A Computational Study of Cross-situational Lexical Learning of Brazilian Portuguese

Pablo Faria
University of Campinas
R. Sérgio Buarque de Holanda, 571
Campinas, Brazil
pablofaria@gmail.com

Abstract

In this paper, a particular algorithm for lexical acquisition – taken as a problem of learning the mapping from words to meanings – is evaluated. The algorithm in Siskind (1996) is adapted to handle more complex input data, including data of Brazilian Portuguese. In particular, the input data in the present study covers a broader grammatical knowledge, showing both polysemy and higher inflectional and agreement morphology. Results indicate that these properties create difficulties to the learner and that more substantial developments to the algorithm are needed in order to increase its cross-linguistic capabilities.

1 Introduction

Computational modeling, as an empirical approach to theoretical problems, has the benefit of demanding clear and exhaustive specification of the problem under consideration (Pearl, 2010; Yang, 2011). In this paper, we consider a computational model of lexical acquisition by a child learning her native language. Lexical acquisition is taken here as a problem of learning the mapping from words to meanings based on a cross-situational strategy. Simply put, cross-situational lexical learning is the strategy by which word-to-meaning mappings are learned by assigning to a given word the meanings which are consistent across the situations where the word is heard. One computational modeling of this strategy is provided in Siskind (1996).

We present an implementation of Siskind's algorithm, which is part of a broader computational model of first language acquisition presented in Faria (2013). It is evaluated against informationally and morphologically more complex input data, including data of Brazilian Portuguese. As shown below, both aspects have an impact on the learner's performance. Consequently, a better understanding of them is necessary in order to progress towards learning models with wider grammatical and languages coverage.

The reader is referred to Siskind's (1996) arguments on the empirical plausibility of the model and for it being an approximation to the empirical problem of lexical acquisition through *cross-situational learning* which is taken in the psycholinguistic literature as a plausible learning strategy (Pinker, 1989; Fisher et al., 1994). Nonetheless, as stressed by Siskind, it is not claimed that the child employs the particular heuristics presented here. The main goal, instead, is to provide a proof of existence for an algorithm that solves approximations to the problem.

2 Lexical acquisition in the model

The lexical acquisition procedure presented in this paper is part of a broader first language acquisition model (Faria, 2013) which aims to simulate the acquisition of word to meaning mappings as well as syntactic knowledge. The model was also aimed at dealing with Brazilian Portuguese (BP) input data as well as with some issues of word order which were evaluated through an artificial corpus built with English vocabulary but displaying a strictly head-final order. Given its characteristics, the model can be included among somewhat similar studies found in the literature, such as Berwick (1985), Gaylard (1995), and Villavicencio (2002), among others.

The procedure is based on Siskind's (1996) heuristics, adapted in order to meet the goals of the modeling. One goal is to account for a greater variety of grammatical phenomena.[1] A second goal

[1] In Siskind's (1996) study, functional elements, such as articles, have no semantic-conceptual content, being acquired as lexical items that do not contribute to the meaning of sentences. This is a simplification not assumed in the model pre-

Proceedings of the Sixth Workshop on Cognitive Aspects of Computational Language Learning, pages 45–54,
Lisbon, Portugal, 18 September 2015. ©2015 Association for Computational Linguistics.

is to account for a greater variety of languages which, in the present study, consists in extending learning to Brazilian Portuguese, a language which, for being of a different family (as compared to English), shows properties that pose difficulties to the original learning heuristics, as is shown in what follows.

2.1 Summary of Siskind's (1996) simulation

Siskind presents an algorithm consisting of a series of ordered heuristics. The heuristics were conceived to guarantee an efficient and successful learning under different conditions, that is, in the presence of noise (utterances paired with incorrect meanings), "referential uncertainty" (utterances paired with more than one partially correct meaning) and homonymy. The corpus used in the simulations was based on a simple context free grammar which randomly generated only simple declarative sentences, pre-segmented and without adjectives and other adjuncts.

