of this paper. Part of this work has been presented at the 1994 Meeting of the Psychonomic Society.

Correspondence concerning this article should be addressed to Vera Kempe, Department of Psychology, Mahar Hall, SUNY Oswego, 7060 State Route 104, Oswego, NY 13126–3599.

Inflectional morphology represents an ideal testing ground for the controversy between rule-based and associative approaches toward language learning. So far, however, most studies addressing this controversy have focused on the interplay between regular versus irregular forms. This contrast is relevant for English morphology but does not do justice to the complexity of inflectional systems found in many other languages. Furthermore, most research on the processing of inflectional morphology has been concerned with L1 acquisition. As a result, inflectional processing in adult L2 learning has not received much attention. In this study, we investigate the acquisition of morphological case marking in adult L2 learners in order to determine which factors guide the learning process.

LEARNING OF INFLECTIONAL PARADIGMS

Traditional language teaching places an emphasis on the regularities underlying the inflectional system of a foreign language (Matthews, 1974). In psycholinguistics, this emphasis has been echoed in work describing inflectional morphology in terms of a system of rules that is assumed to be part of a speaker's language competence (Marcus et al., 1995; Pinker, 1984; Pinker & Prince, 1988). The approaches to paradigm formation developed by MacWhinney (1978) and Pinker (1984) view the learning of inflections as a process of discovering the grammatical dimensions underlying an inflectional paradigm (e.g., number, gender, person, case, or tense) through systematic hypothesis testing. For agglutinative languages, in which a single inflection corresponds to a single dimension and inflections can be combined additively, paradigm learning involves a gradual expansion from a unidimensional to a multidimensional paradigm. Thus, the learner starts out with an initial hypothesis about the relation of a certain inflection to a grammatical dimension. Gradually, more hypotheses about the associations between inflections and grammatical dimensions are introduced. In fusional languages, inflections are associated with combinations of grammatical dimensions like *nominative singular* or *third person plural*, which correspond to the different cells of an inflectional paradigm. Here, the hypothesis-testing process must be applied iteratively by adding dimensions until the correct number is found. Thus, the learner starts out with a simplified hypothesis about what dimension is marked by inflectional change. As the learner notices that the paradigm exhibits more inflectional change than is predicted by this simple hypothesis, he or she must add more dimensions and test whether the paradigm conforms to the adjusted dimension space. The postulation of such an iterative hypothesis-testing process leads to the prediction that the more complex a paradigm, the longer it should take to learn it.

It is important to point out that the discovery of the correct grammatical dimensions is hindered by the fact that, in many paradigms, the same inflection may be associated with multiple cells in a paradigm. Thus, not always are all combinations of grammatical dimensions marked by unique inflections. Sometimes underlying dimensions are partly neutralized in the paradigm. Sometimes different cells are marked syncretically by the same inflection. Neutralization and inflectional syncretism obscure the underlying dimensions so that the learner has to rely on large sets of data in order to discover regularities. According to the theory of paradigm formation, the difficulty of learning a paradigm depends not only on the number of dimensions and cells but also on the amount of neutralization or inflectional syncretism.

The alternative view on the acquisition of inflections minimizes the importance of rule learning in favor of associative learning. This approach has been put forth within a variety of frameworks such as the Network Model (Bybee, 1985, 1995), the Competition Model (MacWhinney & Bates, 1989), and various connectionist models (e.g., Rumelhart & McClelland, 1987; MacWhinney, Leinbach, Taraban, & McDonald, 1989; Plunkett & Marchman, 1991). Common to all these frameworks is the view of paradigms as epiphenomena that emerge from distributional characteristics in the language input. In these models, the notion of rules is abandoned and learning is assumed to take place by gradual strengthening of the association between co-occurring elements of the language. Whereas the Network Model and the connectionist models focus on the structure of the lexicon and the intralexical relations relevant for learning and representing inflectional paradigms, the Competition Model focuses on inflections as cues to underlying thematic roles and pragmatic functions. It is this functional perspective that will be adopted in the present study. Moreover, we will demonstrate below that the functional perspective lends itself to connectionist simulations in a straightforward way.

Within the framework of the Competition Model, the ease of learning an inflection is determined by its *cue validity*. Cue validity, in turn, depends on how often an inflection occurs as a cue for a certain underlying function (*cue availability*) and how reliably it marks this function (*cue reliability*). Learning is assumed to progress more quickly if an inflection is readily available in the language. But availability alone does not facilitate learning if the inflection cannot reliably be associated with a certain underlying function.

Cue validity, as defined in the Competition Model, can be quantified using data from language corpora in order to predict cue strength as measured in empirical tests. The present study will compare predictions derived from the Competition Model with predictions derived from the rule-based approach. In the first part of the paper, we will contrast a corpus-based measure of cue validity with an estimate of paradigm complexity and determine which one is better suited to explain the empirical learning trajectories. In the second part, we will use connectionist modeling to specify in greater detail the ways in which cue availability and cue reliability interact.[1] The performance of this model will then be compared to the empirical data in order to explore whether connectionist learning models constitute a viable approach to aspects of L2 learning.

SECOND LANGUAGE LEARNING

Although researchers are willing to question the role of rules in L1 learning, there seems to be a greater willingness to accept the importance of formal paradigm learning for adult L2 learners. This is because adult L2 learners are more likely to utilize a mature set of cognitive abilities, including metalinguistic knowledge and problem-solving strategies that are not available to the child. However, the effect of these factors in adult learning is often viewed as inhibitory. It has been argued (Felix, 1987; Johnson & Newport, 1989; Krashen, 1981) that the use of problem-solving strategies and formal rules by adult learners is responsible for their inability to acquire a fully natural command over the L2. For example, in a framework that assumes a Universal Grammar (UG), the increasing reliance on problem-solving strategies in adult language learning is often attributed to the decay of the mechanisms that had been operative in L1 learning during the critical period (Clahsen & Muysken, 1986; Schwartz, 1993). Within a UG framework, adult L2 learning differs radically from L1 acquisition in children. If rule-based learning is indeed the prevailing learning mechanism in adults, then we would expect that the processing of paradigmatically complex inflectional markings would pose a particular challenge for the adult L2 learner.

The associative approach, on the other hand, views both L1 and L2 learning as input-based and assumes that cue validity determines the learner's ultimate success in acquiring a paradigm (MacWhinney, 1987a; MacWhinney & Chang, 1995). No principled qualitative difference between L1 and L2 learning is postulated. If it is possible to show that adult L2 learning depends more on cue validity than on paradigm complexity, this would suggest that a strong input-based, associative component is also evident in adult L2 learning. The outcome of the debate between those who emphasize the role of the input and those who view adult L2 learning as rule-based is of great importance to L2 theory as well as instructional methodology (Ellis, 1994).

THE CROSS-LINGUISTIC APPROACH

The concepts of paradigm complexity and cue validity highlight very different aspects of inflectional morphology. The former refers to the configuration of inflections in a multidimensional paradigm space. The latter emphasizes the strength of the association between an inflection and an underlying function. In many instances, complexity and cue validity may be orthogonally related to each other. In other words, regardless of whether a paradigm is complex or simple (i.e., based on a large or small number of dimensions), the validity of each individual inflection may be higher or lower depending on whether this inflection marks underlying functions consistently and whether it occurs with sufficient frequency in the language. Two research strategies can be chosen to contrast the effects of paradigm complexity and cue validity on L2 learning: the construction and investigation of learning of artificial languages; or the cross-linguistic comparison of L2 learning, which exploits the natural variation between languages. In this study, we have opted for the cross-linguistic approach because of the higher ecological validity of investigating natural languages.

In order to assess the independent effects of paradigm complexity and cue validity, one would ideally aim at a research design that permits a full crossing of these two factors. For just two levels of complexity and two levels of cue validity, at least four different inflectional paradigms need to be investigated empirically. However, natural variation in complexity and cue validity is gradual and confounded with a host of other variables. This makes the selection of four suitable languages very difficult. In the present study, we start with the comparison of two languages, while recognizing the need to extend this comparison to additional languages of diverse typological forms. We investigate the learning of the Russian and German declension paradigms by native speakers of English. In order to ensure comparability of these rather different paradigms, we restrict ourselves to the investigation of the nominative and the accusative case in sentences in which these morphological cases express the same underlying functions in both languages. We will show that the German paradigm is of lower complexity but that the individual inflections are also low in cue validity. If paradigm complexity is the main factor in learning, German case marking should be acquired faster than Russian case marking. We will also show that the Russian paradigm is higher in complexity but that the individual inflections are also higher in cue validity. If cue validity guides learning, Russian case marking should be learned faster. We then report an experiment that is designed to evaluate these contrasting hypotheses.

THE INFLECTIONAL PARADIGMS OF RUSSIAN AND GERMAN

Tables 1 and 2 display the inflectional paradigms of Russian and German. Russian (Table 1) has six cases: nominative, genitive, dative, accusative, instrumental, and prepositional. Nouns are marked by a set of suffixes that code simultaneously for the morphosyntactic and semantic dimensions of case, number, gender, and animacy. Note that some of the cells contain multiple suffixes. This reflects the fact that Russian nouns fall into different declension types that do not entirely conform to the gender distinction. Russian has two declension types for feminine nouns but only one declension type for masculine and neuter nouns. Gender differences between the masculine and neuter nouns become apparent only in the nominative and accusative cases. The German system (Table 2) is less complex in terms of underlying dimensions. German has three genders and two numbers but only four cases. In German, marking is conveyed primarily by the article and sometimes by combinations of the article and a suffix on the noun.

As can be seen from Tables 1 and 2, both languages exhibit a fairly extensive pattern of neutralization. For example, gender is entirely neutralized in the German plural. Animacy is entirely neutralized in all Russian cases except

Table 1. Russian case-marking paradigm

Singular

Case	Feminine Animate		Feminine Inanimate		Masculine Animate	Masculine Inanimate	Neuter Animate	Neuter Inanimate
Nom	a/ja	'ø	a/ja	'ø	ø/'ø	ø/'ø	—	o
Gen	i/y	i	i/y	i	a/ja	a/ja	—	a/ja
Dat	e	i	e	i	u/ju	u/ju	—	u/ju
Acc	u/ju	'ø	u/ju	'ø	a/ja	ø/'ø	—	o/'o/e
Instr	oj/'oj/ej	'ju	oj/'oj/ej	'ju	om/'om/em	om/'om/em	—	om/'om/em
Prep	e	i	e	i	e	e	—	e

Plural

Case	Feminine Animate	Feminine Inanimate	Masculine Animate	Masculine Inanimate	Neuter Animate	Neuter Inanimate
Nom	i/y	i/y	a/ja i/y	a/ja i/y	—	a/ja
Gen	ej ø/'ø	ej ø/'ø	ej ø/'ø ov/'ov/ev	ej ø/'ø ov/'ov/ev	—	ej ø/'ø ov/'ov/ev
Dat	am/jam	am/jam	am/jam	am/jam	—	am/jam
Acc	ej ø/'ø	i/y	ej ø/'ø ov/'ov/ev	a/ja i/y	—	a/ja
Instr	ami/jami	ami/jami	ami/jami	ami/jami	—	ami/jami
Prep	ax/jax	ax/jax	ax/jax	ax/jax	—	ax/jax

Note. The table represents the most common declension types in Russian. Suffixes separated by a slash are allomorphs. The apostrophe denotes palatalization of a preceding consonant. The symbol ø denotes null morphemes.

the accusative for masculine nouns. Nominative and accusative case are neutralized in feminine, neuter, and plural nouns in German and in feminine *i*-declension nouns as well as in inanimate masculine and plural nouns in Russian.

Quantifying Paradigm Complexity

In order to quantify complexity in a way that conforms to the rule-based paradigm formation theories (MacWhinney, 1978; Pinker, 1984), we can isolate three factors that determine the complexity of a paradigm: the number of dimensions, the number of cells, and to what extent the cells in the paradigm are marked by unique inflections. First, the number of dimensions determines how many cycles of hypothesis testing have to be passed through during the course of learning. According to this metric, the Russian system is more complex because animacy is a relevant dimension in addition to number, gender, and case. Secondly, the number of levels on each dimension determines the number of different cells in the paradigm that have to be learned and memorized. In the Russian system, the complete crossing of six cases, two numbers,

Table 2. German case-marking paradigm

	Singular					
Case	Feminine		Masculine		Neuter	
Nom	die	-ø	der	-ø	das	-ø
Gen	der	-ø	des	-s/n	des	-s
Dat	der	-ø	dem	-ø/n/(-e[a])	dem	-ø/(e[a])
Acc	die	-ø	den	-ø/n	das	-ø
	Plural					
Case	Feminine		Masculine		Neuter	
Nom	die	-(PL)	die	-(PL)	die	-(PL)
Gen	der	-(PL)	der	-(PL)	der	-(PL)
Dat	den	-n/(PL)	den	-n/(PL)	den	-n/(PL)
Acc	die	-(PL)	die	-(PL)	die	-(PL)

Note. The symbol ø denotes null morphemes. PL denotes the plural morpheme on the noun for example *-e*, *-(e)n*, *-er*, *-s*, *-ø*. Umlaut-changes in the stem vowel are not related to case marking and, therefore, are not further considered.
[a]The dative inflection *-e* as in *dem Volke* or *dem Manne* is optional and rather archaic.

three genders, and two levels of animacy yields as many as 72 cells. This is clearly more complex than the German system, which has only 24 cells.[2] Finally, the average uniqueness of the inflections can be estimated as the ratio of inflections to cells. The closer to zero the inflections-to-cells ratio, the higher the amount of neutralization or syncretism and the lower the average uniqueness of inflections. If phonologically governed allomorphy in Russian is disregarded, the total number of unique inflections is 15, which yields a ratio of .21. For German, the total number of unique article-suffix combinations is 12, which yields a ratio of .5. This is a clear indicator that the uniqueness of each inflection is lower in Russian, which makes the discovery of the dimensions of the paradigm more difficult.[3]

In sum, Russian appears to have the more complex system by all three paradigm-based complexity measures. Because the study below focuses on nominative/accusative marking, we need to look specifically at the complexity estimations for the reduced nominative/accusative paradigm: The number of dimensions is the same as in the full paradigm. The number of possible cells for this part of the paradigm is 24 in Russian and 12 in German. The inflections-to-cells ratio is .29 in Russian and .33 in German. Thus, the estimations for the reduced paradigm confirm that nominative/accusative marking is more complex in Russian than in German. If the acquisition of new cues is determined by the complexity of the underlying paradigm, learners of German should do far better than learners of Russian in picking up nominative and accusative marking in the new language.

