Computational simulations of second language construction learning

Matusevych, Y.; Alishahi, A.; Backus, A.

Published in:
Proceedings of the fourth annual workshop on cognitive modeling and computational linguistics

Document version:
Publisher's PDF, also known as Version of record

Publication date:
2013

Citation for published version (APA):
Matusevych, Y., Alishahi, A., & Backus, A. (2013). Computational simulations of second language construction learning. In V. Demberg, & R. Levy (Eds.), *Proceedings of the fourth annual workshop on cognitive modeling and computational linguistics* (pp. 47-56). Stroudsburg: Association for Computational Linguistics.

Computational simulations of second language construction learning

Yevgen Matusevych
Department of Culture Studies

Afra Alishahi
Department of Communication and Information Sciences

Ad Backus
Department of Culture Studies

Tilburg University, PO Box 90153, 5000 LE Tilburg, the Netherlands

Y.Matusevych@uvt.nl A.Alishahi@uvt.nl A.M.Backus@uvt.nl

Abstract

There are few computational models of second language acquisition (SLA). At the same time, many questions in the field of SLA remain unanswered. In particular, SLA patterns are difficult to study due to the large amount of variation between human learners. We present a computational model of second language construction learning that allows manipulating specific parameters such as age of onset and amount of exposure. We use the model to study general developmental patterns of SLA and two specific effects sometimes found in empirical studies: construction priming and a facilitatory effect of skewed frequencies in the input. Our simulations replicate the expected SLA patterns as well as the two effects. Our model can be used in further studies of various SLA phenomena.

1 Introduction

Computational models have been widely used for investigating how humans learn and process their native language. Various models of child language acquisition have been applied to studies of speech segmentation (e.g., ten Bosch, Hamme, & Boves, 2008), word learning (e.g., Frank, Goodman, & Tenenbaum, 2009; Fazly, Alishahi, & Stevenson, 2010), induction of linguistic structure (e.g., Elman, 1990), etc. In comparison, the acquisition of second language has received little attention from the modeling community. Most of the child language acquisition models cannot be directly used for investigating how humans process and acquire foreign languages. In order to do so, we have to model the existing knowledge of first language—i.e., bilingualism.

Li (2013) provides a state-of-the-art overview of models of bilingualism. One of his claims is that most existing models account for mature adult speaker's knowledge and do not explain how foreign language develops in learners. In other words, there are several computational models of second language *processing* (e.g., Shook & Marian, 2013; Roelofs, Dijkstra, & Gerakaki, 2013; Yang, Shu, McCandliss, & Zevin, 2013, etc.), but only few of Second Language *Acquisition* (SLA). The latter mostly simulate lexis and semantics acquisition (e.g., Li & Farkas, 2002; Li, 2009; Cuppini, Magosso, & Ursino, 2013, etc.), and those that address a higher level of language structure usually do not model the existing L1 knowledge (e.g., N. C. Ellis & Larsen-Freeman, 2009; Rappoport & Sheinman, 2005; but see Monner, Vatz, Morini, Hwang, & DeKeyser, 2013).

At the same time, a number of theoretical SLA issues are not well explained yet, including general questions such as how existing knowledge of the first language influences the acquisition of second language. To give a specific example, it is not clear yet when L1 structures lead to interference and when they do not.

In this paper, we use an existing model of early acquisition of argument structure constructions (Alishahi & Stevenson, 2008) and adapt it to bilingual input data, which allows us to investigate the acquisition process in second language learners. We demonstrate in a number of computational simulations that our model replicates natural L2 developmental patterns and two specific effects observed in human L2 learners, thus allowing us to make certain predictions about the issues under investigation.

2 Description of the model

A usage-based approach to language claims that humans learn abstract linguistic regularities from instances of language use. Specifically, general argument structure constructions are predicted to emerge over time from individual verb usages which share syntactic and semantic proper-

Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, pages 47–56,
Sofia, Bulgaria, August 8, 2013.

ties. Argument structure constructions, according to Goldberg, Casenhiser, and White (2007), are "pairings of form and meaning that provide the means of expressing simple propositions in a language" (p. 74). Since nearly all human utterances contain propositions, the learner's knowledge of argument structure constructions must reflect the level of his grammatical competence.

The model of Alishahi and Stevenson (2008) is based on this approach: the building block of the model is an *argument structure frame*, a collection of syntactic and semantic features which represents a verb usage. Abstract constructions are formed by detecting and clustering similar frames, and various linguistic tasks are simulated by having the model predict the most suitable values for the missing features in a frame. These components are described in the following sections.

