Edinburgh Research Explorer

Quantifying cross-linguistic influence with a computational model: A study of case-marking comprehension

Quantifying cross-linguistic influence with a computational model: A study of case-marking comprehension

Yevgen Matusevych, Afra Alishahi, Ad Backus

Tilburg University

Abstract

Cross-linguistic influence (CLI) is one of the key phenomena in bilingual and second language learning. We propose a method for quantifying CLI in the use of linguistic constructions with the help of a computational model, which acquires constructions in two languages from bilingual input. We focus on the acquisition of case-marking cues in Russian and German and simulate two experiments that employ a picture-choice task tapping into the mechanisms of sentence interpretation. Our model yields behavioral patterns similar to human, and these patterns can be explained by the amount of CLI: the negative CLI in high amounts leads to the misinterpretation of participant roles in Russian and German object-verb-subject sentences. Finally, we make two novel predictions about the acquisition of case-marking cues in Russian and German. Most importantly, our simulations suggest that the high degree of positive CLI may facilitate the interpretation of object-verb-subject sentences.

Keywords: cross-linguistic influence, computational model, case marking

1. Introduction

1.1. Quantifying cross-linguistic influence

The phenomenon of cross-linguistic influence (CLI) is central to our understanding of bilingual and second language (L2) learning. Languages interact in the bilingual mind, and studies of CLI intend to describe various types of such interaction.[1] One challenging issue that has long interested scholars is measuring the amount of CLI – that is, quantifying the extent to which linguistic representations from one language affect the use of the other language(s). Weinreich (1968) suggested that "no easy way of measuring or characterizing the total impact of one language on another in the speech of bilinguals has been, or probably can be, devised" (p. 63). Measuring the amount of CLI is important to understand to

what extent the knowledge of one language is beneficial (in case of positive CLI) or damaging (in case of negative CLI) for the acquisition of other languages.

One common method to measure CLI is through the so-called error analysis: look at the frequency of linguistic errors in a group of learners with a particular first language (L1) background and estimate the contribution of negative CLI to the non-native L2 use (Born, 1985; Grauberg, 1971; see Palmberg, 1976 for a relevant bibliography). At the same time, CLI is not the only source of non-native language use: other factors such as overgeneralization may play a role, and the non-native use is often caused by a combination of factors (Jordens, 1977). This is why the exact methodology for identifying CLI is not straightforward: it has been argued that one needs to show that the learners within a particular group make similar mistakes, that the mistakes are different across the L1 groups, and that the mistakes have their linguistic equivalents in the learners' L1 (Jarvis, 2000). Given the multitude of interfering variables (e.g., proficiency, learning history, aptitude), it is difficult to identify with confidence all cases of the CLI influence, and to measure the amount of CLI using this method. The same problem persists in more controlled experimental settings, which employ linguistic tasks related to language production or comprehension by bilingual learners (Grosjean, 1998). The number of interfering variables can be reduced in research on multilingual speakers: studying learners' third language use allows for identifying the instances of L1 and L2 influence at individual level (e.g., De Angelis & Selinker, 2001), similar to a within-subject design in experimental studies, but this "individual" approach makes it difficult to generalize over the group of learners.

Another issue related to the described methodologies is that the resulting CLI measures are grounded in language use. This may constitute a methodological challenge whenever such measures are used to *predict* the learner's language use, leading to circular reasoning.

These limitations can be overcome in cognitive computational models of bilingual language learning and use, which allow researchers to look inside the "black box" of linguistic representations. While no computational modeling studies focused on measuring CLI, some of such studies in the field of bilingualism employed quantitative measures that reflected the amount of CLI in the respective models. In particular, Zhao and Li (2010) simulated bilingual acquisition of Chinese and English words using a self-organizing neural network model. The learning process in each simulation yielded a spatial representation (map) of the bilingual lexicon. To explain how their computational model arrived at a particular type of map, the authors computed the average Euclidean distance between lexical translation equivalents in multiple pairs: that is, how far an English word (e.g., *star*) was located from its Chinese

equivalent (*星星* 'star') on the map. A shorter average distance means that many translation equivalents are located next to each other, which is the evidence of high CLI: the location of L1 lexemes has influenced the placement of the corresponding L2 lexemes. Vice versa: a longer average distance corresponds to smaller amount of CLI, because the location of L1 lexemes has not played the determining role in the placement of their L2 equivalents.

In a similar type of model, Shook and Marian (2013) studied bilingual speech comprehension in English and Spanish. They employed an online measure, so-called language activation score. This measure showed how strongly the lexical representations from a particular language (e.g., Spanish) were activated on average, when the model was given a word in either the same or a different language (English). One can argue that the activation score for the non-target language reflects the amount of CLI.

The described measures and the respective models, however, do not go beyond the lexeme level, while there are no computational modeling studies of CLI at the level of abstract constructions. To address this gap in the literature, in this article we use a computational model of learning argument structure constructions from bilingual input. We choose this model, because it has been used for simulating bilingual learning of argument structure constructions (Matusevych, Alishahi, & Backus, 2016b, 2017), and it allows for measuring the amount of CLI in this domain. Our goal in the present study is to demonstrate how the amount of CLI can be measured in the learning and use of such constructions, and how a CLI measure can be employed to explain the patterns of language use observed in the model. More specifically, we study the acquisition and interpretation of case-marking cues in Russian and German transitive sentences: as we show below, this is one of the aspects discussed in the literature, and the relevant experimental results from human participants are available.

1.2. Interpretation of transitive sentences

In some languages, such as English, French, Hebrew, and many others, transitive sentences are characterized by a fixed subject-verb-object (SVO) word order – see example (1) from Yoshimura and MacWhinney (2010). In other languages, the word order is more flexible: German transitive sentences can have SVO (2) as well as OVS word order (3).

(1) The dog chases the bear.

(2) Der   Hund-∅   jägt   den   Bär-en.

      ART.M.NOM.SG dog-M.NOM   chase:3SG   ART.M.ACC.SG   bear-M.ACC

      'The dog chases the bear.'

(3)	Den	Bär-en	jägt	der	Hund-∅.
	ART.M.ACC.SG	bear-M.ACC	chase:3SG	ART.M.NOM.SG	dog-M.NOM

‘The dog chases the bear.’

To correctly interpret OVS sentences, speakers rely on other cues than the word order: morphological case marking, as in (3), but also animacy, noun–verb agreement, and others. However, learners of a language allowing for OVS sentences may rely on the word order cue and misinterpret participant roles in such sentences; this happens both in adult L2 learners (e.g., Isabelli, 2008; Kempe & MacWhinney, 1998; VanPatten, 1996) and in monolingual children learning various languages (e.g., Kim, O'Grady, & Cho, 1995; Schaner-Wolles, 1989; Smolík, 2015). Speaking of young monolingual German children, it has been suggested that they start by acquiring the more prototypical and more frequent SVO form first (Dittmar, Abbot-Smith, Lieven, & Tomasello, 2008). The situation with bilingual and L2 learners is more complex, because CLI may be at play. There are two general views on the role of CLI in the misinterpretation of transitive sentences.

The first view is represented by the First-Noun Principle (e.g., VanPatten, 1996, 2012). According to this principle, learners universally tend to assign the agent role to the first noun or pronoun in a given sentence, while the effect of CLI is negligible. Existing studies have argued that the First-Noun Principle can explain data from L2 learners of various languages: English, French, German, and others (see an overview by Lee & Malovrh, 2009).

The alternative view explains the misinterpretation of OVS sentences by CLI from learners' L1. Under this view, L2 learners adhere to the interpretation strategy which is standard in their L1: if learners do not encounter OVS sentences in their L1, they will misinterpret such L2 sentences as SVO. This general view is compatible with multiple acquisition theories (see an overview by Hanson, Aroline, & Carlson, 2014), but the two accounts mentioned most frequently in this respect are the Unified Competition Model (MacWhinney, 2012) and the L1 Transfer Principle (VanPatten, 2015b).

According to the Competition Model (Gass, 1987; Kempe & MacWhinney, 1998; Kilborn & Cooreman, 1987; McDonald, 1987; Mimica, Sullivan, & Smith, 1994; Morett & MacWhinney, 2013, etc.), learners of both L1 and L2 attend to multiple cues in the input, such as word order, case marking, animacy, and others. Importantly, languages differ in the relative importance of various cues (e.g., case marking plays little role in English), and L1 speakers learn to attend to some cues more than to others. These attentional preferences, or cue strengths, are acquired based on the *validity* of the cues. Validity can be calculated using a linguistic corpus, as a product of two other values: cue *availability* and *reliability*. A cue is

available whenever it is present as a marker of a particular function: for example, the nominal case marking of the subject may help discriminating between this subject and the object in the sentence. A cue is reliable whenever its presence ensures the right choice of the function: for example, the nominal case marking of the object would make the cue unreliable for this sentence. The acquired cue strengths are initially transferred to an L2. As a result, when L1 speakers of a language with fixed SVO word order (e.g., English) start learning an L2 in which OVS sentences are allowed (e.g., German), they fail to attend to case marking and misinterpret OVS sentences as SVO.

The L1 Transfer Principle complements the First-Noun Principle mentioned above. Given the combination of the two, learners still tend to interpret the first noun as the agent of a sentence, yet this general strategy is modulated by their L1 knowledge. As an example, Isabelli (2008) demonstrated that L1 Italian students learning L2 Spanish could interpret Spanish OVS sentences better than their L1 English peers. This is because OVS sentences are common in Italian and Spanish, but not in English.[2] At the same time, L1 Italian speakers still performed lower on OVS sentences than on SVO sentences, which the authors considered as evidence for the First-Noun Principle. The data collected from Italian speakers with less exposure to Spanish could potentially clarify the exact role of each principle.

To summarize, we still need to learn the exact contribution of CLI to the interpretation of transitive sentences, as well as the role of cue competition in this task. To investigate these issues, we simulate an experimental task employed in the two target studies described below, and quantify the impact of CLI in our model's language use with a novel quantitative measure.

The rest of the article is organized as follows. First, we briefly introduce two studies on which we focus in our simulations. These studies investigate the interpretation of transitive sentences with case-marking cues by learners whose L1 does not employ such cues. This is followed by the presentation of our computational model, where we also explain how it allows for quantifying CLI. Next, in two sets of simulations we demonstrate that the model's linguistic behavior in the target task is similar to that observed in human learners. The findings are explained in terms of the amount of CLI. Finally, we make two novel predictions on how the model would perform on the same task when trained on different language pairs, and run two additional sets of simulations to see whether our model supports these predictions. Overall, this gives four sets of simulations:

1. Interpretation of German sentences by bilingual learners whose other language has no case marking (Janssen, Meir, Baker, & Armon-Lotem, 2015, 2016).

2. Interpretation of German and Russian sentences by L2 learners whose L1 has no case marking (Kempe & MacWhinney, 1998).

3. Interpretation of German sentences by Russian–German bilingual learners (novel).

4. Interpretation of Russian sentences by bilingual learners with various additional languages (novel).

2. Target studies on case-marking comprehension

Studies on the interpretation of case-marking cues in transitive sentences have mainly focused on adult L2 acquisition (Kempe & MacWhinney, 1998; McDonald, 1987; Mimica et al., 1994; Morett & MacWhinney, 2013, etc.), while similar studies with early bilinguals have been rare (but see Janssen et al., 2015; O'Shannessy, 2011). We focus on one study from each population: a study with bilingual and monolingual Russian children by Janssen et al. (2015), and a study with adult learners of Russian and German (Kempe & MacWhinney, 1998). In the following sections we explain why we choose these two studies. First, however, we describe a picture-choice task employed in both of them.

2.1. Picture-choice task

In this task, participants hear a sentence and see two pictures containing alternative interpretations of the sentence. The participants have to choose the picture which in their opinion corresponds to the correct interpretation of the sentence. In the two target studies, the picture-choice task is employed to study the comprehension of competing cues, in particular case marking and word order. The target sentences include two nouns (nominative and accusative/dative) and a verb, and the two pictures depict the same event, but the participant roles are swapped in one of the pictures. An example from Janssen et al. (2015):

(4) Петух-∅   трогает   зме-ю.

petuh-∅   trogaet   zmey-u

rooster-M.NOM   touch:3SG   snake-F.ACC

'The rooster touches the snake.'