Quantifying Cue Validity

The Competition Model proposes two factors that determine the validity of a cue: its availability in the language input and the reliability with which it allows the language user to access the underlying function. German and Russian differ in the extent to which they provide nominative and accusative markers as cues for agents and objects in sentences. In order to measure these differences in the input of language learners, we performed a corpus analysis that allowed us to estimate the validity of nominative and accusative markers in the context of other surface cues such as word order, animacy of the nouns, and verb agreement. In each language, we analyzed a corpus of active transitive sentences from five widely used textbooks. The references for the textbooks are given in Appendix A. In each textbook, the number of pages was divided by 40 and all sentences on each *n*/40 page were examined. This resulted in a sample of 560 Russian and 671 German sentences. For Russian, negative sentences permitting both accusative and genitive marking of the direct object were excluded, as were transitive sentences with subject omission.

In estimating availability and reliability, we followed the principles applied by McDonald and Heilenmann (1991) in their corpus analysis for French. Availability of a cue was computed as the number of sentences in which a cue is present, divided by the total number of transitive sentences. Note that availability differs from the notion of frequency in that it refers to the presence of a cue as marker of a particular function. For example, availability of the animacy cue does not refer to the frequency of animate nouns but rather to the frequency of sentences in which the cue is contrastive; that is, one noun is animate and the other one is not. Reliability of a cue is computed as the ratio of sentences in which a cue correctly indicates the agent, divided by the number of sentences in which the cue is present. Validity is obtained by simply multiplying availability and reliability. Table 3 presents the availability, reliability, and validity estimations for word order, animacy contrast, case marking (nominative vs. accusative), and verb agreement in the Russian and German textbook corpora. Word order refers to the particular configuration of the two nouns in a simple transitive sentence regardless of the position of the verb. Thus, if the first noun was the agent, then the configuration cue was counted as reliable. If the second noun was the agent, then the configuration cue was counted as not reliable. In order to avoid confusion with other Competition Model studies in which word order refers to variations of the position of the verb, this cue will henceforth be called noun configuration.

Figure 1 depicts the hierarchy of cues as estimated from the German and Russian L2 textbooks. It demonstrates that the validity of case marking (nominative and accusative marking combined) is much higher in Russian (.97) than in German (.56). This difference is due primarily to differences in availability, because case markers are always reliable cues in the context of transitive sentences. The illustration also shows that German and Russian have very similar

Table 3. Availability, reliability, and overall validity of various cues in German and Russian L2 textbooks

	German number of sentences = 671 5% ambiguous sentences			Russian number of sentences = 560 1% ambiguous sentences		
Cue	Availability	Reliability	Validity	Availability	Reliability	Validity
Configuration						
(Agent first)	1.000	.891	.891	1.000	.895	.895
SVO	.413	.841	.347	.853	.958	.817
SOV	.356	.879	.313	.123	.520	.064
VSO	.231	1.000	.231	.015	.866	.013
Animacy contrast	.785	.990	.770	.791	.981	.776
Case marking						
(Total)	.563	1.000	.563	.968	1.000	.968
Nominative	.467	1.000	.467	.947	1.000	.947
Accusative	.203	1.000	.203	.400	1.000	.400
Verb agreement	.562	1.000	.562	.684	1.000	.684

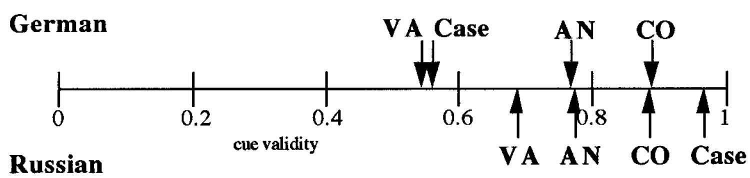

Figure 1. Hierarchy of cues in the German and Russian L2 textbooks (VA, verb agreement; AN, animacy contrast; Case, case marking; CO, noun configuration).

cue validity for animacy contrast and configuration. It is not surprising to find similar cue validity for animacy, because availability and reliability of this cue do not depend on language-specific features but more on universal aspects of language use related to the distribution of animate and inanimate nouns in discourse. The similar levels of cue validity for the configuration cue are due to the fact that, in German and Russian, subjects precede objects in the unmarked, canonical word order.

Major differences exist between the two languages in terms of the cue validities of case marking and verb agreement. Both of these cues have a markedly lower validity in German. For verb agreement, the low validity in German is due to the neutralization of number marking in parts of the German paradigm (MacWhinney, Bates, & Kliegl, 1984). For case marking, the low validity in German is due to the higher amount of nominative/accusative neutralization. In order to understand why this holds even though the average uniqueness of inflections in German was estimated as higher, it is necessary to distinguish between neutralization and syncretism. Neutralization occurs when different levels *within* a dimension are marked by the same inflection, as in the nominative and the accusative cases of German feminine, neuter, and plural nouns. Syncretism refers to the marking of *orthogonal* combinations of dimensions by the same inflection, as in the case of the Russian suffix *-u* which mainly marks either singular feminine accusative nouns or singular masculine dative nouns. The important difference is that neutralization always results in ambiguity whereas syncretism does not. The lower average uniqueness of inflections in the Russian paradigm simply reflects the fact that animacy is always neutralized except in masculine accusative and plural accusative nouns. However, when it comes to the dimension of case, the German paradigm shows more neutralization than the Russian paradigm. Consequently, in German, in order for a transitive sentence to be unambiguously case marked, it must contain at least one masculine noun. Our corpus analysis revealed that 44% of German transitive sentences do not contain masculine nouns and, thus, have no clear nominative or accusative marking. In Russian, on the other hand, neutralization of the nominative/accusative distinction occurs predominantly in inanimate nouns and in a limited set of end-palatalized feminine nouns. However, only 5% of all Russian transitive sentences contain two inanimate nouns. The remaining sentences contain animate nouns, thus making case marking a highly valid cue.

In sum, the corpus-based estimations confirm that the cue validity of nominative and accusative case marking is much higher in Russian than in German. If case markers are learned by gradually strengthening the associations between case markers on one hand and thematic roles on the other hand, then the higher availability of case marking in Russian should result in faster learning of case marking than in German.

STUDY 1: COMPREHENSION OF CASE MARKING

We used a speeded picture-choice task to assess the strength of sentence-comprehension cues in learners of German and Russian. In this task, participants typically listen to simple noun-verb-noun sentences and have to decide as fast as possible which noun refers to the agent of the sentence. The picture-choice task has been widely applied in cross-linguistic research within the Competition Model framework. Studies of L2 learners (Gass, 1987; Harrington, 1987; Kilborn, 1989; Kilborn & Cooreman, 1987; Liu, Bates, & Li, 1992; MacWhinney, 1987a, 1992; McDonald, 1987b; Sasaki, 1991) have demonstrated that, in the early stages, the learner tries to transfer L1 cues to the L2. However, often morphological cues cannot be transferred because there is no match between L1 and L2 in terms of form or function. On the other hand, cues such as animacy, word order, or noun configuration can often be mapped across languages. For these cues, we see evidence of transfer in the initial stages of L2 learning. Research on native speakers of English has shown that word order is the dominant cue in determining the agent in transitive sentences (MacWhinney, Bates, & Kliegl, 1984). Thus, in noun-verb-noun sentences, English speakers consistently choose the first noun as the agent. An L2 learner who starts out with this strategy will be successful in processing L2 sentences that conform with the "agent first" configuration. However, this learner will fail dramatically in L2 sentences that deviate from this pattern. The considerable variability in German and Russian word order allows us to construct grammatical OVS sentences with agents in postverbal position. This configuration may be marked by morphological cues like case marking. Consequently, performance in case-marked OVS sentences will serve as the critical dependent variable that is informative with respect to a learner's mastery of case marking in the L2. For this variable, we can formulate the following predictions: If the acquisition of case marking is predominantly determined by paradigm complexity, we expect fewer errors in case-marked OVS sentences by learners of German. On the other hand, if cue validity guides the process, learners of Russian should exhibit lower error rates in case-marked OVS sentences.

In order to fully understand the processes underlying sentence comprehension, we also need to consider on-line measures of performance. Off-line measures in nonspeeded tasks do not allow us to draw conclusions about the immediate effects of the various cues during processing (Hernandez, Bates, & Avila, 1994; Li, Bates, & MacWhinney, 1993). The speeded task that we use allows us to tap into the time course of sentence processing and provides more detailed information about the on-line effects of various cues. Previous studies of on-line decision latencies in the Competition Model framework (Hernandez, Bates, & Avila, 1994; Kilborn, 1989; Li, Bates, & MacWhinney, 1993; Mimica, Sullivan, & Smith, 1994) have demonstrated that strong cues lead to faster latencies and that cue competition leads to slower latencies. However, there is also evidence (Kail, 1989; Mimica, Sullivan, & Smith, 1994) that processing may be slowed down by the presence of additional cues, even if each of the additional cues is very strong.

Based on these studies, we can formulate specific hypotheses concerning processing speed: First, a strong bias to select the first noun as the agent should lead to fast but incorrect performance in OVS sentences. This is because the learner does not yet expect to encounter any information that might modify this bias, particularly when the relevant cues occur later in the sentence. The stronger these other cues become during the course of learning, the slower the latencies will be, because interpretation will be delayed until more information is gathered from later parts of the sentence. Consequently, at the earliest stages of L2 acquisition, increasing mastery of case marking should lead to a slowing down of performance. Later in learning, decisions will speed up again, because the processing of case markers will become more automatized. If case marking appears early in the sentence, performance may be relatively fast in the advanced stages of learning. Thus, we expect an inverted u-shaped relationship between processing speed and mastery of case marking with fast performance at the outset of learning, relatively slow performance at intermediate stages, and fast performance at higher stages. Language differences in the trajectories of processing speed provide converging evidence as to which learner group acquires case marking faster.

Second, cue validity should affect on-line processing. Very strong cues should lead to larger processing benefits than weaker cues when compared to sentences that do not contain these cues. If case marking has an immediate effect in sentence processing, the processing benefits associated with case markers should be larger in the Russian learner group. Furthermore, the weaker the case-marking cue, the more other cues will be considered during on-line processing. Thus, we can expect to see stronger benefits from noun configuration, or animacy, or both, in the learners of German.

Cross-linguistic comparisons of L2 performance are complicated by a variety of factors. With respect to learning Russian and German, it is important to take into account the fact that speakers of English who learn Russian are faced with the task of acquiring the Cyrillic writing system in order to be able to process written input. Therefore, they are likely to have processed less written input over the same amount of time than learners of German. In order to match the two learner groups, it is necessary to find a valid way of comparing overall language familiarity between different learners as well as between different languages. Because no comparable proficiency tests for Russian and German are available, we constructed our own L2 lexical decision task. The use of this paradigm was based on the assumption that the more L2 words a learner can recognize, the more L2 input has been processed by this learner. Using Signal Detection Theory (Green & Swets, 1966), we calculated d' (d-prime) scores. These scores provide an estimate of language familiarity independent of exposure time. We used this measure as a covariate for the comparison of sentence comprehension performance across languages and across learners. Design and evaluation of the lexical decision task are reported in detail in Kempe and MacWhinney (1996).

Method

Participants. A total of 44 L2 learners were recruited by advertisement at colleges and universities in Pittsburgh and were paid $5.00 for participation. This pool included 22 learners of German (13 men, 9 women) and 22 learners of Russian (12 men, 10 women). All learners were either college students or recent college graduates. Both groups were remarkably homogeneous with respect to age; the mean age of each group was 21.9 years with a range from 17 years to 29 years in both groups.

Materials and Design.

Lexical Decision Task. A total of 362 words and nonwords was selected in each language. To ensure comparability between the Russian and German materials, words were selected on the basis of normative frequency counts for each language (Baayen, Piepenbrock, & van Rijn, 1993; Zasorina, 1977). The raw frequency values provided by these sources were converted into logarithms of the word frequency per million, and the entire frequency continuum was divided into 240 equal units. At each unit, the closest frequency-matching word was selected in both languages. This sampling and matching algorithm was applied in ascending word frequency until the very high-frequency region where exact matches do not exist. The outcome of this sampling procedure was a pool of 181 words in each language, with frequency increasing steadily on a logarithmic scale. Additionally, all frequency-matched words were also matched for word class across languages. The use of words bearing any resemblance to English was thoroughly avoided. Finally, each word was paired with a nonword that obeyed the phonotactic rules of the language and matched its word counterpart in terms of overall phonological shape, derivational type, and word class.[4]

All words and nonwords were randomized individually for each participant and presented in black characters on a light blue screen of a Macintosh Centris 660AV covering a visual angle of approximately 2.5 degrees.