2.1 Argument structure frames

In our SLA model, each frame contains the following features:

- **Head verb** in its infinitive form.
- **Number of arguments** that the verb takes.
- **Semantic primitives of the verb** representing the conceptual characteristics of the event that the verb describes.
- **Semantic properties of each argument** representing its conceptual meaning, independently of the event that it participates in.
- **Event-based properties of each argument** representing the characteristics each argument takes on in the particular event it is participating in.
- **Syntactic pattern** of the utterance.

A sample frame is shown in Table 1. In Section 3.3 we will further explain how values for each frame feature are selected.

Table 1: An example frame extracted from a verb usage *Bill went through the maze.*

Head verb (V.)	*go*
No. of arguments	2
V. sem. primitives	act, move, walk
Arg.1 sem. prop-s	name, male, person, ...
Arg.2 sem. prop-s	system, instrumentality, ...
Arg.1 event prop-s	volitional, sentient, ...
Arg.2 event prop-s	location, path, destination
Syntactic pattern	AGENT V. through LOC.

2.2 Learning Constructions

Alishahi and Stevenson (2008) use an incremental Bayesian algorithm for clustering similar frames into constructions. Each input frame is compared to all the existing constructions and a potentially new one, and is added to the best matching construction:

$$\text{BestConstruction}(F) = \underset{k}{\text{argmax}}\, P(k|F) \qquad (1)$$

where k ranges over the indices of all constructions (index 0 represents the new construction). Using Bayes rule and dropping $P(F)$ which is constant for all k:

$$P(k|F) = \frac{P(k)P(F|k)}{P(F)} \sim P(k)P(F|k) \qquad (2)$$

The prior probability $P(k)$ indicates the degree of entrenchment of construction k:

$$P(k) = \frac{N_k}{N+1}, P(0) = \frac{1}{N+1} \qquad (3)$$

where N_k is the number of frames already clustered in construction k, and N is the total number of frames observed so far. The posterior probability of a frame F is expressed in terms of the (supposedly independent) probabilities of its features:

$$P(F|k) = \prod_{i \in Features(F)} P_i(j|k) \qquad (4)$$

where j is the value of the i^{th} feature of F, and $P_i(j|k)$ is the probability of displaying value j on feature i within construction k. This probability is estimated using a smoothed maximum likelihood formula.[1]

2.3 Bilingual acquisition

We accept the view that L1 and L2 learning have more commonalities than differences (see, e.g., MacWhinney, 2013), thus we do not explicitly encode the difference between the two languages. As the learner perceives a frame, he is not aware of which language the frame belongs to. All the input data are processed equally, so that constructions are formed in the same space and can contain frames from both languages. Such approach allows us to investigate how the existing L1 knowledge influences L2 acquisition.

[1] For single-valued features such as the head verb, likelihood is calculated by simply counting those members of construction k whose value for feature i exactly matches value j. For features with a set value such as the semantic properties of the verb and the arguments, the likelihood is calculated by comparing sets.

2.4 Sentence production

We use sentence production as our main evaluation task for SLA. Given a frame which represents an intended meaning through the semantic properties of an event (or verb) and its participants (or arguments), we want to predict the most probable values for the syntactic pattern feature. Following Alishahi and Stevenson (2008), we estimate the probability that feature i (in our case, the syntactic pattern) displays value j given other observed feature values in a partial frame F as

$$\begin{aligned} P_i(j|F) &= \sum_k P_i(j|k)P(k|F) \\ &= \sum_k P_i(j|k)P(k)P(F|k) \end{aligned} \tag{5}$$

The probabilities $P(k)$, $P(F|k)$ and $P_i(j|k)$ are estimated as before (see Equations 3 and 4). Ranging over the possible values j of feature i, the value of an unobserved feature can be predicted by maximizing $P_i(j|F)$[2]:

$$\text{BestValue}_i(F) = \underset{j}{\text{argmax}}\, P_i(j|F) \tag{6}$$

3 Data

For cognitively plausible computational simulations we had to prepare naturalistic input based on the suitable corpora. While there are available corpora that contain recordings of child-directed speech (MacWhinney, 2000), the resources containing speech addressed to L2 learners appear to be very limited. Therefore, our choice of languages (German as a L1, and English as a L2) was motivated first of all by the data availability. We extracted naturalistic L1 and L2 data from two different sources.