The sentence (4) is accompanied by two pictures (Figure 1), depicting either a rooster touching a snake, or a snake touching a rooster.
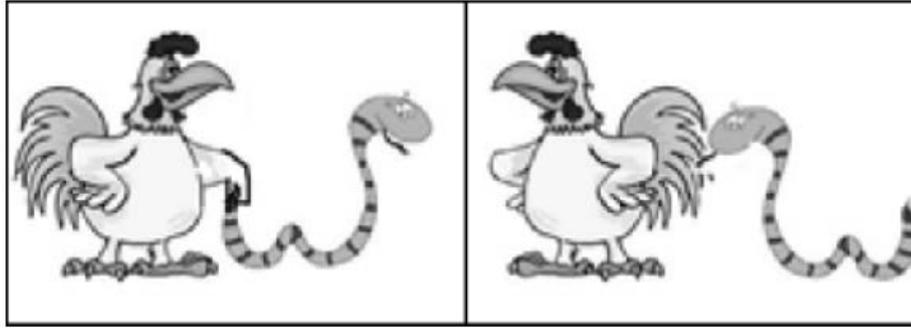
Figure 1: Accompanying pictures in the study of Janssen et al. (2015). Reprinted from "On-line comprehension of Russian case cues in monolingual Russian and bilingual Russian-Dutch and Russian-Hebrew children", 2015, by B. Janssen, N. Meir, A. Baker, and S. Armon-Lotem. In E. Grillo & K. Jepson (Eds.), *Proceedings of the 39th Annual Boston University Conference on Language Development*, p. 272. Copyright 2015 by B. Janssen. Reprinted with permission.

2.2. Bilingual and monolingual Russian children

Janssen et al. (2015) work with Russian monolingual children, as well as with Russian–Dutch and Russian–Hebrew bilingual children. While Russian is characterized by a free word order and systematic case marking of nouns, the opposite holds for Dutch and Hebrew: these two languages have much stricter word orders and no morphological cases on nouns. The case-marking cue is important in Russian: it marks the thematic roles of the nouns. At the same time, in Dutch and Hebrew the word order is often the only cue that allows for distinguishing between SVO and OVS sentences.

In this study, the picture-choice task is employed to investigate whether this difference between Russian and Dutch/Hebrew leads to any differences in sentence interpretation by Russian monolingual and Russian–Dutch or Russian–Hebrew bilingual children. Some of the presented sentences had SVO order, where the word order cue and the case-marking cue supported and complemented each other (the converging cue condition), as in (4) above. Other sentences had OVS word order with the conflicting cues (the conflicting cue condition), such as (5):

(5) Жираф-а   видит   петух-∅.

zhiraf-a   vidit   petuh-∅

giraffe-M.ACC   see:3SG   rooster-M.NOM

'The rooster sees the giraffe.'

In addition to SVO and OVS sentences with a subject and a direct object, noun-verb-noun sentences with an indirect dative object were used, such as (6):

(6) Зме-е   улыбается   жираф-∅.

     zmey-e   ulybaetsya   zhiraf-∅

     snake-F.DAT   smile.at:3SG   giraffe-M.NOM

     'The giraffe smiles at the snake.'

There were 40 stimuli overall: 20 SVO sentences and 20 OVS sentences, the test verbs included four transitive verbs: *любить* 'love', *трогать* 'touch', *целовать* 'kiss', and *видеть* 'see', as well as two intransitive verbs allowing for an indirect dative object (addressee): *улыбаться* 'smile' and *звонить* 'call'.

Both monolingual and bilingual children were expected to perform high in the comprehension of the SVO sentences, but the bilingual children in the conflicting cue condition were predicted to demonstrate lower accuracy rate and longer reaction time than in the converging cue condition, and than the monolingual children in the conflicting cue condition. This is because the bilingual children may transfer the strength of the word order cue from Dutch or Hebrew into Russian, leading to the misinterpretation of the Russian OVS sentences as SVO. These predictions were met in terms of both accuracy and reaction time. Interestingly, no differences between the two bilingual groups were observed, despite the high variation reported for home language use: 61% in the Hebrew group and 17% in the Dutch group.

For us, this study presents an interesting case: first, the authors mention that their results are compatible with both the First-Noun Principle and the Competition Model. Second, this is one of the only two studies on the interpretation of case-marking cues focusing on early bilingual learning. The other one (O'Shannessy, 2011) dealt with rare languages, Lajamanu Warlpiri and Light Warlpiri, for which it was difficult to obtain the data necessary for computational simulations.

2.3. Adult L2 learners of Russian and German

Kempe and MacWhinney (1998) worked with native English adult learners of L2 Russian and L2 German, who had been exposed to the target languages in classroom for 25–26 months. The picture-choice task with transitive sentences was used. Both in Russian and in German, all the sentences had the verb *look for/find* as the predicate: *искать* in Russian, *suchen* in German. The picture-choice task was slightly different in this experiment: the alternative pictures did not depict the full event, but only the two

participants instead, and the learners had to decide which participant was the agent, defined as "who or what did the looking or finding" (Kempe & MacWhinney, 1998, p. 557). The 32 Russian and German test sentences were mutual translations of each other: 12 SVO sentences with case marking, 12 OVS sentences with case marking, and 8 SVO sentences fully neutralized in terms of their case-marking cues: these contained two nouns whose nominative and accusative cases were marked with the same morpheme, as in (7).

(7)	Die	Tochter-∅	sucht	die	Mutter-∅.
	ART.F.SG.NOM/ACC	daughter-SG.NOM/ACC	look.for:3SG	ART.F.SG.NOM/ACC	mother-SG.NOM/ACC

'The daughter looks for the mother.'

Using the methodology commonly adopted in Competition Model studies, Kempe and MacWhinney (1998) compute the availability of a cue as the number of sentences in which the cue is present divided by the total number of transitive sentences. To compute the reliability of a cue, they divide the number of sentences in which the cue correctly indicates the agent by the total number of sentences in which this cue is present. Based on their calculations, Kempe and MacWhinney (1998) show that the case-marking cue in Russian has a higher validity than in German, and this is why Russian L2 learners are more successful in the acquisition of case marking than German L2 learners: they perform the task faster (in terms of decision latencies) and more accurately than German L2 learners.

We choose this study because of its similarity to the study of Janssen et al. (2015): both employ the picture-choice task, and both focus on the comprehension of case marking in Russian. These similarities will help us to make some informed predictions about the interpretation of case-marking cues, and pre-test these predictions using computational simulations. The main difference between the two experiments is the age of the subjects, which we can also take into account in our computational simulations by manipulating the overall amount of input the model is exposed to.

3. Computational model

3.1. Model overview

In this section, we provide a brief conceptual introduction of the model, while a more complete formal description is given in the subsequent sections. The computational model we employ here learns argument structure constructions from the input data. Technically speaking, it uses unsupervised Bayesian

clustering to group incoming instances into constructions in an iterative manner. The idea comes from the study of Anderson (1991), who implemented the model of human category learning. Alishahi and Stevenson (2008, 2010) were the first to simulate the learning of argument structure constructions using the same algorithm, while Matusevych et al. (2016b, 2017) adapted the model for bilingual learning. In all the mentioned studies, the model was demonstrated to replicate certain types of human behavior in the respective domain. In the present study, the learning mechanism is updated to suit the acquisition of languages with free word order. Note that this model is not intended to simulate human construction learning in all its complexity; rather, it only employs the mechanism of bottom-up statistical learning.

The idea behind the model comes from Goldberg's construction grammar. It has been proposed that the learning of argument structure constructions can be seen as a categorization process (Goldberg, 1995; Goldberg, Casenhiser, & Sethuraman, 2004): abstract constructions emerge as a result of generalizing over individual instances. Similarly, our computational model is exposed to a number of individual verb usages, and categorizes them based on their similarity and the size (or the degree of entrenchment: see, e.g., Schmid, 2016) of the existing clusters.

L1 and L2 verb usages are not explicitly marked as such. This means that, if the model encounters an L2 usage very similar to an existing L1 construction, it may add the L2 usage to the L1 constructions, making this construction "blended". Since the blended construction already contains an L2 usage, it is more likely to attract additional L2 usages, so that the process is reinforced. At the same time, in some simulations there may be very few or no blended clusters, depending on the exact input usages and the order of their presentation. This approach follows the theories which suggest that abstract constructions may be shared between L1 and L2 (e.g., Bernolet, Hartsuiker, & Pickering, 2013). This idea is supported by existing data on cross-linguistic structural priming (e.g., Bernolet, Hartsuiker, & Pickering, 2013) and on the use of L2 constructions to comprehend L1 sentences (Higby et al., 2016).[3]

Speaking of the input to our model, individual verb usages consist of multiple features, representing an utterance and its perceptual context. Importantly for this study, while the information about the animacy of the linguistic referents is provided within one of the features (animate is one of the existing argument role properties, see section 3.2.2 below), the animacy cue does not have a status of an independent feature in the model, following the existing implementations. This makes it difficult to provide clear predictions regarding the impact of animacy on sentence interpretation, as in Kempe and MacWhinney (1998), and we refrain from carrying out the analysis of animacy in this study.

Another important issue is how the effect of CLI is manifested in this model. Whenever the model has to make a decision regarding a particular instance, it examines the full repertoire of the acquired constructions ("the constructicon") to collect evidence in favor of each alternative. Because the constructicon contains both L1 and L2 knowledge, the effect of CLI is always present, in theory. However, in practice, the amount of CLI varies: it can be negligibly small when L1 and L2 are very dissimilar and there are no blended clusters, or it can be very high when the languages are very similar and/or there are many blended clusters. All these issues are described in more detail in the following sections.

3.2. Input to the model

3.2.1. Input representations

The input to the model consists of individual verb usages, which we call argument structure (AS) instances. Each AS instance comprises multiple independent features: lexical, semantic, and syntactic. Further, we make two important distinctions: between distributional features (*FD*) and symbolic features (*FS*), and between global (*FG*) and local features (*FL*). Consider an example instance in Table 1.

Symbolic features carry values expressed by a single symbol (e.g., head predicate: *touch*; number of arguments: 2). In contrast, each value of a distributional feature is a set of elements (e.g., head properties: {ACTION, CAUSAL, MANIPULATE, PHYSICAL}). As for the global vs. local features, the former relate to the utterance or the described event as a whole (e.g., the head predicate), while the latter are tied to a particular participant of the event (e.g., an argument or its lexical meaning).

As we demonstrate in the next section, these two distinctions are important in the formal model. In particular, introducing the notion of local features helps us to simulate the learning of free word order languages in a more naturalistic manner. First, however, we briefly describe how the data sets for the model were obtained.

3.2.2. Data collection

In this study, we use a part of the available corpus described in Matusevych, Alishahi, and Backus (2016a). Specifically, we employ four small data sets of child-directed speech: Russian, German, English, and French. The sentences in these data sets were extracted from the respective corpora in CHILDES data base (MacWhinney, 2000),[4] and approximately 500 verb usages in each language were manually

annotated with the features listed in Table 1. The lexical meanings of noun arguments were automatically extracted from a lexical database WordNet (G. A. Miller, 1995).

Table 1: An AS instance for the Russian equivalent of the sentence *The snake touches the rooster*.[5]

Feature	Type	Value
Head predicate	Global, symbolic	*touch*
Head properties	Global, distributional	{ACTION, CAUSAL, MANIPULATE, PHYSICAL}
Head position	Global, symbolic	2
Number of arguments	Global, symbolic	2
Arg.1	Local, symbolic	*snake*
Arg.2	Local, symbolic	*rooster*
Arg.1 case	Local, distributional	{NOM}
Arg.2 case	Local, distributional	{GEN, ACC}
Arg.1 lexical meaning	Local, distributional	{DIAPSIDE, REPTILE, …, CAUSAL AGENT}
Arg.2 lexical meaning	Local, distributional	{CHICKEN, DOMESTIC FOWL, …, CAUSAL AGENT}
Arg.1 role properties	Local, distributional	{ACTING, ANIMATE, …, VOLITIONAL}
Arg.2 role properties	Local, distributional	{ANIMATE, CONCRETE, …, TOUCHED}
Arg.1 preposition	Local, symbolic	N/A
Arg.2 preposition	Local, symbolic	N/A
Arg.1 position	Local, symbolic	1
Arg.2 position	Local, symbolic	2

3.3. Learning process

3.3.1. Key components

During the learning, the model receives input instances one by one, and the learning consists in grouping them into clusters, which potentially correspond to argument structure constructions. The initial state of the model's knowledge is a single empty cluster. While the first instance is always placed into such an empty cluster, for any subsequent instance *I* each existing cluster *C* is considered, including an empty one. The goal is to find the "best" (most probable) cluster C_{best} for the encountered instance *I*.

$$C_{best}(I) = \arg\max_{C} P(C \mid I) \tag{1}$$

The conditional probability in (1), $P(C \mid I)$, cannot be estimated directly; therefore, the Bayes rule is applied:

$$P(C \mid I) = \frac{P(C)P(I \mid C)}{P(I)} \tag{2}$$

The denominator in (2), which is the probability of the (given) instance, has the same value for all clusters and does not affect the decision. This is why it can be excluded from the computation:

$$P(C \mid I) \propto P(C)P(I \mid C) \tag{3}$$

Equation (3) has two components: the prior probability of a cluster, $P(C)$, and the conditional probability of the instance given the cluster, $P(I \mid C)$.