Picture-Choice Task. The stimulus materials consisted of simple active transitive noun-verb-noun sentences that were grammatically correct in Russian and German. Within each language, the sentences varied according to the following five factors: animacy of the first noun (N1-Animacy: animate vs. inanimate), animacy of the second noun (N2-Animacy: animate vs. inanimate), case marking of the first noun (N1-Marking: reliably marked vs. neutralized), case marking of the second noun (N2-Marking: reliably marked vs. neutralized), and noun configuration (SVO vs. OVS). In this design, the configuration factor specifies the type of marking on the two nouns. For SVO sentences, case marking on the first noun was nominative and on the second noun it was accusative. In OVS sentences, the pattern was the opposite; case marking on the first noun was accusative and on the second noun it was nominative. If the sentence does not contain a reliable case marker, there is no structural difference between the SVO and OVS conditions so that the configuration manipulation is neutralized for these sentences, thereby rendering the conditions structurally identical. Language was a between-subjects factor.

The sentences were composed from combinations of four animate nouns, four inanimate nouns, and two verbs. The nouns and their English translation equivalents are given in Appendix B. In order to minimize semantic differences between languages, exact translation equivalents for all 10 words were used so that the same words appeared in the same conditions in Russian and in German. Thus, the German translation equivalents of "mother," "daughter," "flower," and "cake" are feminine resulting in nominative/accusative neutralization. In the Russian counterparts, there is also nominative/accusative neutralization in the translation equivalents of "mother" and "daughter" because they exhibit final consonant palatalization, and in the translation equivalents of "flower" and "cake" because they are masculine inanimate. The pictures of the four inanimate nouns were taken from the Snodgrass and Vanderwart (1980) materials. The pictures of the four animate nouns were taken from other sources. The verbs were "looking for" and "finding." All nouns were singular and all verbs were presented in the present tense so that verb agreement was not available as a cue.

For each condition, four sentences were constructed by first combining the relevant nouns and then counterbalancing the two verbs. This resulted in a total of 128 grammatically correct sentences. Half of the sentences were semantically implausible, similar to English sentences like "The cake finds the father." In these sentences, animacy is in competition with case marking or noun configuration or both. The inclusion of semantically implausible sentences was necessary for a full crossing of the two levels of animacy of the two nouns. Appendix C lists examples of sentences for each condition.

The eight nouns and two verbs were digitized with a 22-kHz sampling rate by a female native speaker and then combined into the 128 experimental sentences using SoundEdit16. Combining single word recordings into sentences ensured that the intonation pattern was identical for all sentences and that no prosodic cues were available to the listener. The digitized sentences were presented in randomized order.

Procedure. All participants were tested individually. They were seated in front of the computer and given the L2 lexical decision task. Participants were instructed to read each item and to decide whether they knew the word or not. They were asked to press the left button if they did not know the word and the right button if they knew the word. The word remained visible on the screen until the participant made a decision. Each trial was followed by a 200-ms inter-trial interval. The participants' responses and response latencies were registered. The completion of the lexical decision task required approximately 15 minutes.

After the lexical decision task, participants were given a 5-minute break, which was followed by the familiarization phase. Each picture was presented one at a time in the middle of the screen. The participants were asked how they would name the picture in the foreign language. After the participants had completed their answer, the correct name was displayed and participants were told that only this name would be used to label the picture through the entire experiment. During the second pass, participants again named the pictures and verified their answers by comparing them with the label. Next, participants saw pairs of pictures on the screen. As soon as the pictures were displayed, the name of one of the two pictures was presented through headphones. Participants placed their left and right index fingers on the left and right buttons of the button box and were instructed to press the button corresponding to the named picture as fast as possible. If the right picture was named, the right button had to be pressed and vice versa. The purpose of this preparation phase was to acquaint the participants with the speeded forced-choice task and to further familiarize them with the picture names. The participants saw all 54 possible pairwise combinations of the eight pictures.

In the main experiment, participants were told that they were going to hear a series of simple transitive sentences accompanied by the pictures of both nouns. Their task was to choose the agent of each sentence as fast as possible by pressing the spatially closer button. Agenthood was explained as “who or what did the looking or finding.” Again, if the picture on the right side of the screen depicted the agent of the sentence, the right button had to be pressed and vice versa. Participants were told that they should make a choice regardless of whether the sentence made sense to them or not. For each participant, 10 practice sentences were selected randomly from the larger pool of sentences. After the 10 practice trials, all 128 sentences were presented. Participants' choices and decision latencies were registered. All instructions were given in English. The entire session lasted approximately 45 minutes.

The presentation of the stimuli in the lexical decision task and in the picture-choice task was controlled by the PsyScope experimental control program (Cohen, MacWhinney, Flatt, & Provost, 1993). Latencies were measured to the nearest millisecond using the CMU-button box timer, in which three response buttons are arranged horizontally. Only the left and right buttons were used in this experiment.

Results

The presentation of the results will be structured as follows: First, we will describe the basic results of the lexical decision task. Second, we will describe a series of regression analyses. The first regression analysis is designed to determine the relationship between language familiarity and case-marking mastery for each language in order to establish which learner group is faster in the acquisition of case marking. The second and third regression analyses are designed to determine the relationship between language familiarity, case-marking mastery, and processing speed. Finally, we will present a series of ANOVAs and ANCOVAs performed on a subgroup of learners that was matched for language familiarity in order to compare the on-line effects of individual cues across languages.

Because sentences were presented auditorily, latencies were affected not only by sentence processing but also by the duration of the stimuli. The duration of verbs and nouns versus article-noun combinations was different for each sentence. This presents problems for the statistical analyses. In order to provide the reader with a better understanding of this problem, Figure 2 illustrates the time course of the stimuli in Russian and German. Generally, the Russian sentences tended to be shorter than the German sentences. At the same time, the case markers appeared earlier in German than in Russian because, in German, case is marked on the articles that precede the nouns, whereas in Russian it is marked on suffixes at the end of the nouns. Thus, if the latencies were adjusted either to the overall sentence duration or to the onset of the case markers, both adjustments would still somewhat distort the data. Here, we chose a compromise. In the regression analysis as well as in all figures, we adjusted the latencies to the end of the sentence by subtracting

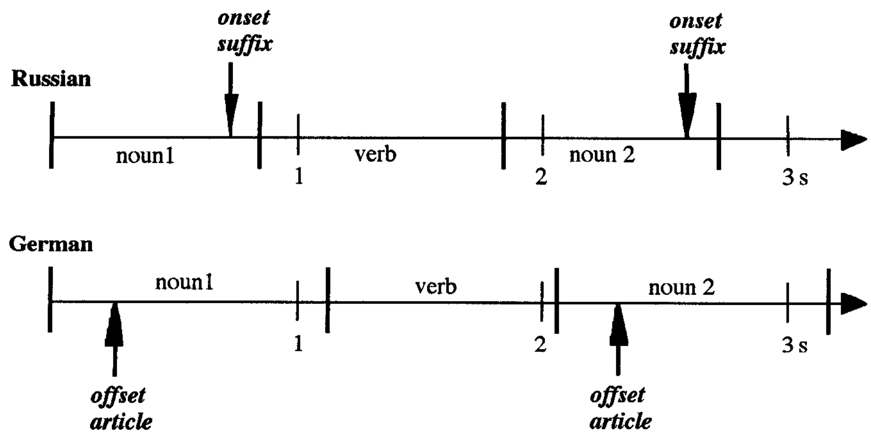

Figure 2. Position of case-markers (articles in German, suffixes in Russian) in relation to the actual duration of the two nouns and the verb.

sentence duration from the raw latency. This adjustment did not allow us to account for the onset of case markers during the sentence. Therefore, in all analyses that were designed to determine the on-line effects of cues, ANCOVAs were performed in order to separate the effects of noun and verb duration on the raw latencies before exploring the effects of the experimental factors. This was only possible in the analyses with items as random effects. Because of this duration confound, we will refrain from presenting numeric differences between various conditions as well as from presenting direct comparisons of latencies across languages. All that can safely be interpreted are language differences in the patterns of interactions of the various factors.

Lexical Decision Task. For each subject, the proportions of hits and false alarms in the lexical decision task were used to calculate d' as a measure of word sensitivity in the L2.[5] The learners of German had a significantly higher word sensitivity ($d' = 1.59$) than the learners of Russian, $d' = 1.29$; $F(1, 42) = 5.3$, $p < .05$, despite nearly identical amounts of exposure time as reported in the questionnaire (German: 26.5 months, Russian: 25.2 months, $p > .8$). Even though similarity to English was avoided in the word materials, this result is hardly surprising, given that the learners of German have a learning advantage due to the familiarity of the graphemic system.

Picture-Choice Task.

Regression Analysis on the First-Noun Choices. First-noun choices in case-marked OVS sentences are the result of the learners' failure to use case markers correctly and constitute an incorrect response. In order to assess the language differences in case-marking mastery, we collapsed the proportions of first-noun choices across all OVS sentences that contained at least one reliable case marker over all levels of first- and second-noun animacy. This pro-

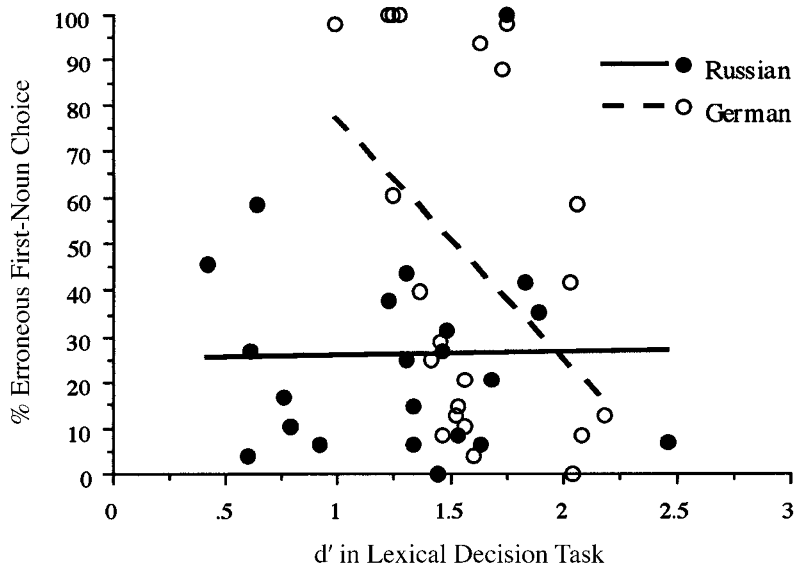

Figure 3. Scatterplot and fitted curves of the regression of errors in case-marked OVS-sentences on word sensitivity (*d′*), collapsed over all levels of case marking, for German and Russian.

portion was then arcsin-transformed and entered as dependent variable into a multiple regression analysis with *d′* and Language (coded as a dummy variable) as predictor variables. *D′* was entered first into the regression and yielded no effect on the proportion of first-noun choices. Language, which was entered next, accounted for 13% of variance, $F(1, 41) = 6.1$, $p < .01$. Finally, the interaction of Language with *d′* accounted for an additional 9% of variance, $F(1, 40) = 4.4$, $p < .05$. Figure 3 depicts the percentages of errors in case-marked OVS sentences as a function of *d′* in the learners of Russian and German. The results indicate that the learners of Russian made fewer errors in case-marked OVS sentences and that the error rate was not affected by their familiarity with the language. The learners of German made significantly more errors, and the error rates decreased with increasing language familiarity. These results lead to the conclusion that case inflections are learned faster in Russian.

Regression Analyses on the Latencies. All raw latencies over 4,000 ms were truncated. In each learner group, there were two very slow subjects for whom the truncation resulted in a loss of 20% of the data points. For all other subjects, the truncation affected only a total of 1% of the data points.

The first set of regression analyses examined the effect of case-marking mastery and language familiarity on processing speed in the learners of Russian and German. Two stepwise regression analyses were performed: one with the overall end-adjusted latencies as dependent variable, and the other with the end-adjusted latencies for case-marked OVS sentences as dependent variable. Language (coded as a dummy variable with values of 0 for German and values of 1 for Russian), *d′* in the lexical decision task, and case-marking mastery, coded as percentage of correct second-noun choices in case-marked OVS sentences, were entered as predictor variables. All effects and interactions that did not account for a significant proportion of variance were removed from the equation. Because the outcome of the two regression analyses was virtually identical, we will only report the results for the overall end-adjusted latencies.

Language accounted for 42% of variance, $F(1, 42) = 30.9$, $p < .001$. This main effect was due to the later onset of case markers in Russian and will not be considered further. At the second step, case-marking mastery was entered, which explained an additional 19% of variance, $F(1, 41) = 20.3$, $p < .001$. The positive regression coefficient of 9.2 confirmed that case-marking mastery and decision latencies were positively related: The better the learners' performance on case-marked OVS sentences, the longer their overall latencies. At the third step, the interaction between Language and case-marking mastery was entered, which accounted for an additional 4% of variance, $F(1, 40) = 4.8$, $p < .05$. This indicates that the relationship between case-marking mastery and decision latencies was different in the two language groups. Although latencies increased with increasing case-marking mastery in the learners of German, the effect was the opposite in the learners of Russian, as evidenced by the negative regression coefficient of –9.7. Finally, the interaction between Language and *d′*, which was entered at the last step, accounted for an additional 6% of variance, $F(1, 39) = 7.5$, $p < .01$. The negative regression coefficient of –435 indicates that learners of Russian, but not learners of German, exhibited a decrease of latencies with increasing language familiarity.

Figure 4 illustrates the relationship between case-marking mastery and overall processing speed. In order to examine the inverted u-shaped relationship between case-marking mastery and processing speed, a polynomial regression was performed on the overall decision latencies for both groups combined. Case-marking mastery was entered as the predictor variable. If the relationship is indeed curvilinear, then the quadratic term should account for a significant proportion of variance over and above the linear term. This regression yielded a significant effect of case-marking mastery, $F(1, 41) = 16.7$, $p < .001$, which accounted for 38% of variance. The quadratic term was also significant, $F(1, 41) = 7.5$, $p < .01$, and accounted for an additional 10% of variance. Figure 4 shows that more learners of German placed on the ascending side of the inverted u-shaped trajectory, which also supports the idea that learners of German had not yet reached the same level of case-marking mastery as the learners of Russian.