3.1 Data sources

L2 English data were extracted from the Flensburg classroom corpus (Jäkel, 2010) that contains transcripts of lessons of English (as a foreign language) taught to children in German schools that cover all school age groups. We estimated the total number of occurrences of different verbs in the corpus. From 20 most frequent verbs we selected 6 that represented syntactically and semantically different linguistic constructions, since constructional variability was one of the crucial factors for the model. The verbs are: *go*, *come*, *read*, *show*, *look* and *put*. For each verb, we extracted all its occurrences from the corpus.

For L1 we used German data extracted from the CHILDES database (MacWhinney, 2000), namely from adults' speech directed to three children: Caroline (age from 0;10 to 4;3; von Stutterheim, 2004), Kerstin (from 1;3 to 3;4; M. Miller, 1979) and Leo (from 1;11 to 4;11; Behrens, 2006). In the same manner as for the English data, we selected six verbs—*machen 'to make'*, *kommen 'to come'*, *gucken 'to look'*, *gehen 'to go'*, *sehen 'to see'* and *geben 'to give'*— and extracted all their occurrences from the three corpora. Since the corpora were of different size, the number of occurrences for some verbs were incomparable between the corpora, thus we balanced the size of the samples used for further analysis by taking equal numbers of random verb uses from each corpus.

3.2 Data annotation

Since the basic input unit for our computational model was a frame, we manually annotated all the verb occurrences in order to extract frames. Approximately 100 instances per verb were annotated using the following general guidelines.

1. Instance grouping is based on the semantics of the main verb and its arguments as well as on the syntactic pattern.
2. We consider only arguments (both obligatory and optional), but not adjuncts, since there is evidence that the two are processed differently (see, e.g., Kennison, 2002).
3. We discard all instances where the main verb was represented by a compound form or by an infinitive, or appeared in a subordinate clause, since in all these cases the "core" frame of the argument structure construction might obtain additional structural or semantic characteristics.
4. We do consider imperatives and questions whose form does not contradict the previous point.
5. We treat German prefixed/particle verbs (e.g., *zumachen 'to close'*) and English compound verbs as an instance of the base verb (in this case, *machen 'to make'*), given that the prefixed/particle verb meaning is compositional and the prefix/particle is actually separated.
6. Considering the previous point, each particle/prefix in our instances represents an in-

[2] A non-deterministic alternative that we have to consider in the future is to sample the feature value from the estimated distribution.

dependent semantic component (see, e.g., Dewell, 2011, for detailed explanation), and we treat them as separate arguments.

7. We discard all the instances in which the verb is used in a formulaic sequence (e.g.,, *Wie geht's? 'How are you?'*), because formulaic sequences are believed to be processed and acquired as a whole (e.g., Wray, 2005; Bannard & Lieven, 2012).
8. Finally, we eliminate the case marking in German and use the Nominative case for all the arguments, because this feature is not crucial for our model, and there is evidence that German children before the age of 7 mostly rely on other features such as word order (Dittmar, Abbot-Smith, Lieven, & Tomasello, 2008).

3.3 Frame extraction

From the annotated data samples, for each verb we extracted frames and their respective frequencies of occurrence. Following Alishahi and Stevenson (2010), the semantic primitives of verbs and their arguments were semi-automatically extracted from WordNet (G. A. Miller, 1995), and the event-based properties of the arguments were manually compiled.

The syntactic pattern of the frame not only shows the order of the arguments, but also implicitly includes information about their semantic roles, i.e., AGENT, THEME, LOCATION, etc. Note that these semantic labels are used only for distinguishing between similar syntactic patterns with the verb in the same position but swapped arguments (cf. *[so] schnell geht es* vs. *es geht [so] schnell* 'it goes [so] fast'—both patterns occur rather frequently in German).[3]

Based on the manually extracted frames, an input corpus of verb uses was automatically generated for each set of experiments. The frequency of occurrence of each frame determined the probability of selecting this frame, and the same method was used for selecting specific arguments.

[3]Although the inclusion of semantic labels into syntactic pattern makes the learning task easier, there is, in fact, no agreement yet on how exactly children acquire the noncanonical word order. They must rely on pragmatics, and this phenomenon most thoroughly has been studied in the generative tradition under the name of scrambling, but still various explanations were proposed (see, e.g., Mykhaylyk & Ko, 2011). Due to this uncertainty, we found it acceptable to provide the learner with the means to distinguish between the patterns like in the example above, since it was highly important for German with its partially free word order.

4 Simulations and results

In this section we report on computational simulations that we ran using our model and the described input data. We investigate general L2 developmental patterns, priming effects in SLA, and the impact of skewed input on the learner's L2 proficiency. Although the latter two are not SLA phenomena per se and can be observed in L1 learners as well, they have been discussed in SLA domain and suit well our methodological framework.