The prior is set to be proportional to the number of AS instances previously put into this cluster, $|C|$, which is normalized by the total number of instances encountered so far $(N+1)$, see equation (4). The idea is that frequent categories (clusters) are more entrenched than non-frequent ones: the learner can access frequent clusters easier, and is more likely to add the new instance into such clusters.

$$P(C) = \frac{|C|}{N+1} \tag{4}$$

An empty cluster is also considered for each incoming AS instance, with potentially one member: the current instance.

The conditional probability in (3), $P(I \mid C)$, accounts for the degree of similarity between the new instance and each cluster. The main difference of the present model from its earlier versions relates to how such similarity is computed, which we explain in the next section.

3.3.2. Interpreting instances

In the existing studies with this model, the similarity between an instance and a cluster was compared in terms of each feature independently: how similar are the verb meanings in the instance and the cluster, the first arguments, the second arguments, and so on. This general approach is preserved in this study. However, consider the following two sentences (8–9) and imagine that the model first encounters an instance based on sentence (8), places it into an appropriate cluster, and then encounters an instance based on sentence (9). Without an ability to "swap" the arguments for the purpose of comparing their similarity, the model would not be able to compare the first argument giraffe in (8) to the second argument giraffe in (9), and the two instances would most probably not be grouped together, despite having nearly identical meanings.

(8) Жираф-а   видит   петух-∅.

      zhiraf-a   vidit   petuh-∅

      giraffe-M.ACC   see:3SG   rooster-M.NOM

      'The rooster sees the giraffe.'

(9) Петух-∅   видит   жираф-а.

      petuh-∅   vidit   zhiraf-a

      rooster-M.NOM   see:3SG   giraffe-M.ACC

      'The rooster sees the giraffe.'

This is why we need to ensure that the model is able to compute the similarity not only between the local features of the first argument in a new instance and in each cluster *C*, but also between features of arguments with different indexes: first to second, first to third, and so on. Such a mechanism is essential for languages with free word order.

Therefore, multiple possible interpretations *i* of the instance *I* are considered in the model. Each interpretation *i* carries exactly the same feature values as *I*, but the indexes of the local features *FL* in *i* may be swapped. In simple terms, whenever the model encounters an instance extracted from the sentence (8), it considers its original order of arguments, but also the reversed one (9). This is not to say

that this mechanism simulates what human learners do at the implementational level: it is unlikely that humans mentally swap the arguments to consider all the alternative word orders. However, humans must be able to see similarities between sentences such as (8) and (9), and this is argued to be reflected in the resulting cognitive representations: think of the notions of alternations and allostructions in construction grammar (Cappelle, 2006; Perek, 2015).

In formal terms, let us denote the value of a particular local feature in the interpretation i as FL_k^i, the value of the respective feature in the instance I as FL_k^I, and the set of all permutations for this feature $S(FL_k^I)$. Then the set of all possible interpretations $\mathbb{P}(I)$ can be defined as provided in (5).

$$\mathbb{P}(I) = \left\{ i : \forall FL_k^i \in S(FL_k^I), \forall FG_k^i = FG_k^I \right\} \tag{5}$$

This way, the model considers each possible argument order, and selects the one with the highest similarity to one of the existing clusters. This maximal similarity value is considered to be the resulting conditional probability, see equation (6).

$$P(I \mid C) = \max\left(\left\{ P(i \mid C) : i \in \mathbb{P}(I) \right\}\right) \tag{6}$$

The overall similarity value between an interpretation i and a cluster C is taken to be a product of similarities of individual features, but the individual values for all the symbolic features FS are weighed by a parameter w,[6] while the distributional features FD preserve their original similarity values (equation 7). This is necessary, because otherwise the symbolic features related to the sentence form (lexical arguments, argument positions, etc.) dominate the clustering process, and the model's decisions are informed mainly by the form of the instances, but not their meaning.

$$P(i/C) = \prod_{k=1}^{\left|FD^i\right|} P\left(FD_k^i \mid C\right) \left(\prod_{k=1}^{\left|FS^i\right|} P\left(FS_k^i \mid C\right) \right)^w \tag{7}$$

Finally, the independent similarities for symbolic and distributional features are computed differently, see (8–9).

$$P\left(FS_k^i / C\right) = \frac{\left|\{FS_k^i / FS_k^i \in FS_k^C\}\right| + \lambda}{\left|FS_k^C\right| + \lambda \left|FS_k\right|} \tag{8}$$

In equation (8), the term $\left|\{FS_k^i / FS_k^i \in FS_k^C\}\right|$ denotes how many times FS_k^i (the value of the feature FS_k observed in the interpretation i) occurs in the cluster C, and the term $\left|FS_k^C\right|$ (the total number of occurrences of the target feature in C) serves as the normalizing factor. The smoothing parameter λ is introduced both in the numerator and the denominator, but in the latter case it is multiplied by the total number of different values of the target feature in the data set. This method would not be robust for calculating the similarity in the distributional features, because their values consist of sets, and the set equality is very unlikely to hold, so that $\left|\{FS_k^i / FS_k^i \in FS_k^C\}\right| = 0$. This is why the method given in (9) is used:

$$P\left(FD_k^I / C\right) = \left(\prod_{e \in FD_k^i} P(e/C) \times \prod_{e \in FD_K \setminus FD_K^i} P(\neg e / C) \right)^{\frac{1}{|FD_k|}},$$ (9)

where $P(e/C)$ and $P(\neg e / C)$ are computed in the same way as in (8), replacing FS_k^i with the respective element e, see equation (10):

$$P\left(e_k^i / C\right) = \frac{\left|\{e_k^i / e_k^i \in FD_k^C\}\right| + \lambda}{\left|FD_k^C\right| + \lambda\left|FD_k\right|}$$ (10)

3.4. Simulated picture-choice task

At any point, the learning process can be paused, and the model is tested on the picture-choice task. The model receives a set of test stimuli, each of which includes a pair of alternatives (Table 2), and has to choose the correct one in each pair. Note that each alternative instance comprises all the features used in the input: lexical, syntactic, and semantic. The alternatives within each pair are identical, and the only difference is in the assignment of the argument roles. As it can be seen from Table 2, the role properties of the two arguments are swapped, to simulate what in human experiments is a pair of images with the participant roles reversed.

Given the two alternatives, the model computes their probability given the acquired knowledge, which can be expressed as the sum of the respective probabilities over all the acquired clusters:

$$P\left(I_A\right) = \sum_C P\left(I_A / C\right)P\left(C\right)$$ (11)

Table 2: A pair of test instances for the Russian equivalent of *The rooster touches the snake*.

Feature	Alternative 1	Alternative 2
Head predicate	*Touch*	*touch*
Head properties	{ACTION, CAUSAL, MANIPULATE, PHYSICAL}	{ACTION, CAUSAL, MANIPULATE, PHYSICAL}
Head position	2	2
Number of arguments	2	2
Arg.1	*rooster*	*rooster*
Arg.2	*snake*	*snake*
Arg.1 case	{NOM}	{NOM}
Arg.2 case	{ACC}	{ACC}
Arg.1 lexical meaning	{CHICKEN, DOMESTIC FOWL, …, CAUSAL AGENT}	{CHICKEN, DOMESTIC FOWL, …, CAUSAL AGENT}
Arg.2 lexical meaning	{DIAPSIDE, REPTILE, …, CAUSAL AGENT}	{DIAPSIDE, REPTILE, …, CAUSAL AGENT}
Arg.1 role properties	**{ANIMATE, CONCRETE, …, TOUCHED}**	**{ACTING, ANIMATE, …, VOLITIONAL}**
Arg.2 role properties	**{ACTING, ANIMATE, …, VOLITIONAL}**	**{ANIMATE, CONCRETE, …, TOUCHED}**
Arg.1 preposition	N/A	N/A
Arg.2 preposition	N/A	N/A
Arg.1 position	1	1
Arg.2 position	2	2

To compute the two probabilities in (11), we use the same methods as during the learning: equation (6) for computing the conditional probability $P(I_A / C)$, and equation (4) for the cluster's prior probability $P(C)$. After evaluating the probability of each alternative, the model selects the more probable one. As we mentioned earlier, CLI may be a factor affecting the model's choice. We next propose a measure of CLI.

3.5. Measuring the amount of CLI

The model accumulates evidence supporting each alternative from all the acquired clusters. At the same time, some clusters contribute to the decision substantially more than others, either because they are similar to the test instance, or because they are strongly entrenched in the model's knowledge. Besides, the amount of the non-target language instances in each acquired cluster differs: some clusters are based on the instances of a single language (L1 or L2), while others are blended – that is, based on data from both languages (see Figure 2). To summarize, there are two components that determine the amount of CLI given an instance I: the contribution of each cluster to the model's choice, and the number of the non-target language instances in the cluster.

If we denote the language of an instance *I* as *L(I)*, then the amount of CLI can be defined as follows:

$$CLI(I) = \sum_{C} P(I/C) P(C) \frac{\left|\{J \mid J \in C, L(J) \neq L(I)\}\right|}{|C|}, \tag{12}$$

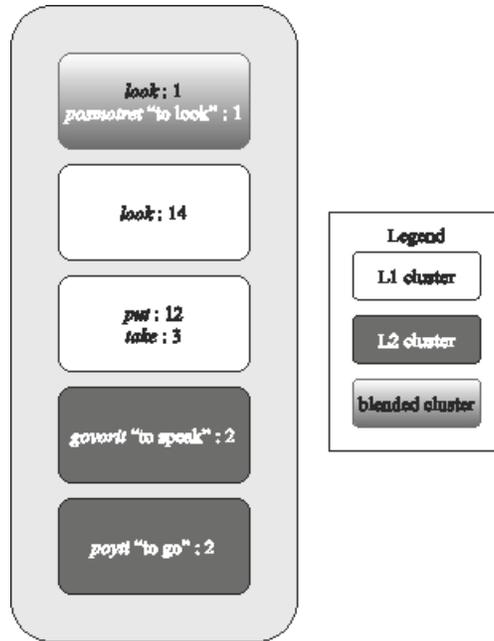where the fraction denotes the proportion of instances from the non-target language in the cluster *C*.

Figure 2: A subset of five clusters emerged in a bilingual Russian–English simulated learner (only head predicates are shown, other features omitted for simplicity).

In the picture-choice task, each pair has a correct alternative $I_{correct}$, and an incorrect alternative $I_{incorrect}$. Using equation (12), we can compute the amount of CLI independently for each alternative. In this particular task the two alternatives are competing, and the support from L1 for $I_{correct}$ can be seen as positive CLI, while the support from L1 for $I_{incorrect}$ is negative. This is why the best way to quantify the impact of CLI in the picture-choice task is to measure the difference in the amount of CLI between the two alternatives:

$$\Delta CLI(I) = CLI(I_{correct}) - CLI(I_{incorrect}) \tag{13}$$

A positive value of $\Delta CLI(I)$ would mean that the positive effect of CLI prevails, while a negative value shows that CLI is damaging for the model's decision on a particular pair of instances.

4. Simulations and results

This section presents our computational simulations of the two target experiments. This is followed by two more simulations, which allow us to make novel predictions regarding the comprehension of case-marking cues in additional language pairs.

4.1. Simulation set 1

In this experiment, we study whether our computational model performs similar to humans in the picture-choice task. Based on Janssen et al.'s (2015) results, we expect that the model will reach higher accuracy in the converging cue condition than in the conflicting cue condition. We also interpret the results in terms of CLI.

4.1.1. Simulation details

The 40 Russian stimuli from Janssen et al.'s (2015) experiment were obtained from the authors and annotated in the same way as our input data set. We had neither Hebrew nor Dutch data to simulate the same language pairs as in the original experiments, yet the results of Janssen et al. were consistent across the two groups of bilinguals, which suggests that the findings generalize on other bilingual children, as long as they speak Russian and an SVO language without case marking. Among our data sets, English and French are such languages; therefore, we simulate Russian–English and Russian–French bilinguals, in addition to Russian monolinguals.

Both in monolingual and bilingual simulations the model received a total of 400 AS instances (value established empirically): for monolinguals, these were Russian instances only, while for bilinguals the

input included Russian and English/French instances in equal proportion. After that, the model in each condition performed the picture-choice task on the 40 test instances.