Taken together, the regression analyses reveal the following picture: The English-speaking participants had learned Russian case marking faster than German case marking. Although the learners of German had received more language input (as indicated by their higher sensitivity in the lexical decision task), they were just starting to abandon their strong "agent first" bias and to

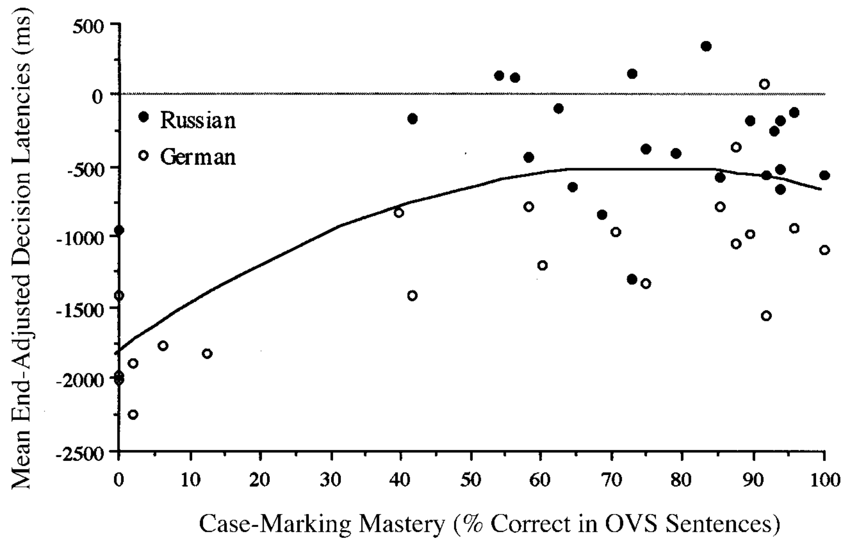

Figure 4. Curvilinear relationship between case-marking mastery and overall processing speed in both learner groups combined.

attend to the case markers. This was leading to a slowing down of processing. The learners of Russian had already reached a stage where case marking is processed correctly and becomes more automatized with increasing language familiarity. Placing these two learner groups onto one continuum of case-marking mastery demonstrates that processing speed follows an inverted u-shaped trajectory over the course of learning.

ANOVAs on First-Noun Choices. In order to compare on-line processing effects across languages, it was necessary to match the two groups for language familiarity. For this purpose, six learners of Russian with d'-scores below 0.9 as well as five learners of German with d'-scores above 2.0 were excluded. Additionally, one learner of Russian with an exceptionally high d'-score (2.47) was also excluded. The matched subgroup consisted of 15 learners of Russian and 17 learners of German, all with d'-scores between 0.9 and 2.0 and a mean d' of 1.4 in both groups. All analyses of variance reported below were performed on the matched learner subgroups.

The arcsin-transformed proportions of first-noun choices of the matched subgroups were submitted to a 2 (Language) × 2 (N1 Marking) × 2 (N2 Marking) × 2 (Configuration) × 2 (N1 Animacy) × 2 (N2 Animacy) ANOVA. Note that SVO and OVS sentences without case marking, which were identical in surface form, were treated as different conditions in the analyses, conforming to their status in the experimental design. All effects that reached significance both in the analysis by subjects and in the analysis by items are displayed in Table 4. The condition means for the matched subgroups are presented in Appendix D.

We will first discuss the effects that are common to both learner groups

Table 4. Significant effects in the ANOVA of the arcsin-transformed proportions of first-noun choices

Effects	F_1 (1, 30)	F_2 (1, 193)
LANG	6.9*	95.4***
M1	84.9***	237.0***
M2	8.6**	13.2***
CO	83.0***	1328.2***
A1	14.3***	22.9***
M1 × M2	16.1***	17.8***
M1 × CO	90.5***	263.2***
M2 × CO	36.9***	135.3***
M2 × A1	10.3**	9.7**
M1 × M2 × CO	26.3***	113.1***
M1 × A1 × A2	7.7**	5.4*
M2 × A1 × A2	7.2*	11.6**
M1 × M2 × A1 × A2	4.3*	8.3**
M1 × M2 × A1 × A2 × LANG	6.9**	6.7**

Note. *$p < .05$, **$p < .01$, ***$p < .001$

and will then turn to the language differences. The large effect of Configuration indicates that learners of both languages tended to choose the first noun in SVO sentences and the second noun in OVS sentences. The effects of N1 Marking and N2 Marking as well as their interactions with Configuration demonstrate that these choices were influenced by the inflections on the first and the second noun. If the first noun was reliably marked as nominative or the second noun was reliably marked as accusative, learners tended to choose the first noun as agent. If the first noun was reliably marked as accusative or the second noun was reliably marked as nominative, learners tended to choose the second noun as agent. Thus, the agent choices were in accordance with case marking. As can be seen from Figure 5, the most difficult case-marked OVS sentences were those with case-neutralized first nouns and nominative-marked second nouns. These sentences proved also to be difficult for native speakers because they require a restructuring of the initial interpretation of the case-neutralized first noun as agent (Kempe & MacWhinney, 1997). The late appearance of a nominative marker on the second noun signals that the initial interpretation was wrong and needs to be revised. Consequently, listeners who are biased toward first-noun choices and tend to be fast in their decisions will miss the cue on the second noun and misinterpret the sentence.

The analyses also revealed some differences between the two language groups: The main effect of Language indicates that, overall, learners of German made more first-noun choices than learners of Russian. The interaction of Language and Configuration reached significance in the analysis by items,

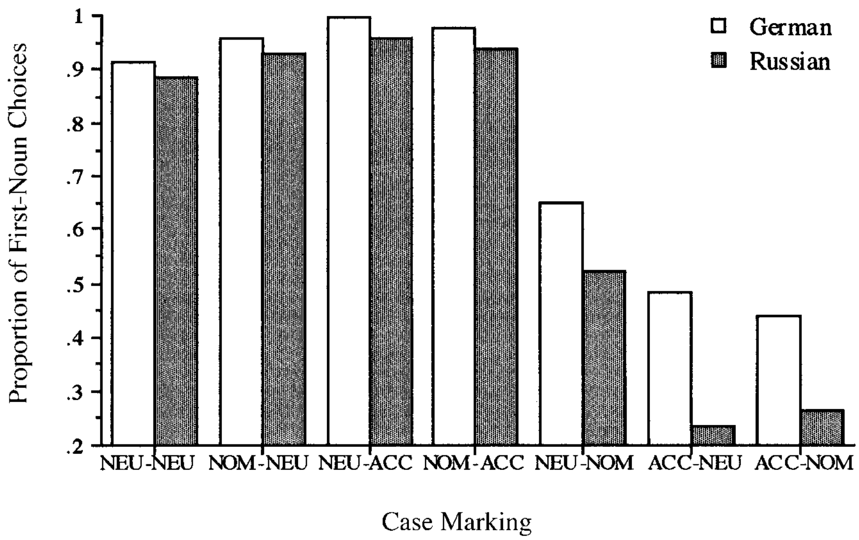

Figure 5. Proportion of first-noun choices as a function of case marking for the learners of Russian and German (NEU, case neutralized; NOM, nominative marked; ACC, accusative marked). Case-marking labels for the first and the second noun are separated by a hyphen.

$F_2\,(1, 96) = 6.1$, $p < .05$, and fell short of significance in the analysis by subjects, $F_1(1, 30) = 3.5$, $p = .07$. This indicates that the German learners' tendency of making first-noun choices was more pronounced in case-marked OVS sentences, as Figure 5 shows. The ANOVA also revealed an effect of N1 Animacy. Overall, first-noun choices were more likely to occur if the first noun is animate. This suggests that the learners' decisions are aided by nonsyntactic information like noun animacy. The three-way, four-way, and five-way interactions involving N1 Animacy and N2 Animacy were somewhat peculiar. Closer inspections of the individual condition means revealed that the higher order interactions were caused by the fact that the learners of Russian made fewer first-noun choices than the learners of German in all but sentences with inanimate first nouns, animate second nouns, and no case markers. In these sentences, the proportion of first-noun choices was approximately the same in both languages. However, taken together, all the higher order interactions involving animacy accounted for only a total of 2% of the experimental variance and will not be considered any further.

In sum, the first-noun choices indicate that both learner groups tended to base their agent choices on the case marking and animacy cues.

ANCOVAs on Decision Latencies. For each language group, we performed a 2 (N1 Marking) × 2 (N2 Marking) × 2 (Configuration) × 2 (N1 Animacy) × 2 (N2 Animacy) ANCOVA with the duration of the nouns and the verb as covariates

Table 5. Significant effects and effect sizes in % of experimental variance in the ANCOVAs on the latencies in the familiarity-matched Russian and German learner subgroups

Language	Effects	F_2 (1, 93)	Effect size (% of exp. variance)
Russian	N1dur	48.3***	9
	N2dur	17.1***	3
	M1	270.1***	51
	M2	16.4***	3
	CO	16.4***	3
	M1 × M2	26.1***	5
	M1 × CO	25.1***	5
	M1 × A1	23.2***	4
	M2 × A2	6.3*	1
German	N1dur	38.0***	15
	M1	121.4***	48
	M2	8.6**	3
	A1	36.8***	15
	A2	7.0**	3
	M1 × M2	11.3**	5
	M1 × CO	11.3**	5

Note. Duration of first noun (N1dur), duration of verb (Vdur), and duration of second noun (N2dur) were entered as covariates. *$p < .05$, **$p < .01$, ***$p < .001$.

and items as random effect. The mean end-adjusted latencies per condition for the matched subgroups are presented in Appendix D. Recall that the end-adjusted latencies do not account for the noun and verb duration differences. Table 5 presents the significant effects obtained in the ANCOVAs as well as the corresponding effect sizes computed as percentages of experimental variance accounted for by each factor or interaction.

Because of differences in the position of case markers, word duration had different effects on the decision latencies in the two languages. In Russian, there was an effect of the duration of the first as well as the second noun, which reflects the contribution of the suffixes at the end of the nouns. In order to fully utilize case marking in comprehension, the listener had to delay the decision until the end of the second noun if no information was given on the first noun. In German, no effect of the duration of the second noun was found, which is a consequence of the fact that the case marker appears on the article before the second noun. The duration of the noun itself appears not to influence the latencies, which suggests that decisions can be made immediately after encountering the case marker on the second article.

Next, we describe the effects related to case marking and then turn to effects related to animacy. Figure 6 shows the decision latencies as a function of case marking, collapsed over all levels of N1 Animacy and N2 Animacy. A comparison of the effect sizes of case marking reveals that N1 Marking had a

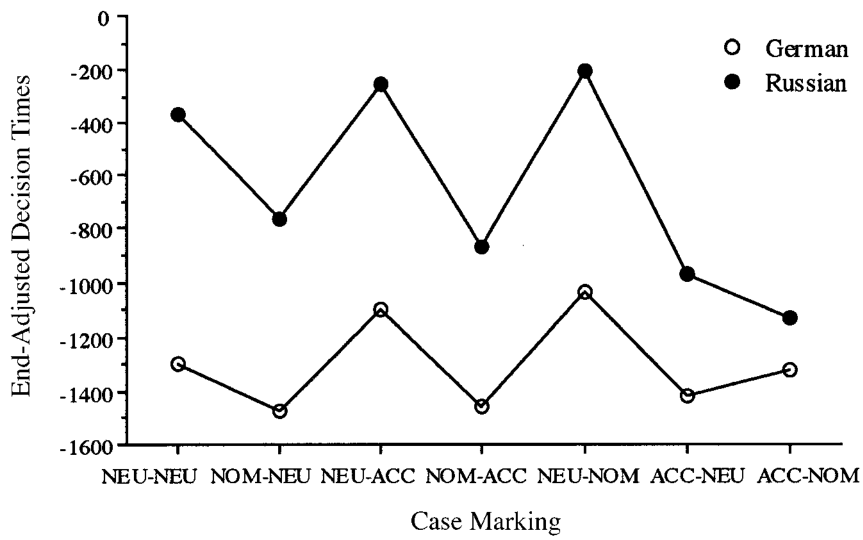

Figure 6. End-adjusted decision latencies as a function of case marking for the learners of Russian and German (NOM, nominative marked; ACC, accusative marked; NEU, case neutralized). Case-marking labels for the first and the second noun are separated by a dash.

large effect in both languages. Thus, a reliable case marker on the first noun reduced the decision latencies significantly compared to sentences with no reliable case marker on the first noun. The interaction of N1 Marking with Configuration indicates that these processing benefits from first-noun marking were more pronounced in OVS sentences in which the first noun is accusative marked. The analyses also revealed a main effect of N2 Marking in both groups. This effect, in conjunction with the interaction of N1 and N2 Marking, indicates that, if the first noun was case neutralized and the second noun was reliably marked, processing was slower than if both nouns were neutralized. If the first noun was reliably marked for case, case marking on the second noun had no additional effect.

The most important language differences in processing speed were related to animacy. We found main effects of first- and second-noun animacy for German, which suggests that decisions were faster if the first noun was animate or if the second noun was inanimate, regardless of the case marking on the nouns. In Russian, there was an interaction between case marking and animacy of the second noun, which indicates that decisions were faster if the second noun was case neutralized and inanimate than if it was case neutralized and animate. No animacy effect was found if the second noun was reliably case marked. This suggests that animacy of the second noun can aid the decision process, but only in the absence of case marking. However, some of the effects of animacy in Russian were exactly opposite to German. Thus, decisions were faster if the first noun was case marked and inanimate than if it was case marked and animate. We will return to this finding in the Discussion.