4.1 L2 general development

Despite numerous attempts to capture and describe the dynamics of SLA, scholars admit that there is no 'typical' profile of general L2 development (for an overview, see Hasko, 2013). This is because many variables are involved, such as the learner's L1, the age of L2 onset, amount of input, type of instruction (if any), etc. They cause significant differences between individual learners and specific linguistic phenomena.

Generally, L2 develops gradually, and second language learners rarely achieve native-like L2 proficiency. To demonstrate that our model follows these patterns, we ran a number of simulations to compare how L1 and L2 proficiency changes over time. In our scenario, the learner was first presented 500 L1 verb uses in small steps (25 times 20 frames). After each step his L1 proficiency was tested in the following way. The learner was presented with 20 test frames in which the syntactic pattern was removed, and had to predict the most suitable syntactic pattern, relying on his current knowledge. We should note that because German has partially free word order, our German data contained a substantial number of frame groups consisting of two or more frames that were almost identical and differed only in the order of arguments in their syntactic patterns (i.e., AGENT *verb* THEME and THEME *verb* AGENT). These patterns are very close both linguistically (i.e., they carry very similar meanings) and algorithmically (i.e., the learner's preference for one of them is determined only by their respective frequencies of occurrence in the input). Therefore, asking the learner to predict the exact pattern would not be a fair task. For this reason, during the evaluation we only checked whether the pattern produced by the learner contained exactly the same set of arguments (and, possibly, the same preposition) as the target pattern. Thus, AGENT *kommen* THEME, *kommen* AGENT THEME, and THEME *kommen* AGENT were considered equal for the purpose of evaluation.

After the initial 25 steps of L1 training and testing, the learner was presented 500 more frames (25 times 20) which could be either from L1 or from L2 data in proportion 3 (L1) to 1 (L2). This way we simulated a common situation when a child starts learning a foreign language at school, thus being exposed to input from both languages, but L1 input prevails. The results averaged over 10 simulations (Figure 1) demonstrate that the L2 proficiency does not achieve that of L1.[4]

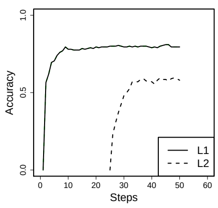

Figure 1: L1 and L2 development over time

We explain the lower L2 proficiency by two factors. First, by the moment when the learner started receiving L2 input, L1 constructions were already formed in his memory, so the L1 entrenchment prevented L2 constructions to fully emerge. Second, even within the period of SLA the amount of L2 input was 3 times smaller compared to that of L1. To investigate whether both factors were indeed important, we tried to eliminate each of them separately, i.e., to present both L1 and L2 from the very beginning keeping the ratio 3:1 (Figure 2, left), or to set an equal ratio while keeping the late age of L2 onset (Figure 2, right). As we can see, in neither case does the L2 proficiency reach that of L1. However, when both factors are eliminated—that is, from the very beginning the learner receives mixed L1/L2 input in equal proportion—he reaches comparable levels of L1 and L2 proficiency (Figure 3).

Additionally, we tried to separately manipulate each of the two parameters keeping the other one constant. We expected that (1) the lower the L2 age of onset, the higher the learner's proficiency at each moment of time with the L1/L2 ratio set at 3:1, and (2) the smaller the L1/L2 ratio (down to 1, when the amount of input is equal), the higher

[4]After presenting 4,000 more L2 frames to the learner this pattern was still observed, and neither L1 nor L2 proficiency converged to 1.

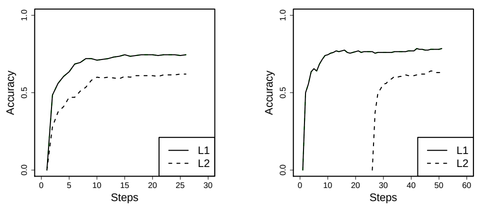

Figure 2: L1 and L2 proficiency provided equal age of onset (left) or input ratio (right)

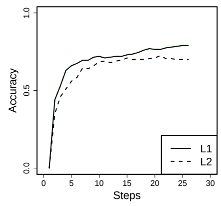

Figure 3: L1 and L2 proficiency provided equal learning conditions

the learner's proficiency at each moment of time with the age of onset set at 500 frames. We found no evidence for either effect. Part of the explanation might be that there was a substantial overlap between L1 and L2 syntactic patterns (especially considering we treated patterns as sets of elements irrespective of word order). Therefore the learner's existing L1 knowledge may indirectly have contributed to the L2 proficiency, in a pattern known as "positive transfer" (see, e.g., Benson, 2002). This can be demonstrated by comparing the initial slopes of L2 development lines in Figure 1 and Figure 2a. In the former case, representing L2 exposure after L1 constructions have already been entrenched, L2 acquisition goes faster in its initial stages, because the learner has, in fact, already acquired a number of syntactic patterns that are shared by the two languages. Monner et al. (2013), who computationally studied the effect of French L1 entrenchment on Spanish L2 grammatical gender learning, explain an exception in their results in similar fashion. However, this requires further investigation, possibly in simulations involving two languages that are typologically more distant.