4.1.2. Results
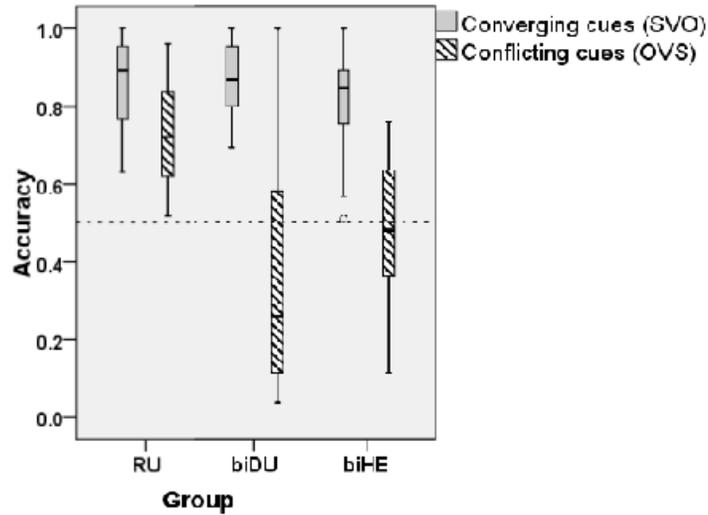
(a) Original results of Janssen et al. Reprinted from "On-line comprehension of Russian case cues in monolingual Russian and bilingual Russian-Dutch and Russian-Hebrew children", 2015, by B. Janssen, N. Meir, A. Baker, and S. Armon-Lotem. In E. Grillo & K. Jepson (Eds.), *Proceedings of the 39th Annual Boston University Conference on Language Development*, p. 273. Copyright 2015 by B. Janssen. Adapted with permission.
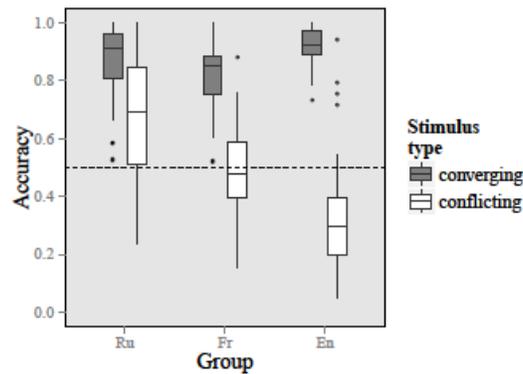
(b) Results of our simulations.

Figure 3: Simulating the experiment of Janssen et al. (2015).

Figure 3 provides a visual comparison of our results vs. human data from Janssen et al. (2015). There are three groups in each Figure: Russian monolinguals and two groups of bilinguals – Russian–Dutch and Russian–Hebrew (in the original study), or Russian–French and Russian– English (in our simulations). Each group is tested in two conditions: on the stimuli with converging cues and with conflicting cues. The accuracy is measured as the ratio of the correct choices to the total number of replies. We can observe the following similarities between the two studies:

1. All groups of learners in both studies perform high in the converging condition: see the gray bar plot in each pair.

2. Monolingual Russian learners (human as well as simulated) perform above chance in the conflicting condition, although not as high as in the converging condition: see the pair of bar plots on the left.

3. All bilingual learners perform either at chance or below chance in the conflicting condition: see the white bar plots in the two pairs on the right.

To investigate whether these similarities are statistically significant, we fit a logistic regression model to the data, which predicts the odds of making the right choice from three variables used by Janssen et al. (2015): group (Russian monolinguals vs. English bilinguals vs. French bilinguals), stimulus cue condition (converging vs. conflicting), and stimulus case contrast (nominative–accusative vs. nominative–dative), with all the interactions between these variables.[7] The summary is provided in Table 3.

When interpreting the results, it is important to keep in mind three points. First, the reference level in the Table is the Russian monolingual group, conflicting cues and nominative–accusative case contrast. Second, to make the results more interpretable, we report them in terms of the probability of selecting the correct alternative in a pair of instances, $P(I_{correct})$. Finally, only some pairwise comparisons between various factor levels are reported in the Table: to obtain the missing comparisons, we use *lsmeans* package for *R* (Lenth, 2016).

First, there is a significant effect of type, which means that simulated Russian speakers interpret the nominative–accusative stimuli with conflicting cues less accurately than such stimuli with converging cues: $P(I_{correct}) = .84$ vs. .97. Our post-hoc pairwise comparisons confirm that this effect is significant in all the other group–case conditions.

Table 3: Summary of the logistic regression model *(in)correct ~ group × type × case*, fitted to the data from our simulation of Janssen et al.'s (2015) experiment.

Variable	β	SE	P	$P(I_{correct})$[a]
(Intercept)[b]	1.67	0.10	<.001	.84
Group:En	−1.12	0.23	<.001	.63
Group:Fr	−0.91	0.22	<.001	.68
Type:Conv	1.73	0.22	<.001	.97
Case:DAT	−1.12	0.21	<.001	.63
Group:En × Type:Conv	−0.33	0.26	.20	.88
Group:Fr × Type:Conv	−1.37	0.25	<.001	.75
Group:En × Case:DAT	−1.24	0.27	<.001	.14
Group:Fr × Case:DAT	−0.41	0.24	.01	.32
Type:Conv × Case:DAT	−0.86	0.24	<.001	.81
Group:En × Type:Conv × Case:DAT	4.10	0.31	<.001	.94
Group:Fr × Type:Conv × Case:DAT	3.08	0.29	<.001	.86

[a] This variable shows the resulting probability of selecting the correct alternative in a particular condition: for example, the value .88 in the line "Group:En × Type:Conv" means that the English group selects the correct alternative on a test stimulus with converging cues (and nominative–accusative contrast, which is the baseline) with the probability of 88%. Each $P(I_{correct})$ value is computed using an inverse-logit transformation on the value of the respective β-coefficient, and adding it up to the identically transformed baseline probability: intercept for the main effects, main effects for the two-way interactions, and so on.

[b] Intercept corresponds to the probability of choosing the correct alternative by the Russian monolingual group on the stimuli with conflicting cue type and nominative–accusative case contrast.

More importantly, we observe a significant effect of group. The monolinguals perform significantly more accurately than Russian–English and Russian–French bilinguals on the nominative–accusative stimuli with conflicting cues: $P(I_{correct})$ = .84 vs. .63 and .68, respectively. The post-hoc comparisons yield the same effect for all the other types of stimuli, apart from the ones with converging cues and nominative–dative case contrast. Together with the main effect of case reported in the Table, this suggests that the Russian monolinguals could not successfully acquire the nominative–dative cue contrast. This differs

from the human subject results reported by Janssen et al. (2015). However, an analysis of the input data to our model explains this difference: the dative case occurs only 25 times in our Russian data, and not a single time in a noun-verb-noun sentence. Given such input, it is unsurprising that the model could not successfully acquire the nominative–dative cue contrast.

Despite this difference, our main finding in terms of the competition of the two cues, case marking and word order, is compatible with Janssen et al.'s (2015) results: OVS sentences are interpreted less accurately than SVO sentences, and this difference is most evident in bilingual learners. Given the competition of cues in our model, this result supports the explanation provided by the Competition Model. Next, we will investigate whether the results can be explained in terms of CLI.
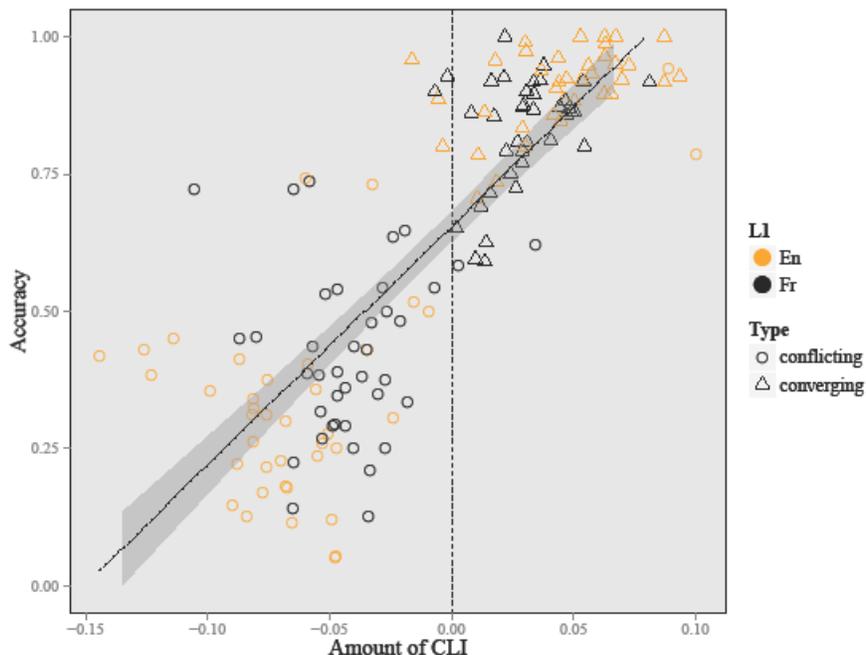
4.1.3. Analysis of CLI

Figure 4: Average accuracy vs. amount of CLI per stimulus in simulation set 1, with a fitted linear regression line.

We use the $\Delta CLI(I)$ measure introduced in section 3.5. Our main prediction concerns the bilinguals' interpretation of the OVS sentences: we expect the negative effect of CLI to prevail over its positive effect. This is why we first zoom in on the conflicting cue condition. The arithmetic mean of $\Delta CLI(I)$ is

negative in this condition for each group of bilinguals: –0.06 for the English group, and –0.05 for the French group. This is different from the converging cue condition, in which the corresponding values of $\Delta CLI(I)$ are positive: 0.04 and 0.03. Although the difference is not large in absolute terms, the signs of the means are opposite, and the Mann–Whitney U test shows that the difference is statistically significant: $U = 2{,}079{,}000$, $p < .001$. The difference between the two types of stimuli is clearly visible in Figure 4: the average accuracy tends to be higher for those stimuli which yield more positive CLI. All together, this supports our prediction that the negative CLI prevails in OVS sentences, leading to their misinterpretation.

To test whether $\Delta CLI(I)$ adds any explanatory power to the regression model reported in the previous section (Table 3), we updated the model by including various interactions between $\Delta CLI(I)$, group, type, and case: the resulting model was *(in)correct ~ $\Delta CLI(I) \times$ group $\times$ type $\times$ case – $\Delta CLI(I) \times$ type – $\Delta CLI(I) \times$ case*. In the model fitted to the data, the coefficients for the predictors and their interactions differed to a certain extent in their absolute values from those in the original model, but these differences were small and did not affect the main results – for brevity we do not report the full model. Most importantly, the amount of CLI had a significant effect on the accuracy of the two bilingual groups on the sentences with conflicting cues, judging by the respective β-coefficients. Also, the comparison between the two regression models, with and without $\Delta CLI(I)$, in terms of the corrected Akaike information criterion (AICc) demonstrated that the model which takes into account the amount of CLI predicted the data better: ΔAICc = 568. This suggests that our ΔCLI measure is able to capture the amount of CLI, as well as its effect on the model's choice in the target task.

To summarize, the results of our simulation were similar to those reported by Janssen et al. (2015), although due to the lack of dative nouns in our input data the model could not successfully acquire the dative–nominative contrast. Taken into account the type of our computational model, this result supports the competition of cues as a plausible explanation for the misinterpretation of OVS sentences. Our analysis of CLI showed that the $\Delta CLI(I)$ measure could serve as an additional independent predictor of the model's accuracy in the target task.

In the next experiment, we simulate a different population of learners, and further investigate the role of CLI in the target task.

4.2. Simulation set 2

In our second set of simulations, we proceed with the experiment of Kempe and MacWhinney (1998). Just as in the previous section, we first test our model by simulating the picture-choice task in the two populations from the target experiment: adult L2 Russian learners and L2 German learners. Second, we investigate whether the impact of CLI on the comprehension of case-marking cues in Russian is manifested in these two populations. Ultimately, this set of simulations will also allow us to make more informed predictions about case-marking comprehension in other language pairs.

We start, however, with an additional data analysis. Kempe and MacWhinney (1998) report that the validity of the case-marking cues in Russian is higher than in German, which makes German case-marking cues more difficult to acquire and comprehend. Following their method (see section 2.3), we calculated the validity of case-marking and word order cues for all the transitive sentences in our data sets. The overall pattern (Table 4) is in line with what Kempe and MacWhinney report for their language samples, although the absolute values differ, probably due to the small number of target sentences in our data set (40 in Russian and 70 in German).

The validity of the case-marking cues, especially the accusative, is lower in German than in Russian – this is why we expect that our model will interpret Russian OVS sentences more successfully than German OVS sentences, just as the human participants in Kempe and MacWhinney's experiment.

Table 4: Availability, reliability, and validity of the case-marking cues in transitive sentences in our data sets.