Functional words, such as determiners, were assumed not to contribute meaning to sentences. The MLU of sentences varied from 4.99 to 6.29 and all sentences had between 2 and 30 words and no more than 30 conceptual symbols. Simulations evaluated different parameterizations for (i) the size of the vocabulary (1000 to 10,000), (ii) the degree of referential uncertainty (i.e., the number of meanings paired with an utterance), (iii) the noise rate (0 to 20%), (iv) the number of conceptual symbols (250 to 2,000), and (v) the mean rate of homonymy in the corpus (between 1 and 2).

Results showed that the parameters (ii) and (iv) seem not to affect the convergence of the learning process. Therefore, the apparent complexity of the discourse context and that the potentially infinite number of concepts we may entertain seem to be efficiently handled by a cross-situational learning strategy. All other parameters had an impact in the learning curve, but the rate of homonymy was crucial: while 10,000 words were sufficient for convergence given a rate of 1 (i.e., no homonymy), 900,000 words were necessary for convergence given a rate of 2. Learning is slow for the first 25 words and increases until most of the vocabulary is learned. In late stages, words can be learned even with one exposition.

Finally, Siskind emphasizes limitations of the algorithm. First, it assumes strict homonymy, that is, words may have completely distinct meanings, but not partially distinct. Thus, polysemy may pose difficulties to the author's heuristics. The semantic-conceptual representation is simplified, not only for leaving aside the semantic content of functional words, but also as a consequence of a restricted grammatical coverage.

2.2 Lexical processing

In this model, lexical recognition and acquisition are part of the same process. At any given moment, the recognition of an utterance consists in obtaining the cross product of the sense sets of its words – Siskind names each combination as a "possible sense assignment" (PSA) – and, once the set of PSAs is obtained, identifying the PSA that is both consistent with the utterance (i.e., all words contribute to its meaning) and, in the case that more than one PSA is consistent, has the highest confidence factor (explained later).

2.3 The input data

The input data in this study is different from Siskind's (1996). First, it better reflects the distribution of types of utterances found in child directed data (Hoff-Ginsberg, 1986; Cameron-Faulkner et al., 2003). Second, by assumption, it more appropriately reflects the nature of the data that a child is exposed to.

2.3.1 Distribution

Hoff-Ginsberg (1986) studies the effects of functional and structural properties in the speech of mothers on the syntactic development of their children. Part of the author's findings is presented below, summarized in Table 1.

Measure	*M*
Measures of syntactic complexity	
MLU	4.47
VP/utterance	.95
NP/utternace	1.60
Auxiliaries/VP	.29
Words/NP	1.33
Frequencies of sentence forms (% of all utterances)	
Declaratives	25
Yes/no questions	15
Wh- questions	17
Imperatives	8
Interjections	17

Table 1: Structural Properties of Mothers' Speech in Hoff-Ginsberg (1986).

sented here.

Cameron-Faulkner et al. (2003) provide a slightly more detailed description of these structural properties, as shown in Table 2. Fragments are utterances with one or more words, the latter consisting of NPs (43%), VPs (23%), PPs (10%) and other (24%). Complex constructions are sentences with sentential complements, as in "*I think it's going to rain*", and subordinate adverbial clauses introduced by *because*, *if* and *when*.

Type	Mean proportion	Tokens
Fragments	.20 (.13–.32)	3351
One word	.07	
Multi-word	.14	
Questions	.32 (.20–.42)	5455
Wh-	.16	
Yes/no	.15	
Imperatives	.09 (.05–.14)	1597
Copulas	.15 (.08–.20)	2502
Subject–predicate	.18 (.14–.26)	2970
Transitives	.10	
Intransitives	.03	
Other	.05	
Complex	.06 (.03–.09)	1028

Table 2: Survey of Child Directed Speech in Cameron-Faulkner et al. (2003).