4.2 Priming effects in L2

Structural priming effects, when speakers tend to recreate a recently encountered linguistic structure in further language use, have been demonstrated both in first (e.g., Bock, Dell, Chang, & Onishi, 2007; Potter & Lombardi, 1998, etc.) and in second language (e.g., McDonough, 2006; Gries &

Wulff, 2005) as well as across the two (e.g., Loebell & Bock, 2003; Vasilyeva et al., 2010). Some of these effects are explained in terms of construction grammar—primes can activate the respective constructions (see Goldberg & Bencini, 2005).

To give a specific example, Gries and Wulff (2005) asked L1 German learners of English to complete sentence fragments after being exposed to a prime sentence, which contained either a prepositional dative (*The racing driver showed the torn overall to the team manager.*) or a ditransitive construction (*The racing driver showed the helpful mechanic the damaged tyre*). The sentences produced by the learners demonstrated the constructional priming effect in L2 acquisition, which was also supported by corpus and sorting evidence (see Gries & Wulff, 2005, for details).

Since in our model we explicitly assume the existence of constructions in learner's memory, we should be able to observe constructional priming effects in L2. To investigate this, we partially simulated the experiment of Gries and Wulff (2005) computationally. First the model was presented with 250 L1 verb uses[5], after which, like in the previous experiment, L2 was introduced in parallel with L1 in small steps (25 times 10 frames). After each step, the learner was additionally presented with one of two primes. Priming frames, which we took from the actual dataset, were uses of the verb *show* with variable arguments, and the only difference between the two primes was the syntactic pattern—a prepositional dative or a ditransitive (see Table 2).

Table 2: The two primes used.

Head verb (V.)	*show*
No. of arg.	3
V. sem. prim.	act, cause, perceive
Arg.1 sem. prop.	*vary*
Arg.2 sem. prop.	*vary*
Arg.3 sem. prop.	*vary*
Arg.1 ev. prop.	volitional, sentient, ...
Arg.2 ev. prop.	sentient, animate, ...
Arg.3 ev. prop.	perceivable, ...
Synt. pattern	AG. *show* BENEF. THEME or AG. *show* THEME *to* BENEF.

[5]Since the impact of a single priming frame on the learner could be insignificant, we used a smaller step size in these simulations.

In the experiment by Gries and Wulff (2005) learners, after seeing a prime, were presented with a test fragment consisting of an agent and a verb (*The racing driver showed* ...), and were required to continue the sentence. In terms of our model, the test frame consisted of the head verb (*show*) and its semantic primitives, total number of arguments, the first argument (pronoun *you*) and its semantic and event properties. The other features (i.e., syntactic patterns and all the properties of the other two arguments) were missing, and the learner had to predict the best syntactic pattern for the test frame. After the prediction was made, both prime and test frame were discarded in order not to influence further results, and the learning continued.

Since we investigated priming effects in ditransitive (D) and prepositional dative (P) constructions, in the further analysis we only looked at the two respective syntactic patterns in the learner's production. That is, we calculated how many patterns of each type were produced after each prime (i.e., D-patterns after D-prime, P-patterns after D-prime, P-patterns after P-prime, and D-patterns after P-prime). Additionally, we ran an identical baseline simulation where the learner was not primed, being presented a test frame immediately after each learning step. Figure 4 shows how many P- and D-patterns were produced in each of the three conditions (*P-prime*, *D-prime* and *no prime*; the results are averaged over 100 simulations).

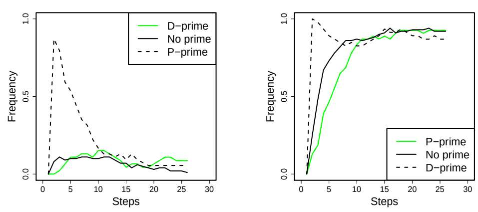

Figure 4: Frequency of prepositional (left) and ditransitive (right) pattern production

As we can see, on the initial 5-10 steps of development both P- and D-patterns were produced substantially more often after the respective matching prime (the jump of the dotted line on each plot) than after the non-matching prime or after no prime. After some time, however, the priming effect was leveled off, presumably because of the exposure to large amounts of training data, and the frequency of production of each of the two patterns aligned with the actual frequency of occurrence of the respective pattern in the training data (31 for D-pattern, 3 for P-pattern).