Cue	German			Russian		
	Availability	Reliability	Validity	Availability	Reliability	Validity
Word order						
(Agent first)	1.00	.97	.97	1.00	.78	.78
VSO	.40	1.00	.40	—	—	—
SVO	.60	.96	.57	.58	1.00	.58
SOV	—	—	—	.43	.47	.20
Case						
(Total)	.80	1.00	.80	1.00	1.00	1.00
Nom.	.77	1.00	.77	.98	1.00	.98
Acc.	.13	1.00	.13	.55	1.00	.55

4.2.1. Simulation details

We annotated the original stimuli available from Kempe and MacWhinney's (1998) study, using the same approach as for our input data sets. Recall that our data sets were obtained from child-directed speech, therefore the L1 input to our model in this experiment may not be as rich as the input that adult speakers are exposed to through the course of their life. Besides, the type of L2 input that adult learners receive differs from child-directed speech. Therefore, we use our data sets as an approximation of the input only, although they are representative in terms of the case marking in Russian and German transitive sentences.

The model was exposed to 600 English instances, followed by 600 instances of mixed input, in which English and Russian (or English and German) were contained in equal proportion. Note that these values are higher than in our previous simulation set, to better approximate adult L2 learning. After that, the model in each condition performed the picture-choice task on the 40 test instances.

4.2.2. Results

Figure 5 provides a visual comparison of our results against Kempe and MacWhinney's (1998) study. Each barplot shows how many times the SVO interpretation was chosen (first-noun-as-subject), normalized by the total number of (simulated) learners; there are seven groups of stimuli in total, depending on the case marking of the first and the second noun in the sentence. The first four groups (NEU-NEU, NOM-NEU, NEU-ACC, and NOM-ACC) represent the SVO pattern, and the other three the OVS pattern. The Figure reveals the following similarities between the original study and our simulation:

1. In SVO sentences (four pairs of bar plots on the left), both Russian and German learners predominantly choose the first noun in the sentence as the agent.

2. In OVS sentences (three pairs of bar plots on the right), Russian learners tend to choose the second noun in the sentence as the agent, while German learners perform close to chance on this type of stimuli.

At the same time, Figure 5 also reveals some differences between the two studies. Most importantly, human participants perform on SVO sentences with fully neutralized case-marking cues (the utmost left pair of bar plots) just as on the other SVO sentences, choosing the first noun as the agent in approximately 90% of cases. In contrast, our model exhibits a less clear preference on this type of stimuli: the proportion of first-noun choice is approximately 70% in each language. We believe it may be either due to the relatively small size of the input data that the model received compared to human speakers, or

due to the model's insufficient attention to the word order cue in isolation; we return to this issue in the discussion.

Another difference relates to the relative accuracy on particular types of Russian OVS sentences. For Kempe and MacWhinney's participants, sentences with the neutralized–nominative case contrast were the most difficult to interpret among the three types of Russian OVS sentences: note that the height of the respective bar plot in Figure 5a is the closest to the chance level. In contrast, our model performed worst on the accusative–neutralized case contrast. We see this difference as an artifact of the particular data sets used in our simulations.

To statistically test the difference in accuracy between the two types of stimuli (OVS vs. SVO sentences) and between the two languages (German vs. Russian), we fit a logistic mixed-effects model to the data, which predicts the odds of making the correct choice from the two mentioned variables and their interaction, with random intercepts over learners and stimuli, and with a random slope of the stimulus type over learners. The model summary is presented in Table 5.

Table 5: Summary of the logistic mixed-effects regression model *(in)correct ~ group × type + (group | simulation) + (1 | sentence)*, fitted to the data from our simulation of Kempe and MacWhinney's (1998) experiment.

Variable	β	SE	p	$P(I_{correct})$
(Intercept)*	−0.19	0.33	.76	.45
Group:Ru	3.04	0.44	<.001	.95
Type:SVO	5.96	1.01	<.001	1.00
Group:Ru × Type:SVO	−3.39	0.97	<.001	1.00

* Intercept corresponds to the probability of choosing the correct alternative by the German monolingual group on the OVS stimuli.
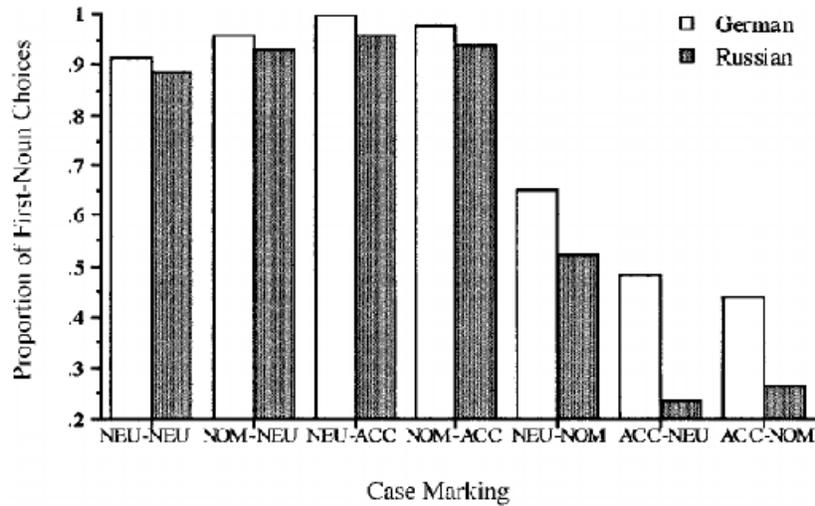
(a) Original results of Kempe and MacWhinney. Reprinted from "The acquisition of case marking by adult learners of Russian and German", 1998, by V. Kempe and B. MacWhinney. *Studies in Second Language Acquisition*, 20, p. 563. Copyright 1998 by Cambridge University Press. Reprinted with permission.
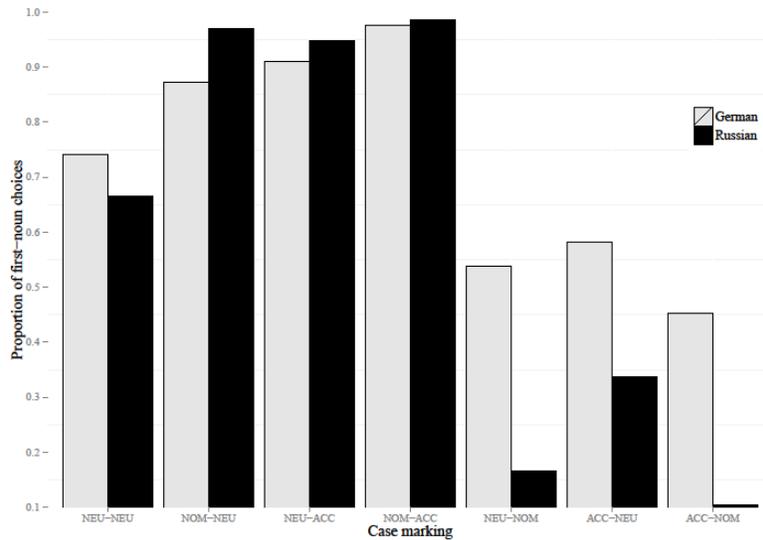
(b) Results of our simulations.

Figure 5: Simulating the experiment of Kempe and MacWhinney (1998).

The results demonstrate a significant effect of language: Russian learners perform significantly more accurately than German learners on the OVS stimuli: $P(I_{correct})$ = .95 vs. .55, while there is no difference on SVO stimuli: $P(I_{correct})$ for both languages is close to 1. There is also a significant effect of sentence type: the performance of the German group on SVO sentences is significantly higher than on OVS sentences: $P(I_{correct})$ = 1.00 vs. .45. Our post-hoc analysis shows that the same effect is significant for Russian learners as well.

Additionally, we compared the performance of Russian and German simulated learners on each of the seven stimulus types shown in Figure 5. A logistic mixed-effects model was fitted to the data on each stimulus type with a fixed effect of language (German vs. Russian), a random intercept and a random slope of language over individual learners, and a random intercept over individual stimuli. The results demonstrated that the difference between Russian and German learners is only statistically significant in the three OVS types (the three pairs of bar plots on the right in Figure 5). These findings are in line with the results of Kempe and MacWhinney (1998) for the accuracy of case-marking comprehension.

To conclude, the results support our prediction about the interpretation of SVO vs. OVS sentences. We proceed with the analysis of CLI in this set of simulations.

4.2.3. Analysis of CLI

Just as in the previous experiment, we investigate whether the choices made by our computational model can be explained in terms of CLI. We fit a regression model similar to the one described in the previous section, which includes ΔCLI and its interactions as additional predictors. The results demonstrate the effect of CLI in OVS sentences: a .1 increase in ΔCLI results in a .04 (German) or .01 (Russian) increase in the probability of making the correct choice for OVS sentences. This is also visualized in Figure 6: we see that the accuracy tends to be higher for the positive values of ΔCLI. The result shows that the CLI measure is highly predictive of the difference between subject groups in the target task.

At the same time, if we focus on OVS sentences and compare the ΔCLI values in Russian vs. German learners, there tends to be no difference: compare the *X*-coordinate of the OVS points (circles) in Figure 6 across the two colors. The Mann–Whitney *U* test also demonstrates no support for the possible difference: U = 1,310,400, p = .24. This suggests that it is not the CLI that explains the difference in the interpretation of OVS sentences by German vs. Russian learners. Instead, this difference must be explained in terms of Russian-to-Russian or German-to-German influence: the higher ambiguity in the

German case system, compared to the Russian system, leads to the observed difference in the model's performance on OVS sentences in Russian vs. German.
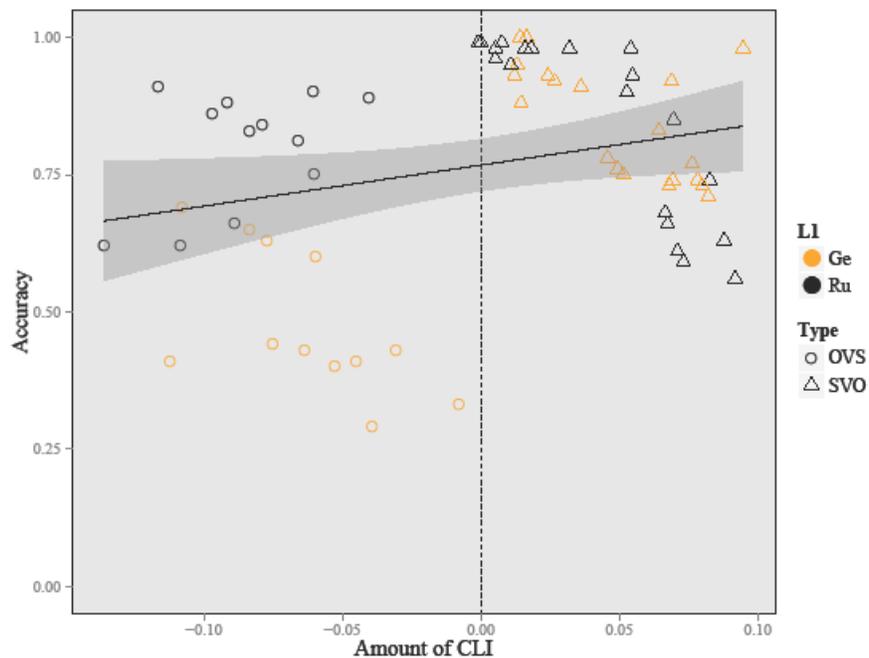

Figure 6: Average accuracy vs. amount of CLI per stimulus in simulation set 2, with a fitted linear regression line.

To summarize, in this set of simulations we demonstrated that our model produces results similar to the human data, when interpreting case-marking cues in Russian and German SVO and OVS sentences. The main discrepancy between our results and human subject results was observed in the interpretation of sentences with fully neutralized case-marking cues. As for the effect of CLI, it was manifested in this set of simulations just as in the previous one, but the amount of CLI could not explain the difference between the accuracy of Russian and German learners. Instead, we attribute this difference to the validity of case-marking cues, suggesting that our computational model is compatible with the Competition Model framework. In the next section we demonstrate how novel predictions can be made based on the outcomes of our two sets of simulations.

4.3. Novel simulations

We can now go beyond the replication setup and make predictions about the interpretation of case-marking cues in other bilingual populations. We make two specific predictions regarding the behavior of

our model and run two additional sets of simulations to test these, followed by an analysis of the results in terms of CLI.

1. Janssen et al. (2015) in their study explain that their result, in particular the low accuracy on the conflicting sentences in bilinguals, may be "due to bilingualism in itself, ... or to the fact that the other language provided no support for case cues" (p. 276). At the same time, the presence or absence of case cues in the other language has been shown to be important: for example, L1 Italian speakers interpret L2 Spanish OVS sentences better than L1 English speakers (Isabelli, 2008). Similarly, we hypothesize that the knowledge of German with its rich case marking can be beneficial for the acquisition of Russian cases, and that Russian–German bilingual children would interpret Russian sentences more accurately than Russian–English or Russian–French bilinguals.