Taken together, the most important results from the latency analyses are that the strongest processing benefits were obtained from case marking on the first noun and that these benefits were even more pronounced if the first noun was accusative marked. Only in German were there on-line effects of animacy of the first noun regardless of case marking, which indicates that, unlike in Russian, animacy is used immediately to supplement the considerably weaker case-marking cue.

Discussion

The cue-validity estimations derived from the Competition Model framework predicted that L2 learners of Russian should be able to correctly rely on case marking at an earlier stage in learning than L2 learners of German. This is exactly what the data show. Even at very low levels of language familiarity, the learners of Russian exhibit a low amount of incorrect first-noun choices in case-marked OVS sentences, and this error rate remains low as the language becomes more familiar. Learners of German, on the other hand, show very high error rates in case-marked OVS sentences at the lower levels of language familiarity, and these error rates decline gradually as language familiarity increases. This indicates that learners of Russian are much quicker in learning to use case marking as a cue to agenthood in sentences with noncanonical configuration, whereas learners of German maintain their strong preverbal subject bias for a longer period and learn to use case marking at a late stage.

The experiment yielded very interesting results with respect to changes in processing speed during the course of learning. We expected overall processing speed to follow an inverted u-shaped curve. This prediction was confirmed by the polynomial regression analysis of case-marking mastery on the overall decision latencies. In learners who have not yet fully mastered case marking, processing speed decreases with increasing case-marking mastery because these learners are just beginning to pay attention to additional cues later in the sentence and to delay their decisions accordingly. For learners that have already mastered case marking, processing speed increases as the use of case markers becomes more automatized with increasing exposure to the L2. In our study, more learners of German could be found on the ascending part of this curvilinear trajectory, which supports the claim that they, as a group, were inferior in learning case marking.

A quantitative analysis of decision latencies in the familiarity-matched subgroups revealed that case marking on the first noun resulted in faster performance. Although the variability in sentence duration precluded a direct comparison of this processing benefit across languages, a closer inspection of Figure 6 suggests that the relative benefit from case marking on the first noun tended to be larger in Russian than in German. Thus, the latency data are sensitive to subtle language differences in processing and provide converging evidence for the claim that case marking is a stronger cue in Russian. The fact that this processing benefit is more pronounced for accusative marking than for nominative marking on the first noun was not predicted from the corpus-based validity estimations. Furthermore, case marking on the second noun speeds up decisions only if the first noun is not reliably marked. This shows that benefits from case marking on the first and second noun do not combine additively. Apparently, local cues like case markers exert their effects immediately, and decisions are made as soon as a strong cue is encountered.

In both languages, the first noun is chosen as the agent more often if it is animate. This seems to indicate that decisions are generally supported by semantic information. However, the latency analyses revealed different on-line effects of animacy in the two language groups: In the learners of German, both the animacy of the first noun and, to a lesser degree, the inanimacy of the second noun led to faster decisions. Thus, the fact that animate nouns are plausible agents and inanimate nouns are plausible objects influences the speed of processing. Evidently, learners of German consider semantic information immediately and regardless of whether an unambiguous case marker is present or not—a strategy that was predicted by the overall lower validity of case markers in German. Low validity of case marking makes it essential to rely on additional information in order for comprehension to function efficiently.

For the learners of Russian, the picture is more complicated. The choice data indicate that the final sentence interpretation by the learners of Russian is affected by animacy information. However, the on-line data revealed that processing speed benefited only from animacy information of the second noun and only when other sources of information (e.g., case marking) were not available. Clearly, animacy information has much less of an immediate online effect in Russian than in German. Furthermore, the learners of Russian exhibited a paradoxical processing benefit when the case-marked first noun was inanimate. Most likely, this effect is a consequence of the fact that animacy is one of the underlying dimensions in the declensional paradigm and, therefore, inevitably confounded with case marking. It is possible that the inflections of the inanimate nouns in this experiment are generally more strongly associated with the nominative or the accusative than the inflections of the animate nouns. In order to check on this possibility, we counted the relative frequencies of inflection-case combinations in a large corpus of 700 inflected Russian nouns sampled from various Russian texts. It turns out that the inflection *-a* was associated with the nominative 35% of the time and with the accusative only 11% of the time (the remaining nouns ending in *-a* are genitive). In the present experiment, the inflection *-a* always marked the nominative in inanimate nouns and the accusative in animate nouns. Similarly, the inflection *-u* was associated with the accusative 74% of the time and with the dative only 26% of the time. In the present experiment, the inflection *-u* on inanimate nouns always marked the accusative. Thus, the experiment mirrored precisely the patterns of co-occurrence that are most frequent in Russian. The inanimate nouns are the ones that happen to have the stronger inflection-case associations, which may explain the faster processing of these nouns. It should be emphasized that these effects are due to the structure of the language and cannot be avoided when animacy and case are manipulated at the same time. They indicate that animacy as an underlying dimension of the Russian case-marking paradigm is much more grammaticalized than in German. Thus, in languages for which animacy is an integral part of the morphosyntax and contributes to the validity of inflectional cues, it seems to play a much lesser role as a supplemental semantic cue. Together with the main finding that the learners of Russian were clearly superior in acquiring case marking in their L2, the animacy effects in Russian also support the notion that the availability and reliability of cues in the input are the crucial determinants for input-driven associative learning in adult L2 learners.

Given this evidence for associative learning in L2 learners, and in light of many approaches that advocate the operation of associative learning mechanisms in children (e.g., Plunkett & Marchman, 1991; Rumelhart & McClelland, 1986, 1987), it is reasonable to expect that the comprehension strategies of advanced L2 learners should resemble those of native speakers. In order to test this assumption, we compared the on-line performance of the eight most successful learners in each language with the on-line performance of 30 native speakers tested with the same task in Kempe and MacWhinney (1997). The qualitative pattern of results in the native speakers was indeed similar: The benefit from first-noun marking was larger in German than in Russian, and there was a benefit from first-noun animacy in German but not in Russian. However, the magnitude of the first-noun–marking benefit was much larger in the L2 learners. Thus, although sentences with case-neutralized nouns were processed with approximately the same speed by native speakers and L2 learners, sentences with case-marked first nouns were processed much faster by the L2 learners. The relative benefit from case-marked first nouns was about 260 ms for German native speakers and about 420 ms for Russian native speakers. For the same sentences, the benefit was 634 ms for the learners of German and 975 ms for the learners of Russian. Although performance of L2 learners and native speakers is qualitatively similar, there are quantitative differences. One explanation for the faster performance of the L2 learners is that they direct more attentional resources to the case markers. We can speculate that this might be a consequence of awareness of inflectional structures induced by explicit learning of morphology as favored in the classroom setting. Further research directed at a detailed comparison of comprehension strategies in native speakers and advanced L2 learners is needed in order to better understand the nature of these differences. Still, the general advantage for learning Russian case marking suggests that even if explicit presentation of rules and regularities is a common part of L2 learning, it does not suppress the learner's sensitivity to distributional characteristics of the input. This resonates well with findings suggesting that explicit-deductive learning is not very effective for prototypical patterns and fuzzy rules (DeKeyser, 1995) like the ones typically found in inflectional paradigms.

STUDY 2: NETWORK SIMULATION

The Competition Model notions of reliability and availability allow us to determine the relative hierarchy of cues within a particular language. However, they do not allow us to predict specific patterns of interaction and joint effects of cues in on-line language processing. Moreover, the Competition Model does not provide us with a full learning model. Work by Taraban and colleagues (Taraban & Palacios, 1993; Taraban & Roark, 1996) has shown that the detailed effects of cue interactions and exemplar frequency in learning are better reflected in connectionist models.

The simulation described below explores whether a connectionist model can reproduce the basic results of this experiment if it is given the statistical distribution of cues in Russian and German as input. So far, connectionist models have not yet received wide attention as tools for explanation and prediction in the area of L2 learning. In fact, we are aware of only three attempts to model phenomena occurring in adult L2 learning using connectionist networks. Two of these models (Gasser, 1990; Sokolik & Smith, 1992) are not designed to simulate specific empirical findings, and the third model (Taraban & Roark, 1996) simulates only a very limited learning situation in which native speakers of English acquired the gender of 36 French nouns. The connectionist simulation described below attempts to model a more complex set of L2 learning data. We use a very simple neural network with only a few units. It is not our intention to present this minimalist model as a realistic simulation of actual L2 learning. Instead, we focus on what effects the different statistical distributions of cues in German and Russian have on the quantitative and qualitative performance of the model in order to explore whether a mechanism that is based on associative learning can produce similar results.

Network Architecture

In neural networks, activation is propagated through a network of interconnected nodes or units. Learning takes place by changing the weights of the connections between the units. Many models of human learning have successfully used a supervised learning algorithm called back-propagation (Rumelhart, Hinton, & Williams, 1986), which modifies the connection weights so as to minimize the difference between the output of the network and a desired target signal. The back-propagation learning algorithm adjusts the connection weights in a way that reduces the overall error in the network. In many models, activation passes from the input units to the output units through one or multiple layers of hidden units. The distribution of activation over the hidden units can be thought of as the internal representation of the acquired knowledge.

In comprehension, linguistic input is perceived in more or less discrete units that are presented over time. In order to simulate the time-sensitive nature of comprehension, we used a recurrent network, which is able to handle

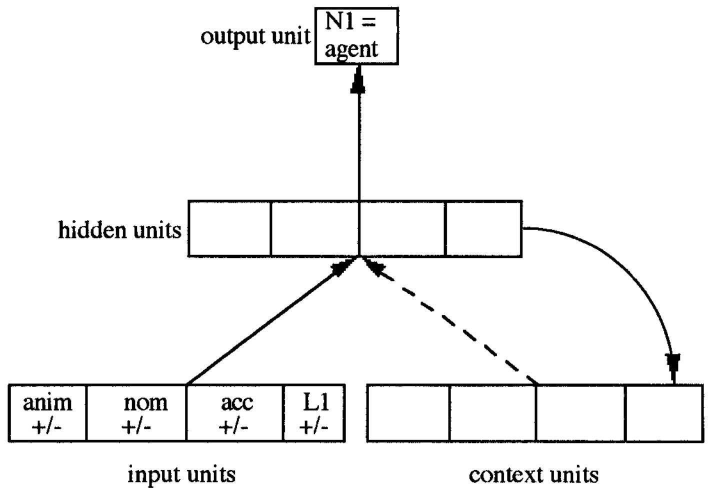

Figure 7. General architecture of the recurrent neural network used in the simulation.

incremental changes in the input over time. Recurrent networks contain feedback loops that allow the network to keep track not only of the present input but also of the context that has been presented and processed earlier. For example, in so-called Elman networks (Elman, 1990), the current activation state of the hidden units serves as context for subsequent inputs so that, at each time step, the network receives not only the current input but also a representation of its prior internal state.

The Elman network used in our simulation consisted of four input units, four context units, four hidden units, and one output unit (see Figure 7). The four input units coded the following features of each noun: animacy (±), nominative marking (±), accusative marking (±), and whether a given noun belongs to the native language or not. Note that this binary coding reflects only the mere presence or absence of nominative or accusative marking and not the specific inflections used to mark these cases. Because the experiment did not contain verb agreement, the input was restricted to the information from the nouns. In the output unit, an activation value of 1 was associated with the first noun as agent and an activation value of 0 was associated with the second noun as agent.

In each trial, the network was first presented with the input pattern corresponding to the first noun. Then, the activation values of the hidden units reached after presentation of the first noun were copied to the context units. In the next step, the activation on the context units was presented together with the input pattern corresponding to the second noun. After each two-noun sequence (i.e., a "sentence"), the context units were cleared and the network was presented with the next sequence of input patterns.

We constructed three training-pattern corpora to capture the typical distributions of cues in English, German, and Russian. The English input patterns resembled English transitive sentences in that there was no case marking on the nouns and the first noun was always the agent. Animacy information was distributed roughly similar to Russian and German. The Russian and the German training patterns mirrored the frequency distribution of all 32 experimental sentence types in the L2 textbook corpus described above. The only difference from the original corpus was that sentence types that were not present in the L2 textbook corpus were presented to the network at least once. Thus, the Russian training patterns consisted of 575 two-noun sequences and the German training patterns consisted of 675 two-noun sequences. The smaller number of Russian training patterns reflects the fact that frequent subject omission in Russian results in fewer NVN sentences than in German.

The network was first trained on the English corpus until it had settled down, which took only three sweeps through the patterns (epochs). Obviously, there is not much learning involved for English; all patterns pointed toward the first noun as agent and were therefore always associated with an output of 1. The purpose of training the network on English was to arrive at internal weights that correspond to the native English agent-first bias. Next, one network was trained on the Russian corpus and the other one on the German corpus. We did not place the network through a period of mixed training in both the L1 and the L2, as has been done in other simulations (Gasser, 1990), because our primary goal was to explore the learning trajectories of the Russian and German networks as a function of the different cue distributions in the languages. The learning rate was set to .01 and momentum to .7. Prior pilot simulations had shown that if these parameters are set to higher values, the networks are not able to escape the local error minimum associated with a rigid agent-first strategy. Each network was trained for 1,000 epochs. Every 100 epochs, the networks were tested on all patterns corresponding to the 32 sentence types. For each pattern, output activation values reached after the first noun and after the second noun were saved and submitted to further analyses.

The dependent variables in our experiment were proportions of first-noun choices and decision latencies. In order to compare the network performance with the human data, it was necessary to convert the output activation into estimations of noun-choice proportions and latencies. These conversions are explained in detail in Appendix E. The estimations take into account the fact that the strength of the response alternative reached after the presentation of each noun determines when during the sentence a decision is likely to be made. Thus, if there is a strong cue on the first noun, a fast choice is executed right after the first noun and cues on the second noun have little additional impact on the speed of the decision. On the other hand, if the response strength after the first noun is low, the response is likely to be delayed until after the second noun. The estimations also take into account that the coherence between the activation after the first and the second noun may affect the speed of decisions. Thus, if the cues on the first noun point to the first noun as the agent and the cues on the second noun point to the second noun with equal strength, additional time and effort are required to restructure the initial interpretation.