On the one hand, the presence of the priming effect in our results is in line with the findings of Gries and Wulff (2005). On the other hand, their participants were advanced foreign learners of English who must have achieved rather high proficiency in L2 by the moment of study, but they were still sensitive to the priming effect—a result that we could not replicate computationally.

4.3 Skewed vs. balanced L2 input

There is an ongoing discussion in the literature on the supposed facilitatory effect of skewed input on constructional acquisition, summarized by Boyd and Goldberg (2009). In monolingual contexts, it has been demonstrated that children (Casenhiser & Goldberg, 2005) and adults (Goldberg, Casenhiser, & Sethuraman, 2004) acquire a novel construction with artificial verbs faster if one verb has higher token frequency in the input compared to the other verbs, and slower in case of balanced input, with all the verbs having equal token frequencies.

As for SLA, N. C. Ellis and Ferreira-Junior (2009) showed that the distribution of verbs/constructions in input to L2 learners is Zipfian, and that the most frequent verb in each construction is acquired first. However, they do not provide evidence for a facilitatory effect of skewed distribution on construction learning. At the same time, there is experimental evidence that high type frequency facilitates the acquisition of *wh*-questions in L2 (McDonough & Kim, 2009).

Year and Gordon (2009) experimentally studied the facilitatory effect of skewed verb frequency in the input on L2 constructional learning. In their study, L1 Korean learners of English were presented with 5 English verbs in the ditransitive construction, where either all the verbs appeared equally often (balanced input), or one verb appeared 6 times more often than the other (skewed input). The learners' knowledge of the construction was assessed in the elicited production and acceptability judgement task. The exposure and testing procedures were distributed over 8 weeks, or over 4 weeks, or over 4 days, depending on the group. Surprisingly, in no group they found the evidence for the facilitatory effect of skewed input. These findings contrast with those in the other studies that we mentioned.

In order to address this issue computationally, we ran simulations using our model. Unlike Year and Gordon (2009) who investigated the acquisition of one construction only, we assessed the general L2 knowledge of all constructions that the learner was exposed to, since our model is perfectly suited for this.

The frequency distribution of verbs in our naturalistic L2 input was not uniform (79-81-61-58-48-29), however the most frequent verb appeared approximately 3 times more often than the least frequent, which was not comparable to the ratio of 1:6 in the study by Year and Gordon (2009). Thus, in addition to the natural data we introduced two more conditions. First, we estimated the distribution of verbs over different constructions in our data and concluded that two verbs—*go* and *show*—accounted for most syntactic patterns in the input. Therefore, to prepare truly skewed input data, we set the frequencies for these two verbs to 30 and for the other verbs to 1[6]. Second, we prepared the balanced input data by setting the frequency of each verb to 1.

Using the three types of input, we ran the exact same simulations as for investigating the general developmental pattern, and compared the learner's L2 proficiency over time in the three conditions. The results are shown in Figure 5.

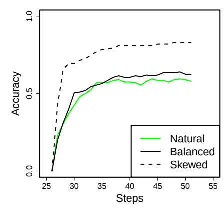

Figure 5: L2 proficiency over time on skewed vs. balanced input

As we can see, the learner's proficiencies with the natural and balanced input data do not differ much. However, the facilitatory effect of the skewed frequencies in the input is very evident. Thus, our findings constrast with the results of Year and Gordon (2009), but are in line with the general trend as summarized by Boyd and

[6]Although the ratio of 30:1 is much higher than that in the experiment being simulated, we had to account for the fact that individual frames within each verb were assigned their own frequencies, so a high-frequency frame of a low-frequency verb could still appear more often in the input than a low-frequency frame of a high-frequency verb. We excluded this possibility by setting the ratio to the high value.

Goldberg (2009). We agree with Year and Gordon's (2009) explanation that the lack of facilitatory effect that they found can be explained by the presentation order of the high-frequency verbs. Goldberg et al. (2007) demonstrated the effect of the presentation order of high- and low-frequency stimuli on the learners' performance. We believe that due to the rather large ratio 30:1 that we set in the skewed data, the two high-frequency verbs prevailed in the L2 input from the very initial stage of L2 learning, therefore our simulations were closer to the "skewed first" condition of Goldberg et al. (2007) than to the "skewed random" condition.