By collapsing their findings, we arrived at the frequencies shown in Table 3, used in the generation of the input data for the model. Frequencies in the interior of each type are not controlled, that is, subtypes have random frequencies. With respect to similar models in the literature, the grammatical coverage is larger, although far from covering the full grammatical knowledge of a speaker.

Type	H-G	Cetal.	This study
Fragments	–	.20	.20
Questions	.32	.31	.32
Wh-	.17	.16	
Yes/no	.15	.15	
Imperatives	.08	.09	.09
Declaratives	.25	.39	.39
Total			1.00

Table 3: Types and frequencies of utterance types assumed in the present simulation. "H-G" stands for Hoff-Ginsberg (1986) and "Cetal." for Cameron-Faulkner et al. (2003).

2.3.2 Linguistic properties

This model embodies a richer diversity of word classes and utterance types. For a detailed view of these, I refer the reader to Faria (2013, p.154-155). A direct consequence is that polysemy in the input is higher. As one example, since inchoative uses of verbs are included in the input, it will have the learner dealing with potentially one extra (non-causative) sense for each verb of change of state. The verb "break", for instance, may appear in "John broke the car" and "The car broke", utterances which by assumption differ in terms of causativity. Thus, one of the goals of this modeling is to evaluate the learner's performance given more polysemy in the input.

2.4 The learning procedure

In the end of this section, an illustration of the functioning of the heuristics is provided. We refer the reader to Siskind (1996) for a lengthy discussion about the reasoning behind each heuristic. In what follows, the heuristics assumed are presented and the main adaptations to the original highlighted. As in the original procedure, for learning to be possible the lexicon LEX is organized in three tables:

1. Table N, which maps a sense to its *necessary* conceptual symbols;
2. Table P, which maps a sense to its *possible* conceptual symbols;
3. Table D, which maps each sense to its possible conceptual expressions.

Word symbols may have more than one *sense*, one for each of its meanings in cases of homonymy or polysemy. The following set of heuristics (rules 1 to 5) is applied to each of the PSAs generated for a given utterance, as explained in the previous section.

Rule 1. *Ignore a PSA when (i) at least one symbol from the meaning of the utterance is absent from all P(w), and (ii) not all N(w) contribute to the meaning of the utterance.*

Rule 2. *For each word w of the utterance, remove from P(w) any symbol not included in the utterance meaning.*

Rule 3. *For each word w of the utterance, add to N(w) any conceptual symbol exclusively in P(w) (thus, absent from the P set of the remaining words).*

Rule 4. *For each word w in the utterance, remove from P(w) any conceptual* *symbol that appears only once in the utterance meaning and is included in the N(w') for some other word w' of the utterance.*

Rule 5. *For each word w in the utterance, if w converged for its conceptual symbol set, that is, N(w) = P(w), remove from D(w) any expression that does not involve the conceptual symbols in N(w); if the word has not yet converged, remove from D(w) any expression that includes a symbol not in P(w).*

The original "Rule 1" in Siskind (1996, p.57) was conceived to deal with referential uncertainty. However, in the present study this parameter is not evaluated. Thus, the original rule being irrelevant, an alternative rule is conceived to deal with the possibility that the words of a sentence may never contribute the whole meaning of an utterance. In the present study, this is a consequence of including conceptual symbols for the utterance type, for instance, DECL for declarative sentences, which have no morphological realization in languages like English and Brazilian Portuguese. The original Rule 1 would discard relevant PSAs because at some point the symbol DECL would be absent from all P(w) (the set P for a word w), that is, at some point there would be no word in any utterance which could possibly contribute DECL.

Siskind proposes a sixth heuristic that is put aside here. Its task is to check if there is at least one combination of the subexpression for the words in the utterance that matches exactly the utterance meaning. Since in this study, words in a given utterance may not contribute all the conceptual symbols present in the utterance meaning, this rule would cause problems to the learner. Although acknowledging that a different version of this rule may still be useful, the learning procedure in the present study has only the five rules shown above.