Performance of the Model

For each pattern, the output activation values were converted into first-noun–choice probabilities and latency estimations as explained in Appendix E. The mean choice probabilities and latency estimations per pattern, collapsed over all epochs, were then compared with the mean first-noun–choice proportions and decision latencies per condition in the familiarity-matched subgroup. These means are presented in Appendix D.

For the first-noun choices, a goodness-of-fit statistic yielded an R^2 of .90 for Russian, $F(1, 30) = 126.5$, $p < .001$, and an R^2 of .64 for German, $F(1, 30) = 54.4$, $p < .001$. Thus, the fit of the model to the choice data is excellent for Russian and very good for German. The fit of the latency estimations was determined after the effects of first- and second-noun duration had been partialled out from the mean latencies per condition. Because the verbs were counterbalanced within conditions, verb duration was a constant and therefore not entered into the analysis. The latency estimations accounted for 22% of variance over and above the noun durations in Russian, $F(1, 28) = 13.9$, $p < .001$, and for 14% of variance, $F(1, 28) = 6.0$, $p < .05$, in German. Again, the model fit was somewhat better for Russian than for German.

In sum, the model explains a significant proportion of variance in the familiarity-matched learner subgroups, both in the noun choices and in the decision latencies. However, overall fit of the model does not allow us to determine to what extend the model exhibits the same qualitative characteristics as the human data. The following comparisons will explore the model's behavior in greater detail. Specifically, we were interested in seeing whether the model replicates the better performance in case-marked OVS sentences in the learners of Russian, the inverted u-shaped relation between case-marking mastery and decision latencies, and the language differences in processing benefits from cues on the first noun.

Case-Marking Mastery in Russian and German. Figure 8 shows the probability of first-noun choices in case-marked OVS sentences over the course of learning. The simulation matches the human data in that the Russian net is clearly superior to the German net in learning case marking. The learning advantage in Russian comes about because the higher frequency of case marking in the Russian input to the network leads to a faster strengthening of the connections between case markers and thematic roles than in German. However, there are differences from the human data in the qualitative pattern of results. If the first-noun choices are broken down by sentence type (see Appendix D), it becomes apparent that the inferior performance of the German net is mainly

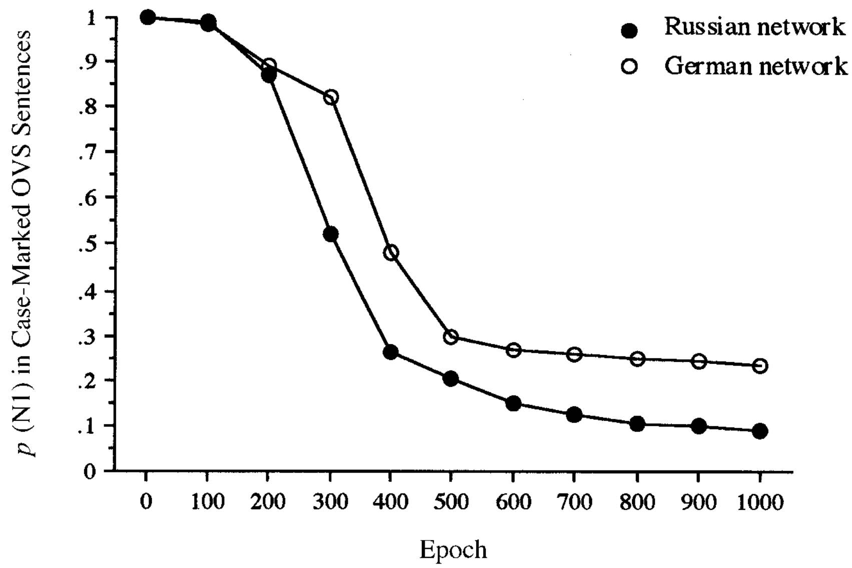

Figure 8. Probability of incorrect first-noun choices in case-marked OVS sentences as a function of learning in the Russian and German networks.

due to sentences where case marking and animacy are in conflict. The network does a particularly bad job when nominative marking of the second noun competes with animacy of the first noun as in sentences like *Die Mutter sucht der Teller* (The-NOM/ACC mother looks-for the-NOM plate). Here, the network never learns to disregard the conflicting animacy cue and to base decisions on case marking.

Case-Marking Mastery and Decision Latency Estimations. In Figure 9, the estimated latencies are plotted against the probability of correct second-noun choices in case-marked OVS sentences for the Russian and German networks. As in the human data, increasing mastery of case marking is associated with increasing latencies at earlier stages and with decreasing latencies at later stages of learning. The increase in latencies reflects the learning of new cues which work to slow down the monolithic application of the agent-first bias. Latencies tend to drop again at higher levels of case-marking mastery, as the newly learned patterns solidify their interrelations. Thus, the network performance exhibits the same inverted u-shaped relation between performance speed and case-marking mastery that was found in the human data.

Figure 10 presents the decision latency estimations as a function of learning. Recall that in the human data learners of Russian started to exhibit a decrease in latencies at lower levels of language familiarity than the learners of German. The same language difference can be found in the network performance for which decision latencies decrease earlier in the Russian network. Again, this finding matches the results obtained from the human data as well.

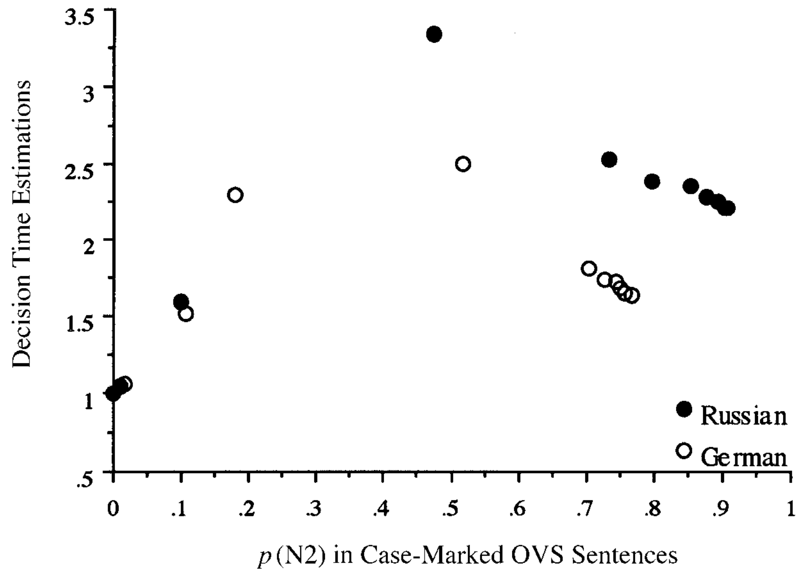

Figure 9. Latency estimations as a function of the probability of correct second-noun choices in case-marked OVS sentences in the Russian and German networks.

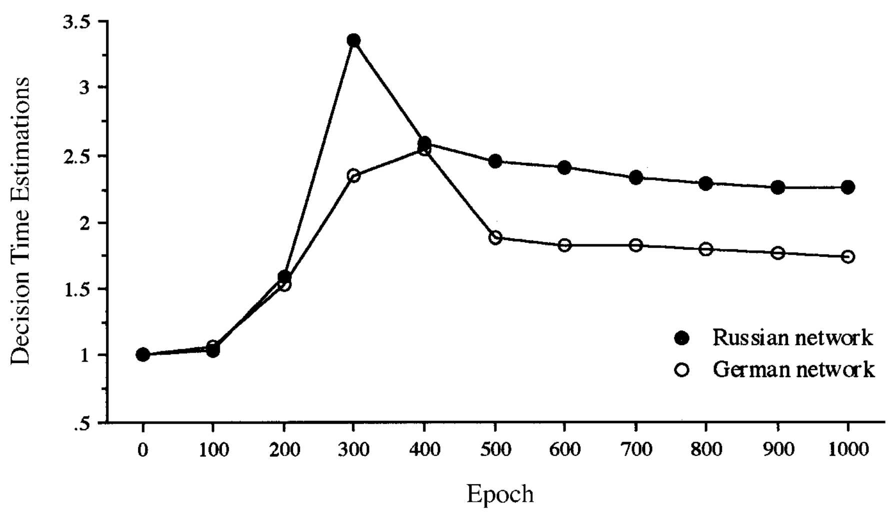

Figure 10. Latency estimations as a function of learning in the Russian and German networks.

Table 6. Significant effects and effect sizes in % of experimental variance in the ANOVAs on estimated latencies in the Russian and German networks

Language	Effects	F_2 (1, 93)	Effect size (% of exp. variance)
Russian	M1	193.8***	91
	CO	5.3*	2
	M1 × CO	8.1**	4
	M1 × A1	5.0*	2
German	M1	30.7***	13
	CO	4.4*	2
	A1	137.0***	57
	M1 × A1	45.2***	19
	CO × A1	9.6**	4

Note. *$p < .05$, **$p < .01$, ***$p < .001$

Effects of Cues on Decision Latency Estimations. In order to compare the effects of the various cues between model and learners, we submitted the latency estimations for each pattern obtained every 100 epochs to a 2 (N1 Marking) × 2 (N2 Marking) × 2 (N1 Animacy) × 2 (N2 Animacy) × 2 (Configuration) ANOVA separately for each language. These analyses are the network equivalent to the analyses of the familiarity-matched learner subgroup. Table 6 presents the significant effects as well as the effect sizes calculated as percent of experimental variance for the Russian and German networks.

As in the human data, there was a benefit in processing speed from first-noun case marking. This benefit was much stronger in the Russian network than in the German network. Compared to the learners, the simulation tends to overestimate the magnitude of this effect in Russian and to underestimate it in German, as can be seen from the proportions of variance accounted for by N1 Marking in the networks (see Table 6) and in the learner data (see Table 5). Figure 11 depicts the latency estimations as a function of case marking, collapsed over all levels of animacy. Although the model correctly predicts a larger first-noun marking benefit in Russian than in German, it tends to exaggerate the language difference. The overestimated effect sizes in the model are in part a consequence of the fact that human data are much noisier. It should be noted, however, that in the simulation the first-noun case-marking benefit was actually greater from nominative marking than from accusative marking. Recall that in the learners we found a greater benefit from accusative marking. Several possible reasons for the model's failure to capture the accusative superiority effect will be discussed below.

Another important parallel between model and data is the benefit from animacy of the first noun, which was observed for the learners of German as well as in the German network but not for the learners of Russian or the Russian network. This supports the idea that lower strength of morphological

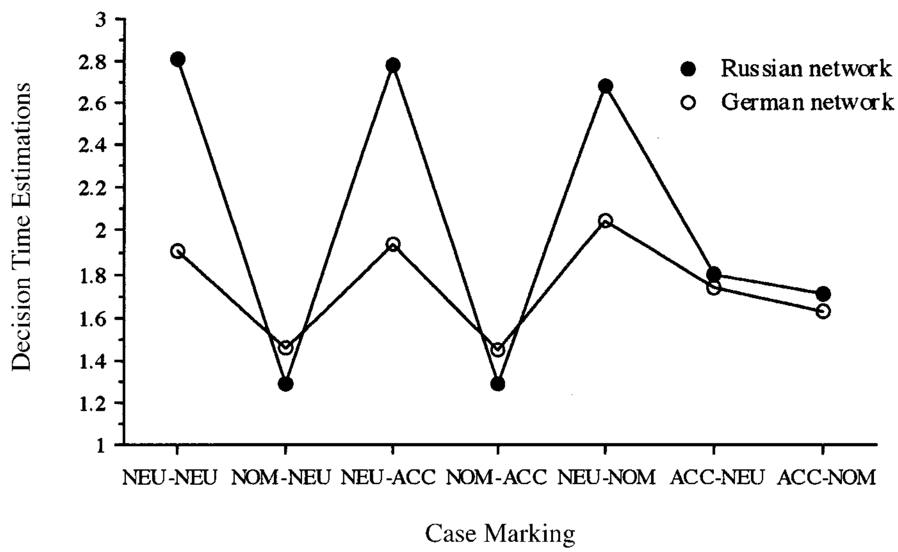

Figure 11. Latency estimations as a function of case marking on the first and second noun in the Russian and German networks (NOM, nominative marked; ACC, accusative marked; NEU, case neutralized). Case-marking labels for the first and the second noun are separated by a dash.

cues is compensated for by greater reliance on semantic cues. Two minor differences between model and data should be noted in this regard: First, the magnitude of the effect was again much larger in the simulation than in the human data. Second, the processing benefit from first-noun animacy in the German network is stronger if the first noun is neutralized than if it is reliably marked, a result that is different from the learners of German, for whom no interaction between case marking and animacy was found.

Finally, the model clearly underestimates the effects of the cues on the second noun. Recall that, in the language learners, latencies increase if the first noun is case neutralized and the second noun is reliably marked. In the simulation, there was virtually no effect of second-noun cues, which can be attributed to the fact that the response alternatives after the first noun were very strong. This suggests that in the networks the weights get consolidated much faster than is the case with human learners.