We have to note, however, that the facilitatory effect observed in our experiment could also be due to the fact that the distribution of the verbs in the test frames was also different for each of the three conditions, since the test data were sampled from the same distribution as the training data. We will further investigate this issue in the future.

5 Discussion

Patterns of second language development have been studied for decades, starting from the morpheme learning studies in 1970s (e.g., Wode, 1976). Although some classroom studies allow SLA theorists to make inferences about general L2 developmental patterns (e.g., R. Ellis, 1994; VanPatten & Benati, 2010), scholars agree that a typical pattern of L2 development can hardly exist due to the inherent complexity of the SLA process. The enormous variability of L2 learning conditions makes it difficult to provide general conclusions about SLA development. Partly for this reason, most longitudinal studies have been focusing on the development of specific linguistic features in small number of individuals (see an overview by Ortega & Iberri-Shea, 2005). DeKeyser (2013) emphasizes the methodological difficulties in this domain, especially when it comes to studying age effects in the second language of immigrant population. The inherent problems of documenting the individuals' language experience and sampling those learners who match a number of specific criteria make the research in this field very laborious and time-consuming.

In contrast, a computational framework can be effectively used for studying the complexities of learning a second language, specifically in relation to the characteristics of the first language. We present a computational model of second language acquisition which investigates grammatical L2 development in connection with the existing L1 knowledge, a setup that has not been properly addressed by the existing computational models of SLA (but see Monner et al., 2013).

We evaluate the model's acquired grammatical knowledge (in the form of emergent argument structure constructions) through sentence production. Our simulations replicate the expected patterns of L2 development, such as gradual emergence of constructions and increased proficiency in sentence production. Moreover, we investigate two specific SLA phenomena: construction priming and the facilitative effect of skewed frequencies in the input.

Priming effects have been demonstrated in second language learners (Gries & Wulff, 2005), although sometimes inconsistently (McDonough, 2006). We replicate a priming effect at the early stages of learning in our simulations, but this effect diminishes as the model receives more input. Systematic manipulation of various (potentially relevant) factors via computational simulation will shed more light on the nature of priming in SLA.

The facilitative effect of skewed input on construction learning has been subject of much debate (Boyd & Goldberg, 2009). Our experiments show that skewed frequencies in the input can improve the performance of the model in sentence production, but more careful investigation of this pattern is needed for a clear picture of the interaction between different parameters.

Although some of our results are inconclusive, we believe that our preliminary experiments clearly demonstrate the opportunities of the model for SLA research. In the future we plan investigating the described and other phenomena more thoroughly. Applying additional methods such as analysis of the frame categorization structure under different conditions, or quantitative comparison of the production data obtained in computational simulations and in the natural learner corpora (Gries & Wulff, 2005), could help us to draw specific implications for the SLA theory.

References

Alishahi, A., & Stevenson, S. (2008). A computational model of early argument structure acquisition. *Cognitive Science*, *32*(5), 789–834.

Alishahi, A., & Stevenson, S. (2010). A computational model of learning semantic roles from child-directed language. *Language and Cognitive Processes*, *25*(1), 50–93.

Bannard, C., & Lieven, E. (2012). Formulaic language in L1 acquisition. *Annual Review of Applied Linguistics*, *32*, 3–16.

Behrens, H. (2006). The input-output relationship in first language acquisition. *Language and Cognitive Processes*, *21*(1-3), 2–24.

Benson, C. (2002). Transfer/cross-linguistic influence. *ELT Journal*, *56*(1), 68–70.

Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, *104*(3), 437–458.

Boyd, J. K., & Goldberg, A. E. (2009). Input effects within a constructionist framework. *The Modern Language Journal*, *93*(3), 418–429.

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science*, *8*(6), 500–508.

Cuppini, C., Magosso, E., & Ursino, M. (2013). Learning the lexical aspects of a second language at different proficiencies: A neural computational study. *Bilingualism: Language and Cognition*, *16*, 266–287.

DeKeyser, R. M. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning*, *63*, 52–67.

Dewell, R. (2011). *The Meaning of Particle / Prefix Constructions in German*. John Benjamins.

Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Development*, *79*(4), 1152–1167.

Ellis, N. C., & Ferreira-Junior, F. (2009). Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics*, *7*(1), 187–220.

Ellis, N. C., & Larsen-Freeman, D. (2009). Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning*, *59*, 90–125.

Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford University Press.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, *34*(6), 1017–1063.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers referential intentions to model early cross-situational word learning. *Psychological Science*, *20*(5).