2. Kempe and MacWhinney (1998) demonstrate that case-marking cues are more difficult to acquire in German than in Russian, and our simulation set 2 validates this result applied to our model. It is therefore reasonable to hypothesize that monolingual German children would perform less accurately on OVS sentences in the picture-choice task when tested on German, compared to Russian monolingual children tested on Russian (i.e., as in the experiment of Janssen et al. and our first set of simulations). At the same time, bilingual French–German and English–German children are expected to perform poorly on OVS sentences, while Russian–German children may benefit from their knowledge of Russian and achieve higher accuracy compared to the two other groups of bilinguals.

Using our computational model, we run two additional simulations to pre-test these hypotheses.

4.3.1. Bilingual Russian–German children

Following the setup described in section 4.1.1, we simulate an additional group within the same experiment: Russian–German bilinguals. This population of simulated learners is tested on the same Russian stimuli as the other three groups (Russian monolinguals, Russian–English and Russian–French bilinguals). The comprehension accuracy for the new group (utmost right plots) against the other groups is shown in Figure 7. It suggests that Russian–German bilingual learners have an advantage in this task over Russian–English and Russian–French learners. This prediction must be tested experimentally with human participants. At the same time, to understand whether this effect in our model is caused by positive CLI from Russian to German, we believe it is useful to look in more detail at the simulated data.
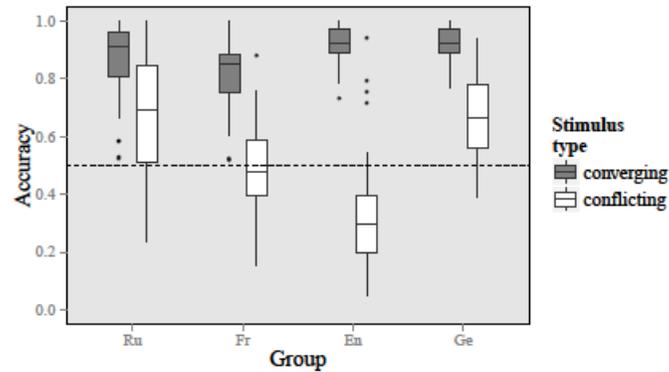
Figure 7: Accuracy of simulated Russian–German bilinguals against the other groups.

First, we statistically test the pairwise differences in accuracy between the Russian–German group and the other groups across the types of our stimuli: a logistic regression model *(in)correct ~ group × type × case* is fitted to the data for all the five groups, just as in simulation set 1 (section 4.1.2), and then all the pairwise contrasts between the Russian–German group vs. each of the other groups are analyzed using *lsmeans* package. The summary of the contrasts is provided in Table 6.

We are mostly interested in the difference between Russian–German vs. the other two groups of bilinguals. Table 6 suggests that the former perform substantially better than the other two groups on dative-verb-nominative sentences, and than the Russian–French group (but not Russian–English) on nominative-verb-accusative and nominative-verb-dative sentences. Because there are no significant differences for the other types of sentences, our hypothesis about the facilitatory effect of the German knowledge is supported only partially.

To investigate the contribution of CLI to this result, we can compare the differences in accuracy to the differences in the amount of CLI across different groups of learners. A linear mixed-effects model has been fitted to the data, predicting the amount of CLI (ΔCLI), and the pairwise contrasts were computed (see Table 6 on the right). We can see that the greatest difference in the amount of CLI is observed for dative-verb-nominative sentences ($\Delta LSM = -0.10$ and -0.06), which is the only type of stimuli on which the Russian–German group scores higher in accuracy than both other bilingual groups. This means that the amount of positive CLI for this type of stimuli is higher in Russian–German group than in Russian–French and Russian–English group, in line with our prediction.

Table 6: Summary of pairwise linear contrasts for accuracy and CLI using least-square means (LSM).

Contrast[a]	Type	Case	Accuracy			CLI		
			ΔLSM	SE	p[b]	ΔLSM	SE	p
Ru–Ge[c]	conflicting	Acc	0.83	0.12	<.001	0.02	0.00	<.01
En–Ge	conflicting	Acc	–0.29	0.11	.07	–0.03	0.00	<.001
Fr–Ge	conflicting	Acc	–0.08	0.11	1.00	–0.02	0.00	<.001
Ru–Ge	converging	Acc	1.28	0.23	<.001	0.00	0.00	1.00
En–Ge	converging	Acc	–0.17	0.16	.96	0.01	0.00	.86
Fr–Ge	converging	Acc	–1.01	0.14	<.001	0.01	0.00	.94
Ru–Ge	conflicting	Dat	0.10	0.08	.91	–0.02	0.00	<.001
En–Ge	conflicting	Dat	–2.25	0.10	<.001	–0.10	0.00	<.001
Fr–Ge	conflicting	Dat	–1.22	0.08	<.001	–0.06	0.00	<.001
Ru–Ge	converging	Dat	–1.36	0.15	<.001	–0.05	0.00	<.001
En–Ge	converging	Dat	0.05	0.18	1.00	0.01	0.00	.01
Fr–Ge	converging	Dat	–0.98	0.16	<.001	–0.02	0.00	<.001

[a] German-Russian bilingual learners are at the second position in each contrast: negative Δ values are associated with the higher estimate of the respective coefficient in the Russian–German group.

[b] The p-values are adjusted for multiple comparisons (so-called multivariate t-probabilities) using the *mvt* method (Lenth, 2016).

[c] "Ru" is the monolingual group, while "Ge", "En" and "Fr" denote the respective bilingual groups.

As for the other types of stimuli, the amount of CLI in German is simply not high enough to facilitate the interpretation of Russian sentences: this can be demonstrated by plotting the average amount of CLI across the three groups of bilinguals, see Figure 8. Note that in the conflicting cue condition, the amount of CLI is always higher in the German group than in the other two groups. However, its absolute value is positive only for the dative-verb-nominative sentences, but not for accusative-verb-nominative. This explains why we observe no differences across the bilingual groups for this latter type of sentences.
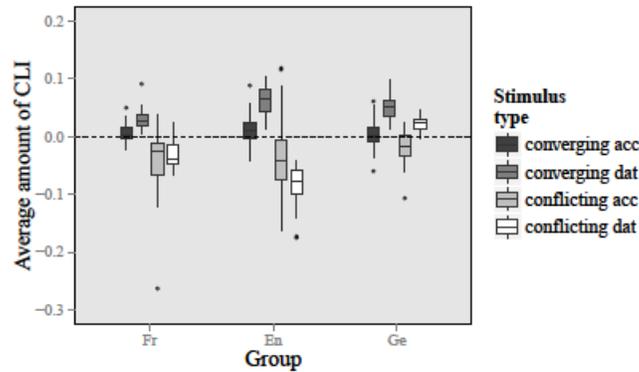
Figure 8: Amount of CLI (ΔCLI) per group, averaged over stimuli, in the Russian picture-choice task.

To summarize, our simulated data only partially confirms our prediction that the knowledge of German can facilitate the interpretation of Russian case marking. The lack of the hypothesized effect can be explained by the amount of CLI across different types of stimuli.

4.3.2. Bilingual and monolingual German children

Until this point, we have simulated the experiment of Janssen et al. (2015) using Russian sentences. Our final prediction, however, concerns case comprehension in German. We use the same setup as in simulation set 1, but this time simulating German monolingual children and three groups of bilinguals: German–French, German–English, and German–Russian. All the four groups are tested on German instances. Ideally, we would translate Janssen et al.'s stimuli into German, however many of such translated sentences would be fully neutralized in terms of their case-marking cues. To give an example, the sentence *Кукла любит жирафа* 'The doll loves the giraffe' would translate into German as *Die Puppe liebt die Giraffe*, where the case of both nouns can be interpreted as either accusative or nominative. Therefore, in this experiment we used a subset of Kempe and MacWhinney's German stimuli, 24 out of 32: the 8 fully neutralized stimuli were eliminated.
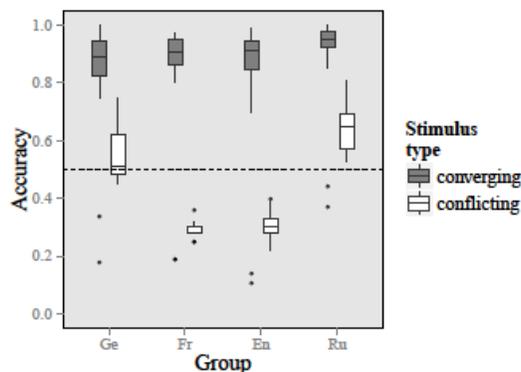
Figure 9: Accuracy per group in the German picture-choice task.

The performance of the four groups is shown in Figure 9 (the significance was tested with a logistic mixed-effects model fitted to the data)[8]. First of all, the results show that German monolinguals (the utmost left pair of plots) perform well in the converging cue condition, but close to chance in the conflicting cue condition. This is in line with the findings of Kempe and MacWhinney (1998) for L2 learners (which we simulated in experiment 2), and also with the existing data suggesting that German children are only able to interpret case marking in OVS sentences around the age of seven (Dittmar et al., 2008). Besides, this result is clearly different from the accuracy of simulated Russian monolinguals in our simulation set 1, who performed well in the conflicting cue condition. This supports our prediction about the accuracy of monolinguals on German vs. Russian OVS sentences.

As for the bilingual groups, both German–English and German–French bilinguals perform in the conflicting cue condition less accurately than monolinguals. Interestingly, our simulated German–Russian bilinguals perform significantly better than German monolinguals in this condition. While this is inconsistent with the view that bilinguals lag behind monolinguals in their language development (Hulk, 2004; Schmitz, 2006), bilingual children have been shown to acquire some grammatical features earlier than monolinguals (Meisel, 1986; Pléh, Jarovinskij, & Balajan, 1987). In our model, this may only happen if bilinguals benefit from positive CLI. To investigate this, we analyze the ΔCLI values for all the groups, focusing on the conflicting condition. The comparison is provided in Figure 10. Note the obvious difference across the groups in the conflicting condition, but not in the converging condition. The direction of this difference is as expected: the effect of CLI is positive in German–Russian bilinguals and negative in the other two groups.
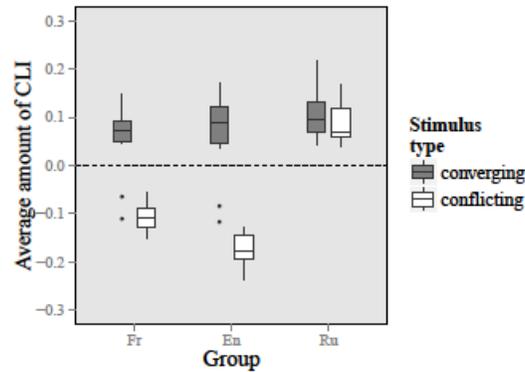
Figure 10: Amount of CLI (ΔCLI) per group, averaged over stimuli, in the German picture-choice task.

To summarize, the results of this simulation set support our prediction: the accuracy of simulated monolinguals on the sentences with conflicting cues is lower in German than in Russian. This is in line with Kempe and MacWhinney's (1998) findings and with the results of our simulation set 2, corroborating the idea in the Competition Model that the validity of case cues is important. The results also confirm our hypothesis about the performance of the bilingual groups on the German OVS sentences: while the simulated German–French and German–English bilinguals performed poorly on such sentences, German–Russian bilinguals benefited from positive CLI and achieved high accuracy.

5. Discussion

Our goal in this article was to demonstrate how the amount of CLI can be measured in a computational model, and how such a measure can be applied for explaining a particular phenomenon, the comprehension of case-marking cues.

5.1. Quantifying the effect of CLI

We introduced a measure of CLI and used it to quantify the CLI effect in the picture-choice task. This measure helped us to determine the contribution of CLI to the observed result. As it was demonstrated, the measure can be used both on the level of a particular group or condition (e.g., average amount of CLI in the interpretation of OVS sentences by Russian–German bilinguals), and on the level of a particular test item (cf. Figure 4), in case the goal is to study the differences between individual sentences.

In this study, we only employed a particular linguistic task related to sentence interpretation, but it is perfectly possible to use this computational model for simulating other linguistic tasks: filling in verbs or

prepositions, verb definition, verb selection, and others (see Matusevych et al., 2016b, 2017). For all these tasks, the same type of measure (*CLI* or *ΔCLI*) can be used to quantify the amount of CLI and shed light on its role in language comprehension and production.

As we demonstrated in this study, this measure can help us explain the observed linguistic behavior of human participants, as well as make informed predictions about such behavior in groups that have not been studied yet (our novel simulations). A potential drawback of this measure is that it may not be possible at the moment to verify it with human participants. As we argue in section 1.1, the existing measures of CLI in human speakers are based on their language use, and not on the actual linguistic representations, as in the present study. At the same time, the development of neuroimaging tools measuring brain activation in bilingual and L2 speakers (e.g., Higby et al., 2016; Jeong, Sugiura, & Sassa, 2007; Yokoyama, Kim, Uchida, Miyamoto, Yoshimoto, & Kawashima, 2013) may allow for a proper evaluation of our CLI measure in the future.