Discussion

The simulation accounted for many important results obtained in the comprehension experiment with learners of Russian and German. First, it correctly captured the finding that case marking is learned faster in Russian than in German. Given that the only difference between the German and Russian networks was in the distribution of cues in the training corpus, this simulation outcome clearly indicates that the frequency of case marking in the input plays a crucial role in acquiring this cue. Second, the network captured the inverted u-shaped relation between case-marking mastery and processing speed found in the L2 learners. In the network, an increase in latencies comes about because the strength of the response alternatives decreases (i.e., moves closer to .5) as the old agent-first bias is abandoned and new cues take over. With learning, the strength of these new cues increases, and the output moves away from .5, which is reflected in faster latencies. Third, the network performance was speeded up by case marking on the first noun, a result that mirrors the human data as well. Moreover, the language differences in the simulation point in the same direction as the learner data: A larger benefit from case marking was found in Russian. Again, this happens because the higher frequency of case marking results in a faster strengthening of the connections between case marking and agenthood. Finally, the simulation correctly captured the language differences in the effects of first-noun animacy in that it exhibited processing benefits for animate first nouns in German. In Russian the high strength of case marking essentially overrides the effects of other supporting cues, whereas in German the low strength of case marking leaves room for other cues to contribute to the output regardless of whether they are morphosyntactic or semantic in nature.

Taken together, the simulation resembles the performance of the L2 learners in many aspects, suggesting that an associative learning mechanism provides a good account for the learning of L2 sentence cues such as case marking. The model is particularly accurate in reproducing patterns of cue interactions that could not be predicted based on the Competition Model. However, certain aspects of the model performance did not match the human data. Below we will briefly discuss these mismatches.

First, unlike the learners of German, the inferiority in learning case marking in the network is mainly due to performance in sentences with a conflict between animacy and case marking. It is possible that this is a consequence of the specific characteristics of our training corpus rather than of the learning principles in connectionist networks. Perhaps the distribution of cues in the training corpus, which was based on the distribution of cues in L2 textbooks, does not accurately reflect the actual learner input. Although the textbooks we examined are representative of the formal input to the L2 learner, they may fail to capture aspects of informal input the learner receives from other sources (e.g., oral input from the teacher). In future research, it would be desirable to complement the analysis of printed input to the learner by an analysis of oral teacher input, which is much more difficult to obtain. In order to check whether the textbooks contained fewer conflict sentences than natural language, we compared the L2 textbook corpora to corpora collected from the writings of native speakers (Kempe & MacWhinney, 1997). We found that L2 textbooks contain fewer sentences with conflicting cues than the native corpora. Specifically, the native German corpus contained about 17% of sentences in which a morphological cue was in conflict with either animacy or configuration as compared to only 10% in the German L2 textbooks. Furthermore, the frequency of the animacy cue is exaggerated in the L2 textbooks, which is a consequence of the fact that textbooks usually introduce a few protagonists and then present many sentences describing their various actions. This distortion does not affect performance of the Russian network, in which case marking is a strong enough cue to win in conflict sentences. However, it makes it more difficult for the German network to resolve the conflict between animacy and case marking. We assume that L2 learners encounter more informal and nativelike language input, which may compensate for the deviations in the textbooks and can explain the differences between network and human data. The simulation results suggest that the lack of conflict sentences can potentially be detrimental for learning. To facilitate the learning of German case, it may be desirable for future authors of L2 learning materials to attempt to preserve the natural distribution of cues and provide the necessary instances of cue conflict.

Second, the network failed to capture the stronger benefit from accusative marking than from nominative marking on the first noun. This also seems to be a consequence of the training input; specifically, of the very limited set of sentence types. The presentation of just NVN sentences naturally restricts the input in an artificial way. It is possible that our corpus did not accurately represent the frequency of accusative marking. For example, Russian null-subject sentences, which contain only the object noun and a verb, were excluded from the corpus analysis in order to maintain comparability with German. Inclusion of these sentences, which are very likely to contain an accusative-marked noun, will increase the overall frequency of accusative and strengthen the benefit from accusative marking in Russian. This explanation, however, does not work for German. All German sentences contain subjects, so that our corpus is likely to reflect the relative frequencies of German case markers quite accurately. We think that the accusative superiority can in part be accounted for by interactions between the configuration-based agent-first bias and case marking. Nominative marking on the first noun conforms with the expected noun configuration and has, therefore, less of an effect than accusative marking. The fact that nominative marking had stronger effects in the simulation than in the human data suggests that the agent-first bias was abandoned much faster by the network than by the L2 learners. This can also be attributed to the limited set of structures in the training input, which rendered word-order information, particularly information from the position of the verb, redundant. Again, including a richer set of different structures in the input should increase the model fit.

Third, there were no effects related to second-noun cues in the networks' performance. A closer inspection of the raw activation values reveals that this is an artifact of the way latencies were estimated. Recall that, in the human data, there was an effect of case marking of the second noun if the first noun was case neutralized. Similarly, in the network, the activation after the first noun is close to .5 if there is no cue on this noun or if the cue is not strong enough. If the second noun contains a strong cue, the strength of the response increases but so does the absolute difference between a1 and a2. In the latency estimations, these effects cancel each other out (see Appendix E). The fact that case marking on the second noun exerts a detectable effect in the human data suggests that the mismatch between the output activation reached after the first noun and the one reached after the second noun has a much larger impact on the latencies than reflected in our latency estimation. In order to account for the additional time needed to alter an initial interpretation more accurately, the difference between the activation reached after the first and after the second noun should be weighted by a constant, the value of which will have to be determined in future research.

Finally, the Russian network did not exhibit any effects related to inflectional syncretism. This is simply a consequence of the fact that the input coded only the presence or absence of reliable nominative versus accusative marking rather than the actual inflections of the nouns. An exploration of the specific effects of inflectional syncretism was beyond the scope of the current simulation.

GENERAL DISCUSSION

This study addresses two approaches to the acquisition of inflectional cues by adult L2 learners. Rule-driven acquisition, which is determined by the complexity of an inflectional paradigm, should lead to an advantage in learning the less complex German case marking. Our assessment of paradigm complexity was based on the principles of paradigm formation (MacWhinney, 1978; Pinker, 1984), which assume that acquisition of inflectional morphology requires the discovery of the grammatical dimensions underlying the inflectional paradigm.

Associative learning, on the other hand, is determined by the strength of inflections as cues to sentence interpretation. Using cue-validity estimations based on the Competition Model, we predicted that Russian case marking should be acquired faster. This is what the data show: Russian case marking is learned faster if language familiarity is controlled for. The differences between Russian and German in the validity of case marking arise from differences in the availability of case markers in the input, which, in turn, are determined by the different patterns of neutralization within the paradigms: Whereas the overall amount of neutralization and syncretism is higher in Russian, neutralization of case is actually more frequent in German. Thus, for inflectional marking to be reliable, the uniqueness of the markers is not important but how they are distributed in the paradigm is. Inflectional syncretism keeps the number of inflections small by maintaining the reliability of marking, whereas case neutralization keeps the number of inflections small by introducing ambiguity into the system.

We were able to show that the differences between Russian and German in the validity of inflectional cues can be detected in on-line processing: Generally, strong case markers lead to processing benefits, particularly if they appear early in the sentence. Moreover, sufficiently high strength of case marking, as in Russian, can override the effects of other cues so that they do not affect on-line processing. On the other hand, low strength of case marking gives room for the influence of supporting cues as indicated by the benefits from first-noun animacy in German. Here, the immediate on-line effect of a semantic cue like animacy can be taken as evidence against the operation of an encapsulated morphosyntactic module during comprehension and in support of interaction between all types of information available in the input. In sum, our results argue for a learning mechanism that is sensitive to the distributional characteristics of cues in the input regardless of whether they are morphological, syntactic, or semantic in nature.

The basic results of our experiment are compatible with a connectionist approach toward language learning. We used a connectionist simulation to confirm and extend the predictions of the Competition Model. The match between human data and simulation performance supports the notion that L2 learning in adults has a large associative component. It also demonstrates that connectionist models are useful tools for exploring the learning of inflectional morphology. Given that connectionist models have been used to model various aspects of L2 acquisition in children (Plunkett & Marchman, 1991; Rumelhart & McClelland, 1986, 1987), the successful application of connectionism to L2 learning suggests that associative learning represents a general mechanism that operates in both children and adults.

It can be argued that the language difference in the use of case markers does not necessarily arise from differences in the statistical distribution of case marking in the two languages but may be related to differences in metalinguistic awareness of case marking. In particular, the fact that both English and German have determiners that mark definiteness may obscure the fact that the German determiner carries other morphological functions as well. More generally speaking, the perceived distance between the L1 and the L2 is smaller for the learners of German, which might encourage transfer and minimize these learners' awareness of the morphological function of the determiner. Although we cannot completely exclude the possibility that this factor might have affected the results, we can present two arguments that make this an unlikely explanation. First, the connectionist network mimics the obtained effects without any information about the syntactic structure of the case markers and by keeping the similarity between L1 and L2 constant for both languages. Thus, an explanation based on the statistical distribution of case marking in the language is the more parsimonious one because it can account for the performance of both the learners and the model. Second, the differential effects of animacy information in the two languages cannot be explained in terms of perceived distance between L1 and L2 but are compatible with an associative learning mechanism that is sensitive to the statistical properties of the language.

In conclusion, we would like to point to some directions for future research on L2 learning of inflectional morphology that are derived from the present study. First, the approach of contrasting a case-marking system low in complexity and low in cue strength with a case-marking system high in complexity and high in cue strength does not allow us to assess the independent contributions of complexity versus strength in learning. All we can say based on these data is that the strength of individual cues appears to have a stronger effect on learning than the complexity of the entire paradigm. However, we cannot exclude the possibility that both factors combine additively so that learning is still easier in less complex paradigms regardless of neutralization and, ultimately, cue strength. This question remains a topic for future research that needs to examine a broader range of inflectional systems. Our simulation has demonstrated that connectionist models can be used for exploration and hypothesis generation if the morphological variation between languages is confounded by other factors or if the appropriate languages are not accessible for empirical study.

Second, this research has focused exclusively on comprehension, arguing that it is the validity of inflections as cues to sentence and discourse interpretation that determines the speed of learning. However, the learner also needs to be able to use those inflections correctly when attempting to express thoughts in the L2. Thus, the cross-linguistic comparison of the use of inflections, and morphosyntactic cues in general, in language production becomes critical for a full understanding of the L2 learning process. The relations between the functional structure of an intended message and the inflectional means that can be used to express these functions in production may not be isomorphic. In other words, ambiguity as found in case neutralizations is detrimental for comprehension but may be beneficial for production because it involves fewer inflectional choices for expressing a variety of functions. Thus, the distributional characteristics relevant for comprehension and for production may well differ, which could account for the lag between comprehension and production commonly observed in L2 learning. Although the present study makes no claims about the active use of case marking in the production of Russian and German by L2 learners, it suggests that connectionist models are valuable tools for exploration because they are able to capture the effects of statistical characteristics of the language in various tasks.

(*Received 7 May 1997*)

NOTES

1. The Competition Model also assumes that processing characteristics may affect the acquisition of cues (MacWhinney, 1987b, 1992; McDonald, 1986, 1987a, 1989). These factors, which are subsumed under the term *cue cost*, depend on perceptual salience and memory demands. For the purpose of the present study, we will ignore any language differences in cue cost. This is justified by the fact that perceptual salience of inflections is increased and memory load is decreased in written language, which constitutes a significant part of the input to L2 learners.

2. Russian has almost no animate neuter nouns, so the animacy distinction for neuter nouns may be disregarded. Still, the remaining total of 60 cells is nonetheless a clearly higher number of cells than in German.

3. This estimate does not account for the different types of plural formation in German. If those combinations of articles and suffixes that are related to plural formation were counted as well, the inflections-to-cells ratio would be even higher, supporting the idea that the amount of neutralization and syncretism is lower in German than in Russian.

4. The word class of nonwords can be determined based on the derivational suffix if the word is a pseudo-derivative. Furthermore, nonsense words can unambiguously be recognized as nouns in written German because the rules of German orthography require capitalization of all nouns.

5. The latency data in the lexical decision task are not relevant for assessing word sensitivity and will not be discussed further. The interested reader is referred to Kempe and MacWhinney (1996).

REFERENCES

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX Lexical Database*. Philadelphia: Linguistic Data Consortium.

Bates, E. & MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of language processing* (pp. 3–73). New York: Cambridge University Press.

Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam: Benjamins.

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

Clahsen, H., & Muysken, P. (1986). The availability of UG to adult and child learners: A study of the acquisition of German word order. *Second Language Research*, *2*, 93–119.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphical system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavioral Research Methods, Instrumentation, and Computation*, *25*, 257–271.

DeKeyser, R. M. (1995). Learning second language grammar rules. *Studies in Second Language Acquisition*, *17*, 379–410.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Elman, J. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–212.

Felix, S. W. (1987). *Cognition and language growth*. Dordrecht: Foris.

Gass, S. (1987). The resolution of conflicts among competing systems: A bidirectional perspective. *Applied Psycholinguistics*, *8*, 329–350.

Gasser, M. (1990). Connectionism and universals of second language acquisition. *Studies in Second Language Acquisition*, *12*, 179–199.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley.

Harrington, M. (1987). Processing transfer: Language-specific strategies as a source of interlanguage variation. *Applied Psycholinguistics*, *8*, 351–378.

Hernandez, A., Bates, E., & Avila, L. (1994). On-line sentence interpretation in Spanish-English bilinguals: What does it mean to be "in between"? *Applied Psycholinguistics*, *15*, 417–446.

Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99.

Kail, M. (1989). Cue validity, cue cost, and processing types in French sentence comprehension. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of language processing* (pp. 77–117). New York: Cambridge University Press.

Kempe, V., & MacWhinney, B. (1996). The crosslinguistic assessment of foreign language vocabulary learning. *Applied Psycholinguistics*, *17*, 149–183.

Kempe, V., & MacWhinney, B. (1997). *On-line processing of morphological cues: Evidence from Russian and German*. Manuscript submitted for publication. University of Toledo.

Kilborn, K. (1989). Sentence processing in a second language: The timing of transfer. *Language and Speech*, *32*, 1–23.