Goldberg, A. E., & Bencini, G. M. L. (2005). Support from language processing for a constructional approach to grammar. In A. Tyler (Ed.), *Language in Use: Cognitive and Discourse Perspectives on Language and Language Learning*. Georgetown University Press.

Goldberg, A. E., Casenhiser, D., & White, T. (2007). Constructions as categories of language. *New Ideas in Psychology*, *25*(2), 70–86.

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, *15*(3), 289–316.

Gries, S. T., & Wulff, S. (2005). Do foreign language learners also have constructions? *Annual Review of Cognitive Linguistics*, *3*(1), 182–200.

Hasko, V. (2013). Capturing the dynamics of second language development via learner corpus research: A very long engagement. *The Modern Language Journal*, *97*(S1), 1–10.

Jäkel, O. (2010). Working with authentic ELT discourse data: The Flensburg English Classroom Corpus. In R. Vogel & S. Sahel (Eds.), *NLK Proceedings 2010* (pp. 65–76). Universität Bielefeld.

Kennison, S. M. (2002). Comprehending noun phrase arguments and adjuncts. *Journal of Psycholinguistic Research*, *31*(1), 65–81.

Li, P. (2009). Lexical organization and competition in first and second languages: computational and neural mechanisms. *Cognitive Science*, *33*(4), 629–664.

Li, P. (2013). Computational modeling of bilingualism: How can models tell us more about the bilingual mind? *Bilingualism: Language and Cognition*, *16*, 241–245.

Li, P., & Farkas, I. (2002). A self-organizing connectionist model of bilingual processing. In R. R. Heredia & J. Altarriba (Eds.), *Bilingual Sentence Processing* (Vol. 134, pp. 59–85). Elsevier Science.

Loebell, H., & Bock, K. (2003). Structural priming across languages. *Linguistics*, *41*, 791–824.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.

MacWhinney, B. (2013). The logic of the unified model. In S. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 211–227). Taylor & Francis Group.

McDonough, K. (2006). Interaction and syntactic priming: English L2 speakers' production of dative constructions. *Studies in Second Language Acquisition*, *28*(2), 179–207.

McDonough, K., & Kim, Y. (2009). Syntactic priming, type frequency, and EFL learners' production of wh-questions. *The Modern Language Journal*, *93*(3), 386–398.

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, *38*(11), 39–41.

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

Monner, D., Vatz, K., Morini, G., Hwang, S.-O., & DeKeyser, R. M. (2013). A neural network model of the effects of entrenchment and memory development on grammatical gender learning. *Bilingualism: Language and Cognition*, *16*, 246–265.

Mykhaylyk, R., & Ko, H. (2011). Optional scrambling is not random: Evidence from English-Ukrainian acquisition. In M. Anderssen, K. Bentzen, & M. Westergaard (Eds.), *Variation in the Input* (Vol. 39, pp. 207–240). Springer.

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, *25*, 26–45.

Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, *38*(3), 265–282.

Rappoport, A., & Sheinman, V. (2005). A second language acquisition model using example generalization and concept categories. In *Proceedings of the Second Workshop on Psychocomputational Models of Human Language Acquisition* (pp. 45–52). Omnipress Inc.

Roelofs, A., Dijkstra, T., & Gerakaki, S. (2013). Modeling of word translation: Activation flow from concepts to lexical items. *Bilingualism: Language and Cognition*, *16*, 343–353.

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

ten Bosch, L., Hamme, H. V., & Boves, L. (2008). A computational model of language acquisition: Focus on word discovery. In *Proceedings of Interspeech 2008* (pp. 2570–2573). ISCA.

VanPatten, B., & Benati, A. (2010). *Key Terms in Second Language Acquisition*. Bloomsbury.

Vasilyeva, M., Waterfall, H., Gámez, P. B., Gómez, L. E., Bowers, E., & Shimpi, P. (2010). Cross-linguistic syntactic priming in bilingual children. *Journal of Child Language*, *37*(5), 1047–1064.

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

Wode, H. (1976). Developmental sequences in naturalistic L2 acquisition. *Working Papers on Bilingualism*, *11*, 1–31.

Wray, A. (2005). *Formulaic Language and the Lexicon*. Cambridge University Press.

Yang, J., Shu, H., McCandliss, B. D., & Zevin, J. D. (2013). Orthographic influences on division of labor in learning to read Chinese and English: Insights from computational modeling. *Bilingualism: Language and Cognition*, *16*, 354–366.

Year, J., & Gordon, P. (2009). Korean speakers' acquisition of the English ditransitive construction: The role of verb prototype, input distribution, and frequency. *The Modern Language Journal*, *93*(3), 399–417.