5.2. CLI and "agent-first" in case-marking cue comprehension

Speaking about the linguistic phenomenon of interest – case-marking cue comprehension – it was demonstrated in this study that our probabilistic computational model performed in the target task similar to human learners: early bilinguals in the experiment of Janssen et al. (2015) and L2 learners in the experiment of Kempe and MacWhinney (1998). In section 1.2 we outlined two general views on the role of CLI in case-marking cue comprehension by bilingual or L2 learners. Our results suggest that CLI is an important factor that affects this kind of comprehension: first, an effect of the amount of CLI was observed in all our simulations; second, the performance of simulated Russian–German speakers on Russian and German OVS sentences (recall our novel sets of simulations) was higher compared to bilinguals whose other language had no case-marking cues (i.e., French or English).

At the same time, our results do not rule out the existence of an independent "agent-first" cognitive strategy, not driven by language input, in line with VanPatten's First-Noun Principle. There is no agreement in the literature whether such a strategy is plausible. Yasunaga, Yano, Yasugi, and Koizumi (2015) provide an overview of the issue and argue against the universal strategy. On the other hand, studies of non-verbal behavior, in which speakers of different languages use the same order of gestures to describe an event, provide one type of indirect evidence that an agent-first strategy may exist (Goldin-Meadow, So, Özyürek, & Mylander, 2008). Another type of evidence is the fact that the subject precedes the object in the majority of the world's languages (e.g., the SVO word order is more common than OVS, and SOV is more common than OSV). Our study does not provide conclusive evidence on the issue, but it

may be the case that the mismatch between the performance of our model and Kempe and MacWhinney's participants on fully neutralized sentences (recall section 4.2.2) comes from the fact that our model does not employ an independent "agent-first" strategy, while human speakers may do so. At the same time, there are alternative explanations of the mentioned mismatch, which we discuss in section 5.5.

Finally, recall from section 1.2 that there are at least two theories promoting the role of CLI: the Competition Model and the L1 Transfer Principle. In our model, various features compete with each other, which makes it compatible with the Competition Model. This similarity is also supported by the data of our simulation set 2: while the amount of CLI could not explain the difference in the performance of German–English vs. Russian–English learners on OVS sentences, our analysis of the cue validity in the German and Russian input data suggested that the model was sensitive to the cue validity, at least for the case-marking cues. This being said, our study supports MacWhinney's Competition Model as the explanation of the misinterpretation of OVS sentences. At the same time, the results do not necessarily challenge VanPatten's L1 Transfer Principle. To our knowledge, the cognitive mechanisms behind this principle have not been described in detail, and this is why it is challenging to verify of falsify this principle.

5.3. CLI in argument structure constructions

The present study also sheds some light on how CLI may occur at the representational level. Given the learning mechanism implemented in our computational model, as well as the type of CLI measure used, there are two ways for the CLI measure to obtain higher values. First, the learner may have the two languages separated in the existing constructions: that is, some constructions are based on L1 only, and others on L2 only. In order for the CLI value to be high, in this case the similarity between a test instance in the target L2 language and some of the existing L1 (non-target language) constructions must be rather high. This is possible, but improbable.

A more likely alternative explanation is that some constructions are blended, as it was shown in Figure 2. There may be substantial variability across simulations in this respect: the number of blended clusters depends on the similarity between L1 and L2 data sets used for training the model, but also – to some extent – on the exact sample of the training data in each simulation and the order of presentation of instances. Given a test instance in L2, the model sometimes makes its choice based on such blended constructions, hence the high CLI value. This supports the view that constructional representations may be shared across languages (Bernolet, Hartsuiker, & Pickering, 2013; Higby et al., 2016; Salamoura & Williams, 2007).

5.4. Limitations of the study

It is important to keep in mind that in the present study we did not consider all the cues that affect sentence comprehension. In particular, we did not take into account the pragmatics of the utterance, expressed in the intonational cues: such cues have been shown to be highly informative for the interpretation of participant roles by monolingual German children (Grünloh, Lieven, & Tomasello, 2011).

Another factor that was left out of our simulations is the lexical similarity between languages. In the study of Isabelli (2008), L1 Italian speakers could interpret L2 Spanish OVS sentences more accurately compared to L1 English speakers, and positive CLI was suggested to be responsible for this. However, as VanPatten (2015a) noticed, this may also be due to the better familiarity of L1 Italian speakers with L2 Spanish personal pronouns used in the experiment. This is why lexical similarity, including the potential effect of cognates, must be taken into account. In our computational model this effect could be captured only for the words spelled identically in the two target languages, such as *giraffe–Giraffe* in English and German. At the same time, there are more cognates between English/French and German, compared to Russian – the results in our simulations (in particular, simulation set 2) may have differed to a certain extent, had the effect of cognates been taken into account.

Another limitation of the present study is that the language pairs in our simulation 1 were not the same as in Janssen et al.'s study. The model may behave differently on different language pairs: for example, the model trained and tested on Russian data performed differently in our simulations than the model trained and tested on German data (novel simulation 1 vs. novel simulation 2). Following Kempe and MacWhinney's theory, we explain this by the different strengths of the case-marking cue characterizing German and Russian. This explanation, however, does not hold for English, French, Hebrew, or Dutch: these languages do not use case marking on nouns (with few exceptions). This is why there might be no reason to assume that the model trained on Russian–English or Russian–French data would perform differently from the model trained on Russian–Dutch or Russian–Hebrew data. This is also indirectly supported by the following facts: first, Janssen et al. (2015) in their study did not observe differences between the Russian–Hebrew group and the Russian–Dutch group; and second, in our simulations, we did not observe qualitative differences between the Russian–English group and the Russian–French group.

Finally, the use of computational modeling has its own merits and flaws, and we discuss this in the next section.

5.5. Computational modeling: methodological implications

The computational model used in the present study leaves out many essential details characterizing the cognitive process of human language acquisition. Most prominently, it only focuses on the bottom-up mechanism of statistical (implicit) acquisition from input data. Other cognitive mechanisms are not accounted for in this model. To give an example, no distinction is drawn between early and late L2 learners in terms of the learning mechanisms involved in the learning process. This is why it is important to keep in mind that computational modeling is only a tool that helps us explain the existing data and/or make informed predictions, which should be tested with human participants in the future.

With this in mind, we can discuss the particular discrepancy between the results of our simulation 2 vs. Kempe and MacWhinney's experimental data: recall that our model did not perform similarly to human participants on fully neutralized sentences. There is something that prevents the model from choosing the first noun as the agent in such sentences as frequently as humans do. To investigate this, we have additionally looked at the performance of the model on individual neutralized sentences, as well as in individual simulations. The variation between individual sentences was substantially smaller than between simulations: in some simulations it was always the second noun that was selected as the agent. We believe this could happen in case there were very few or even only one construction similar to the neutralized sentences, and for each sentence the model made its decision mostly based on the evidence from this construction. If agent-second sentences prevailed in this construction, the model would choose the second noun as the agent. Training the model on a larger corpus could help the model to build richer linguistic representations, so that it would make the decision for each test sentence based on a greater number of constructions, preventing the situation described above. Another solution could be to adjust the feature weights in the model. In the current setup, all features in the input (i.e., word order, head verb, prepositions, etc.) equally affect the model's decision; the only mechanism allowing for setting the weights differently is the parameter w, which helps to distinguish between distributional and symbolic features, as it is explained in section 3.3.2. However, it may be the case that some features (e.g., word order) are always more important for human speakers than others. If we had available data on the relative importance of the features, we could set the respective weights for individual features, for example forcing the model to prioritize the evidence coming from the word order.

Another issue worth bringing up here again is that animacy was not an independent feature in the model. While it may contrast with the approach taken in Kempe and MacWhinney's study, the set of features representing the input data for our model (recall Table 1) was largely adopted from the previous studies

with this model, and setting the animacy as an independent feature would be a questionable solution from this perspective. The role of the animacy cue in this computational model must be additionally investigated in the future.

5.6. Modeling bilingual populations

The model's focus on the mechanism of statistical learning, mentioned in the previous section, affects the way bilingual populations are simulated in this study. Here, the only difference between early and late simulated learners was the amount of their exposure to L1 prior to the L2 onset. Clearly, in reality the differences between early and late L2 learners are much more substantial and may relate to the cognitive mechanisms involved in the learning process, to the role of CLI, to the learning style (e.g., type of instruction), motivations, and so on. On the one hand, the fact that all these variables are eliminated from our computational model makes it a rather simplistic learner. On the other hand, once we adopt such a simplistic view on bilingualism promoted by this model, we can simulate various groups of statistical learners – monolinguals, early and late bilinguals, various L2 learners – and see whether the mechanism of statistical learning alone can capture certain patterns reported in the existing studies with human participants. This approach also makes us leave aside the issue of directionality of CLI: in this model, L1 and L2 can equally affect each other; and as long as the time of onset of the two languages is the same, it does not matter which language is considered L1 and which L2. At the same time, we completely agree that the mentioned issues are important for the study of bilingualism and SLA, and an ideal computational model would need to account for all of them.

6. Conclusion

In this study we used an existing computational model of bilingual argument structure acquisition and demonstrated how the amount of CLI can be measured in this model. The proposed measure was employed to investigate the interpretation of case-marking cues in Russian and German noun–verb–noun sentences. Based on the experimental studies of Janssen et al. (2015) and Kempe and MacWhinney (1998), in the two sets of simulations we demonstrated that our model yielded the types of linguistic behavior similar to humans in the picture-choice task. We then demonstrated that the interpretation of the target sentences in our simulations was more accurate when the amount of CLI was positively large. The findings were used to make two novel predictions not yet tested with human participants. One of them related to the performance of an additional group of Russian–German speakers in the target task with Russian test sentences, while the other one concerned the performance of various groups of learners on German test sentences. The predictions were largely supported by the novel simulations, which suggests it

is worth carrying out the respective experiments with human participants in the future, to see whether the predicted effects are borne out. The findings of the study are discussed in the context of various theories of sentence interpretation, such as the First-Noun Principle and the Competition Model. Our findings support the theoretical framework promoted in the Competition Model, although a cognitive strategy similar to the First-Noun Principle might still be at play in the task of sentence interpretation.

Acknowledgments

References

Alishahi, A., & Stevenson, S. (2008). A computational model of early argument structure acquisition. *Cognitive Science, 32,* 789–834. doi: 10.1080/03640210801929287

Alishahi, A., & Stevenson, S. (2010). A computational model of learning semantic roles from child-directed language. *Language and Cognitive Processes, 25,* 50–93. doi: 10.1080/01690960902840279

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429. doi:10.1037/0033-295x.98.3.409

Bar-Shalom, E., & Snyder, W. (1996). Optional infinitives in Russian and their implications for the pro-drop debate. In M. Lindseth & S. Franks (Eds.), *Formal Approaches to Slavic Linguistics: The Indiana Meeting* (pp. 38–47). Ann Arbor, MI: Michigan Slavic Publications.

Behrens, H. (2006). The input-output relationship in first language acquisition. *Language and Cognitive Processes, 21,* 2–24. doi: 10.1080/01690960400001721

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2013). From language-specific to shared syntactic representations: The influence of second language proficiency on syntactic sharing in bilinguals. *Cognition, 127,* 287–306. doi: 10.1016/j.cognition.2013.02.005

Born, R. (1985). Error types and negative transfer in compositions of third, fourth, and fifth semester German students. *Die Unterrichtspraxis/Teaching German, 18,* 246–253. doi: 10.2307/3530457

Cappelle, B. (2006). Particle placement and the case for "allostructions". *Constructions, Special Volume 1.* Retrieved from http://journals.linguisticsociety.org/elanguage/constructions/article/view/22.html

De Angelis, G., & Selinker, L. (2001). Interlanguage transfer and competing linguistic systems in the multilingual mind. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Crosslinguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives* (pp. 42–58). Clevedon: Multilingual Matters.

Dittmar, M., Abbot-Smith, K., Lieven, E. V. M., & Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Development, 79,* 1152–1167. doi: 10.1111/j.1467-8624.2008.01181.x

Gass, S. M. (1987). The resolution of conflicts among competing systems: A bidirectional perspective. *Applied Psycholinguistics, 8,* 329–350. doi: 10.1017/S0142716400000369

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: University of Chicago Press.