Kilborn, K., & Cooreman, A. (1987). Sentence interpretation strategies in adult Dutch-English bilinguals. *Applied Psycholinguistics*, *8*, 415–431.

Krashen, S. (1981). *Second language acquisition and second language learning*. New York: Oxford University Press.

Li, P., Bates, E., & MacWhinney, B. (1993). Processing a language without inflections: A reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language*, *32*, 169–192.

Liu, H., Bates, E., & Li, P. (1992). Sentence interpretation in bilingual speakers of English and Chinese. *Applied Psycholinguistics*, *13*, 451–484.

MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development*, *43*(Whole No. 1).

MacWhinney, B. (1987a). Applying the competition model to bilingualism. *Applied Psycholinguistics*, *8*, 315–327.
MacWhinney, B. (1987b). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Erlbaum.
MacWhinney, B. (1992). Transfer and competition in second language learning. In R. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 371–390). Amsterdam: Elsevier.
MacWhinney, B., & Bates, E. (Eds.). (1989). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.
MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*, 127–150.
MacWhinney, B., & Chang, F. (1995). Connectionism and language learning. In A. Nelson (Ed.), *Basic and applied perspectives on learning, cognition and development: The Minnesota Symposia on Child Psychology*, *28*(pp. 33–57). Mahwah, NJ: Erlbaum.
MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. L. (1989). Language learning: Cues or rules? *Journal of Memory and Language*, *28*, 255–277.
Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, *29*, 189–256.
Matthews, P. (1974). *Morphology*. New York: Cambridge University Press.
McDonald, J. L. (1986). The development of sentence comprehension strategies in English and Dutch. *Journal of Experimental Child Psychology*, *41*, 317–335.
McDonald, J. L. (1987a). Assigning linguistic roles: The influence of conflicting cues. *Journal of Memory and Language*, *26*, 100–107.
McDonald, J. L. (1987b). Sentence interpretation in bilingual speakers of English and Dutch. *Applied Psycholinguistics*, *8*, 379–414.
McDonald, J. L. (1989). The acquisition of cue-category mappings. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of language processing* (pp. 375–396). New York: Cambridge University Press.
McDonald, J. L., & Heilenman, K. (1991). Determinants of cue strength in adult first and second language speakers of French. *Applied Psycholinguistics, 12*, 313–348.
Mimica, I., Sullivan, M., & Smith, S. (1994). An on-line study of sentence interpretation in native Croatian speakers. *Applied Psycholinguistics*, *15*, 237–261.
Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a Parallel Distributed Processing Model of language acquisition. *Cognition*, *29*, 73–193.
Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition*, *38*, 43–102.
Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by back-propagation. In D. Rumelhart & J. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.
Rumelhart, D., & McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 216–271). Cambridge, MA: MIT Press.
Rumelhart, D., & McClelland, J. L. (1987). Learning the past tenses of English verbs: Implicit rules or parallel distributed processes? In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition* (pp. 195–248). Hillsdale, NJ: Erlbaum.
Sasaki, Y. (1991). English and Japanese interlanguage comprehension strategies: An analysis based on the Competition Model. *Applied Psycholinguistics*, *12*, 47–73.
Schwartz, B. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition*, *15*, 147–163.
Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.
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.
Taraban, R., & Palacios, J. M. (1993). Exemplar models and weighted cue models in category learning. In G. V. Nakamura, R. Taraban, & D. Medin (Eds.), *The psychology of learning and motivation: Vol. 29. Categorizations by humans and machines* (pp. 91–127). San Diego, CA: Academic Press.
Taraban, R., & Roark, B. (1996). Competition in learning language-based categories. *Applied Psycholinguistics*, *17*, 125–148.
Zasorina, L. N. E. (1977). *Častotnyj slovar' russkogo jazyka*. Moscow: Russkij Jazyk.

APPENDIX A

SOURCES FOR THE CORPUS ANALYSIS OF RUSSIAN AND GERMAN L2 TEXTBOOKS

Russian

Bitekhtina, G., Davidson, D., Dorofeyeva, T., & Fedyanina, N. (1988). *Russian*. Moscow: Russkyj Jazyk.

Leed, R., Nakhimovsky, A., & Nakhimovsky, A., (1981). *Beginning Russian. Vol. 1*. Columbus, OH: Slavica.

Leed, R., Nakhimovsky, A., & Nakhimovsky, A., (1981). *Beginning Russian. Vol. 2*. Columbus, OH: Slavica.

Stepanova, E. M., Ievlev, Z. N., Trushina, L. B., & Baker, R. L. (1984). *Russian for everybody*. Moscow: Russkyj Jazyk.

Townsend, C. E. (1981). *Continuing with Russian*. Columbus, OH: Slavica.

German

Byrnes, H., & Fink, S. (1987). *Wendepunkt: Intermediate German for proficiency*. Boston, MA: Heinle & Heinle.

Crean, J. E., Scott, M., & Briggs, J. (1993). *Deutsche Sprache und Landeskunde*. New York: McGraw-Hill.

Lohnes, W. F. W., Strothmann, F. W., & Petig, W. E. (1989). *German: A structural approach*. New York: Norton.

Moeller, J., Liedloff, H., Adolph, W. R., Kirmse, C., & Lalande, J. F. (1992). *Deutsch heute: Grundstufe*. Boston: Houghton Mifflin.

Terell, T. D., Tschirner, E., Nikolai, B., & Genzmer, H. (1996). *Kontakte: A communicative approach*. New York: McGraw-Hill.

APPENDIX B

NOUNS USED IN THE EXPERIMENTAL STUDY

	Animate nouns		Inanimate nouns	
Language	Case marked	Case neutralized	Case marked	Case neutralized
German	der Vater der Sohn	die Mutter die Tochter	der Löffel der Teller	die Blume die Torte
Russian	otec syn	mat′ doč′	ložka tarelka	cvetok tort
English translation	father son	mother daughter	spoon plate	flower cake

APPENDIX C

EXAMPLE SENTENCES USED IN THE EXPERIMENTAL STUDY

Condition	German	Russian
AA —	Die Mutter sucht die Tochter.	Mat' iščet doč'.
AA N–	Der Vater sucht die Tochter.	Otec iščet doč'.
AA –A	Die Mutter sucht den Sohn.	Mat' iščet syna.
AA NA	Der Vater sucht den Sohn.	Otec iščet syna.
AA —	Die Tochter sucht die Mutter.	Doč' iščet mat'.
AA –N	Die Tochter sucht der Vater.	Doč' iščet otec.
AA A–	Den Sohn sucht die Mutter.	Syna iščet mat'.
AA AN	Den Sohn sucht der Vater.	Syna iščet otec.
AI —	Die Mutter sucht die Torte.	Mat' iščet tort.
AI N–	Der Vater sucht die Torte.	Otec iščet tort.
AI –A	Die Mutter sucht den Teller.	Mat' iščet tarelku.
AI NA	Der Vater sucht den Teller.	Otec iščet tarelku.
AI —	Die Tochter sucht die Blume.	Doč' iščet cvetok.
AI –N	Die Tochter sucht der Löffel.	Doč' iščet ložka.
AI A–	Den Sohn sucht die Blume.	Syna iščet cvetok.
AI AN	Den Sohn sucht der Löffel.	Syna iščet ložka.
IA —	Die Blume sucht die Tochter.	Cvetok iščet doč'.
IA N–	Der Löffel sucht die Tochter.	Ložka iščet doč'.
IA —A	Die Blume sucht den Sohn.	Cvetok iščet syna.
IA NA	Der Löffel sucht den Sohn.	Ložka iščet syna.
IA —	Die Torte sucht die Mutter.	Tort iščet mat'.
IA –N	Die Torte sucht der Vater.	Tort iščet otec.
IA A–	Den Teller sucht die Mutter.	Tarelku iščet mat'.
IA AN	Den Teller sucht der Vater.	Tarelku iščet otec.
II —	Die Blume sucht die Torte.	Cvetok iščet tort.
II N–	Der Löffel sucht die Torte.	Ložka iščet tort.
II –A	Die Blume sucht den Teller.	Cvetok iščet tarelku.
II NA	Der Löffel sucht den Teller.	Ložka iščet tarelku.
II —	Die Torte sucht die Blume.	Tort iščet cvetok.
II –N	Die Torte sucht der Löffel.	Tort iščet ložka.
II A–	Den Teller sucht die Blume.	Tarelku iščet cvetok.
II AN	Den Teller sucht der Löffel.	Tarelku iščet ložka

Note. The letter sequence in the abbreviations of the condition names should be read as follows: N1-Animacy: animate (A) versus inanimate (I), N2-Animacy: animate (A) versus inanimate (I), N1-Marking: nominative (N), accusative (A), or unmarked (–), N2-Marking: nominative (N), accusative (A), or unmarked (–). Configuration (SVO vs. OVS) can be determined from the sequence of case marking (e.g., in SVO, the first noun is nominative or the second noun is accusative).

APPENDIX D

CONDITION MEANS OF FIRST-NOUN (N1) CHOICES AND DECISION LATENCIES IN THE FAMILIARITY-MATCHED LEARNER SUBGROUPS AND IN THE CONNECTIONIST MODEL TRAINED ON RUSSIAN VERSUS GERMAN

	Russian				German			
Condition	N1 choice: data	*p* (N1 choice): model	End-adj. latency: data	Latency estimate: model	N1 choice: data	*p* (N1 choice): model	End-adj. latency: data	Latency estimate: model
AA —	.903	.729	–225	2.914	.974	.954	–1251	1.218
AA N–	.944	.962	–631	1.186	.947	.968	–1440	1.150
AA –A	.917	.766	–290	2.859	.987	.956	–1019	1.217
AA NA	.900	.962	–661	1.184	1.000	.969	–1500	1.148
AA —	.931	.729	–297	2.914	1.000	.954	–1336	1.218
AA –N	.587	.501	–161	2.868	.618	.905	–1045	1.758
AA A–	.258	.389	–805	1.620	.474	.470	–1406	1.571
AA AN	.228	.406	–979	1.617	.414	.392	–1551	1.556
AI —	.929	.813	–529	2.842	.987	.956	–1211	1.219
AI N–	.944	.963	–710	1.186	.987	.969	–1688	1.149
AI –A	.985	.825	–264	2.829	1.000	.956	–1088	1.220
AI NA	.972	.963	–750	1.187	.987	.969	–1624	1.148
AI —	.916	.813	–420	2.842	.918	.956	–1326	1.219
AI –N	.549	.460	–224	2.837	.728	.929	–1041	1.491
AI A–	.250	.457	–939	1.753	.532	.836	–1600	1.930
AI AN	.304	.363	–1043	1.565	.401	.422	–1378	1.550
IA —	.815	.600	–250	2.752	.716	.446	–1149	2.485
IA N–	.873	.927	–834	1.397	.974	.855	–1317	1.788
IA –A	.986	.712	–163	2.715	1.000	.660	–1208	2.661
IA NA	.957	.927	–1090	1.391	.974	.878	–1458	1.765
IA —	.786	.600	–274	2.752	.841	.446	–1342	2.485
IA –N	.505	.374	–192	2.505	.596	.415	–1053	2.466
IA A–	.210	.402	–1052	1.818	.487	.394	–1314	1.702
IA AN	.239	.352	–1302	1.818	.455	.415	–1312	1.690
II —	.888	.773	–457	2.726	.920	.785	–1294	2.689
II N–	.957	.928	–896	1.388	.920	.881	–1450	1.752
II –A	.944	.822	–296	2.710	1.000	.897	–1065	2.630
II NA	.915	.929	–959	1.387	.947	.886	–1243	1.748
II —	.903	.773	–465	2.726	.959	.785	–1461	2.689
II –N	.448	.361	–244	2.508	.653	.383	–990	2.478
II A–	.231	.502	–1093	1.992	.432	.428	–1356	1.748
II AN	.281	.342	–1203	1.816	.479	.360	–1047	1.718

Note. For an explanation of the abbreviations of the condition names, see Appendix C.

APPENDIX E

ESTIMATING THE PROBABILITY OF FIRST-NOUN CHOICES AND THE DECISION LATENCIES FROM THE OUTPUT ACTIVATION OF THE NETWORK

Estimations are based on the strength of the output activation obtained after the first noun (a_1) and after the second noun (a_2). Because the model had only one output unit, the strength of each response alternative (0 vs. 1) after the first (s_1) vs. after second noun (s_2) was calculated as the scaled difference from the midpoint according to the formulas $s_1 = 2\ |0.5 - a_1|$ and $s_2 = 2\ |0.5 - a_2|$.

Estimating the Probability of First-Noun Choices.

The probability of first-noun choices was estimated as $p(N1) = a_1\ s_1 + a_2\ (1 - s_1)$. This formula takes into account that the strength of the response alternative after the first noun determines to what extent the response is affected by cues from the second noun.

Estimating the Decision Latencies.

Decision latencies were estimated using the equation $L_{estim} = d_1 + (1 - s_1) + (1 - s_1)\ (d_2 + (1 - s_2) + |a_1 - a_2|)$, where d_1 and d_2 represent the duration of the first and second noun, respectively, which were both set to a value of 1. The lower the value of s_1, the longer a decision will be delayed and the more the latency will depend on s_2. The higher the value of s_1, the smaller the impact of the activation value reached after the second noun. Therefore, the term that accounts for the effects of the second noun is weighted by $(1 - s_1)$. The difference $|a_1 - a_2|$ represents the additional effort required to revise an initial interpretation. For example, if the first noun is interpreted as the agent ($a_1 > .5$) and the second noun contradicts this initial interpretation ($a_2 < .5$), additional time is needed for the reassignment of thematic roles.

It is perhaps more accurate to assume a nonlinear relationship between response strength and decision latency, but we have found that estimations based on nonlinear relations or networks using cascaded back-propagation do not improve the overall model fit or the fit to specific effects.