Goldberg, A. E., Casenhiser, D. M., and Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics, 15,* 289–316. doi: 10.1515/cogl.2004.011

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences, 105,* 9163–9168. doi: 10.1073/pnas.0710060105

Grauberg, W. (1971). An error analysis in German of first-year university students. In G. Perren & J. L. M. Trim (Eds.), *Applications of Linguistics* (pp. 257–263). London: The University Press.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1,* 131–149. doi: 10.1017/S136672899800025X

Grünloh, T., Lieven, E. V. M., & Tomasello, M. (2011). German children use prosody to identify participant roles in transitive sentences. *Cognitive Linguistics, 22,* 393–419. doi: 10.1515/cogl.2011.015

Hanson, S., Aroline, E., & Carlson, M. T. (2014). The roles of first language and proficiency in L2 processing of Spanish clitics: Global effects. *Language Learning, 64,* 310–342. doi: 10.1111/lang.12050

Higby, E., Vargas, I., Pérez, S., Ramirez, W., Varela, E., Campoverde, G., …, Obler, L. K. (2016). The bilingual's mental grammar system: Language-specific syntax is shared by both languages. Poster presented at Cognitive Neuroscience Society Annual Meeting, April 4, New York, NY.

Hulk, A. (2004). The acquisition of the French DP in a bilingual context. In P. Prévost & J. Paradis (Eds.), *Language Acquisition and Language Disorders* (pp. 243–274). Amsterdam: John Benjamins Publishing Company.

Isabelli, C. A. (2008). First Noun Principle or L1 Transfer Principle in SLA? *Hispania, 91,* 465–478.

Janssen, B. E. (2016). *The Acquisition of Gender and Case in Polish and Russian: A Study of Monolingual and Bilingual Children*. Amsterdam: Uitgeverij Pegasus.

Janssen, B. E., Meir, N., Baker, A., & Armon-Lotem, S. (2015). On-line comprehension of Russian case cues in monolingual Russian and bilingual Russian-Dutch and Russian-Hebrew children. In E. Grillo & K. Jepson (Eds.), *Proceedings of the 39th Annual Boston University Conference on Language Development* (pp. 266–278). Somerville, MA: Cascadilla Press.

Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in them interlanguage lexicon. *Language Learning, 50,* 245–309. doi: 10.1111/0023-8333.00118

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*. New York, NY: Routledge. doi: 10.4324/9780203935927

Jeong, H., Sugiura, M., & Sassa, Y. (2007). Cross-linguistic influence on brain activation during second language processing: An fMRI study. *Bilingualism: Language and Cognition, 10,* 175–187. doi: 10.1017/S1366728907002921

Jordens, P. (1977). Rules, grammatical intuitions and strategies in foreign language learning. *Interlanguage Studies Bulletin, 2,* 5–76.

Kempe, V., & MacWhinney, B. (1998). The acquisition of case marking by adult learners of Russian and German. *Studies in Second Language Acquisition, 20,* 543–587. doi: 10.1017/s0272263198004045

Kilborn, K., & Cooreman, A. (1987). Sentence interpretation strategies in adult Dutch–English bilinguals. *Applied Psycholinguistics, 8,* 415–431. doi: 10.1017/S0142716400000394

Kim, S., O'Grady, W., & Cho, S. (1995). The acquisition of case and word order in Korean: A note on the role of context. *Language Research, 31,* 687–695.

Lee, J. F., & Malovrh, P. A. (2009). Linguistic and non-linguistic factors affecting OVS processing of accusative and dative case pronouns by advanced L2 learners of Spanish. In J. Collentine, M. García, B.

Lafford, & F. Marcos-Marín (Eds.), *Selected Proceedings of the 11th Hispanic Linguistics Symposium* (pp. 105–116). Somerville, MA: Cascadilla Proceedings Project.

Lenth, R. V. (2016). Using **lsmeans**. Retrieved from https://cran.rproject.org/web/packages/lsmeans/vignettes/using-lsmeans.pdf

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (2012). The logic of the unified model. In S. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 211–227). London: Routledge. doi: 10.4324/9780203808184.ch13

Matusevych, Y., Alishahi, A., & Backus, A. (2017). The impact of first and second language exposure on learning second language constructions. *Bilingualism: Language and Cognition, 20,* 128–149. doi: 10.1017/S1366728915000607

Matusevych, Y., Alishahi, A., & Backus, A. (2016a). *Learning Constructions from Bilingual Exposure: Computational Studies of Argument Structure Acquisition*. Utrecht: LOT.

Matusevych, Y., Alishahi, A., & Backus, A. (2016b). Modelling verb selection within argument structure constructions. *Language, Cognition and Neuroscience,* 31, 1215–1244. doi: 10.1080/23273798.2016.1200732

McDonald, J. L. (1987). Sentence interpretation in bilingual speakers of English and Dutch. *Applied Psycholinguistics, 8,* 379–413. doi: 10.1017/S0142716400000382

Meisel, J. M. (1986). Word order and case marking in early child language. Evidence from simultaneous acquisition of two first languages: French and German. *Linguistics,* 24, 123–184. doi: 10.1515/ling.1986.24.1.123

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM, 38,* 39–41. doi: 10.1145/219717.219748

Miller, M. (1979). *The Logic of Language Development in Early Childhood.* Berlin: Springer-Verlag.

Mimica, I., Sullivan, M., & Smith, S. (1994). An on-line study of sentence interpretation in native Croatian speakers. *Applied Psycholinguistics, 15,* 237–261. doi: 10.1017/S0142716400005348

Morett, L. M., & MacWhinney, B. (2013). Syntactic transfer in English-speaking Spanish learners. *Bilingualism: Language and Cognition, 16,* 132–151. doi: 10.1017/S1366728912000107

Morgenstern, A., & Parisse, C. (2012). The Paris Corpus. *Journal of French Language Studies, 22,* 7–12. doi: 10.1017/S095926951100055X

Morgenstern, A., Parisse, C., Sekali, M., Bourdoux, F., & Caet, S. (2004). French Paris Corpus [Electronic database]. Retrieved from http://childes.psy.cmu.edu/data/Romance/French-MOR/Paris.zip

O'Shannessy, C. (2011). Competition between word order and case-marking in interpreting grammatical relations: A case study in multilingual acquisition. *Journal of Child Language, 38,* 763–792. doi: 10.1017/s0305000910000358

Palmberg, R. (1976). A select bibliography of error analysis and related topics. *Interlanguage Studies Bulletin, 1,* 340–389.

Perek, F. (2015). *Argument Structure in Usage-based Construction Grammar: Experimental and Corpus-based Perspectives*. Amsterdam: John Benjamins Publishing Company. doi: 10.1075/cal.17

Pléh, C., Jarovinskij, A., & Balajan, A. (1987). Sentence comprehension in Hungarian-Russian bilingual and monolingual preschool children. *Journal of Child Language, 14,* 587–603. doi: 10.1017/s0305000900010308

Protassova, E. (2004). Russian Protassova Corpus [Electronic database]. Retrieved from http://childes.psy.cmu.edu/data/Slavic/Russian/Protassova.zip

Salamoura, A., & Williams, J. N. (2007). Processing verb argument structure across languages: Evidence for shared representations in the bilingual lexicon. *Applied Psycholinguistics, 28,* 627–660. doi: 10.1017/s0142716407070348

Schaner-Wolles, C. (1989). Strategies in acquiring grammatical relations in German: Word order or case marking. *Folia Linguistica, 23,* 131–156.

Schmid, H.-J. (2016). A framework for understanding linguistic entrenchment and its psychological foundations. In H.-J. Schmid (Ed.), *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge* (pp. 9–36). Washington, DC: De Gruyter Mouton and American Psychological Association. doi: 10.1515/9783110341423-002

Schmitz, K. (2006). Indirect objects and dative case in monolingual German and bilingual German/Romance language acquisition. In D. Hole, A. Meinunger, & W. Abraham (Eds.), *Datives and Other Cases: Between Argument Structure and Event Structure* (pp. 239–268). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/slcs.75.11sch

Shook, A., & Marian, V. (2013). The bilingual language interaction network for comprehension of speech. *Bilingualism: Language and Cognition, 16,* 304–324. doi: 10.1017/S1366728912000466

Smolík, F. (2015). Word order and information structure in Czech 3-and 4-year-olds' comprehension. *First Language, 35,* 237–253. doi: 10.1177/0142723715596098

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *Journal of Child Language, 28,* 127–152. doi: 10.1017/s0305000900004608

VanPatten, B. (1996). *Input Processing and Grammar Instruction in Second Language Acquisition.* Norwood, NJ: Ablex Publishing Corporation.

VanPatten, B. (2012). Input processing. In S. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 268–281). London: Routledge. doi: 10.4324/9780203808184.ch16

VanPatten, B. (2015a). Foundations of processing instruction. *International Review of Applied Linguistics in Language Teaching, 53,* 91–109. doi: 10.1515/iral-2015-0005

VanPatten, B. (2015b). Input processing in adult SLA. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (2nd ed., pp. 113–134). New York, NY: Routledge.

von Stutterheim, C. (2004). German Caroline Corpus [Electronic database]. Retrieved from http://childes.psy.cmu.edu/data-xml/Germanic/German/Caroline.zip

Weinreich, U. (1968). *Languages in Contact: Findings and Problems* (6th ed.). The Hague: Mouton Publishers.

Yasunaga, D., Yano, M., Yasugi, Y., & Koizumi, M. (2015). Is the subject-before-object preference universal? An event-related potential study in the Kaqchikel Mayan language. *Language, Cognition and Neuroscience, 30,* 1209–1229. doi: 10.1080/23273798.2015.1080372

Yokoyama, S., Kim, J., Uchida, S., Miyamoto, T., Yoshimoto, K., & Kawashima, R. (2013). Cross-linguistic influence of first language writing systems on brain responses to second language word reading in late bilinguals. *Brain and Behavior, 3,* 525–531. doi: 10.1002/brb3.153

Yoshimura, Y., & MacWhinney, B. (2010). The use of pronominal case in English sentence interpretation. *Applied Psycholinguistics, 31,* 619–633. doi: 10.1017/s0142716410000160

Zhao, X., & Li, P. (2010). Bilingual lexical interactions in an unsupervised neural network model. *International Journal of Bilingual Education and Bilingualism, 13,* 505–524. doi: 10.1080/13670050.2010.488284

---

[1] We adopt a broad cognitive view on CLI (Jarvis & Pavlenko, 2008), which covers manifestations of CLI both in L2 acquisition and in bilingual language use.

[2] Note, however, that the lexical similarity between Italian and Spanish might be a factor in this example (VanPatten, 2015a) – we return to this issue in the discussion.

[3] This does not necessarily contradict Green's (1998) Inhibitory Control Model, in which each language instance carries a tag: L1, L2, and so on. Green's model deals with lemmas, and in our study most lemmas occur only in one input language – that is, each lemma implicitly carries information about its language. Therefore, the model's behavior would barely change if we would append a language tag to each lemma (e.g., *rooster-L1* instead of *rooster*).

[4] English Manchester corpus (Theakston, Lieven, Pine, & Rowland, 2001); German corpora of Caroline (von Stutterheim, 2004), Kerstin (M. Miller, 1979), and Leo (Behrens, 2006); Russian Protassova corpus (Protassova, 2004) and Tanya corpus (Bar-Shalom & Snyder, 1996); French Paris corpus (Morgenstern & Parisse, 2012; Morgenstern, Parisse, Sekali, Bourdoux, & Caet, 2004).

[5] For simplicity, an English translation of the Russian sentence is used.

[6] The value of this parameter is set empirically, together with the value of the smoothing parameter λ. In all the simulations presented here, we use $\lambda = 10^{-14}$ and $w = 0.2$.

[7] We additionally tried fitting mixed-effects models to the data, but these did not converge. At the same time, even the non-converged models yielded the values of β and *SE* close to those in the reported model, and the significance of the reported effects was not affected.

[8] A question that may arise is why the model performs rather poorly on two sentences in the converging condition: see the outliers in Figure 9, which appear in each group of simulations. These are the sentences *Die Tochter sucht der Löffel* 'The spoon looks for the daughter' and *Die Torte sucht der Löffel* 'The spoon looks for the cake'. Note that both sentences include the word form *der Löffel* at the second position, which denotes two morphological homonyms: singular number nominative case and plural number genitive case. This makes the given form ambiguous in isolation, and the model has to rely on other cues. In other sentences, such as *Den Sohn sucht der Löffel* 'The spoon looks for the son' or *Den Teller sucht der Löffel* 'The spoon looks for the plate', the model successfully resolves this ambiguity by attending to the accusative case marking on the noun at the first position:

den Sohn or *den Teller*. This strategy cannot be used in the outlier sentences, because the case marking of the first noun is also ambiguous: the nominative and the accusative coincide in the word forms *die Tochter* and *die Torte*. Thus, both nouns in each of the two sentences are ambiguous, and the model cannot interpret these sentences correctly. We believe that training the model on a larger corpus could provide it with more usages of nouns in the genitive case, helping to disambiguate the case marking in the target sentences.