Three situations may arise, after an utterance is processed: (i) the algorithm converges to an unique consistent PSA; (ii) it converges to a set of consistent PSAs; and (iii) no PSA is found to be consistent with the utterance meaning. In the first case, the confidence factors for the senses involved are incremented. In the second, the algorithm first identifies the PSA with the highest current confidence factor and then update the confidence factor of the senses involved. In the last case, the algorithm determines the least number of words to be updated in their P and D sets. If it identifies some, the utterance is processed again. Otherwise, the utterance is discarded. As we can see, the confidence factor is a simple measure that allows the learner to converge to more consistent senses while gradually eliminating incorrect lexical entries.

2.5 An illustration

Let us assume that at some given stage, the learner shows the following partial non-converged lexicon:

	N	P
John	{**John**}	**John**, **ball**
took	{CAUSE}	CAUSE, WANT, BECOME, **take**, PAST
the	{}	WANT, **arm**, DEF
ball	{**ball**}	**ball**, **take**

Now, suppose that the learner is presented with the input "John took the ball", paired with the meaning:

(1) DECL(PAST(CAUSE(**John**,
BECOME(DEF(**ball**), **take**))))

Since the N(*the*) is empty, the sole PSA for this input sentence (which includes **John**, CAUSE and **ball**) would be discarded given Rule 1. The algorithm then determines the minimum number of words to be updated in the lexicon, in this case, only the word *the*:

	N	P
John	{**John**}	**John**, **ball**
took	{CAUSE}	CAUSE, WANT, BECOME, **take**, PAST
the	{}	WANT, **arm**, DEF, DECL, PAST, CAUSE, **John**, BECOME, **ball**, **take**
ball	{**ball**}	**ball**, **arm**

Given the new lexicon and assuming the same input, Rule 1 would not filter out its PSA. Now, another inference becomes possible, captured by Rule 2: since the utterance meaning does not contain the symbols WANT and **arm**, they can be excluded from the P sets of the relevant lexical items. After this, a comparison between the P sets of the words is possible: exclusive symbols in the P sets of the words can be copied to their respective N sets, a task carried on by Rule 3. The updated lexicon shows the following configuration:

	N	P
John	{**John**}	**John**, **ball**
took	{CAUSE}	CAUSE, BECOME, **take**, PAST
the	{DEF}	DEF, DECL, PAST, CAUSE, **John**, BECOME, **ball**, **take**
ball	{**ball**}	**ball**

The fourth heuristic compares the necessary symbol sets of the utterance words. In the example, it will detect that **ball** and **John** appear (each) only once in the utterance meaning and that both are, respectively, in N(*ball*) and N(*John*). Thus, the conceptual symbol *ball* can be removed from P(*john*) and P(*the*), as shown below:

	N	P
John	{**John**}	**John**
took	{CAUSE}	CAUSE, BECOME, **take**, PAST
the	{DEF}	DEF, DECL, PAST, CAUSE, BECOME, **take**
ball	{**ball**}	**ball**

Some more input is necessary for a complete convergence. Suppose, now, that the learner receives the utterance "The kids" paired with DEF(**kids**). By applying Rules 1 to 4, the following updated lexicon would be obtained (the entry for *kids* is omitted):

	N	P
John	{**John**}	**John**
took	{CAUSE}	CAUSE, BECOME, **take**, PAST
the	{DEF}	DEF
ball	{**ball**}	**ball**

Note that *the* has totally converged. If exposed again to the utterance (1), Rules 1 to 4 would take the learner to the final partial state below:

	N	P
John	{**John**}	**John**
took	{PAST, CAUSE, BECOME, take}	PAST, CAUSE, BECOME, **take**
the	{DEF}	DEF
ball	{**ball**}	**ball**

The learner is ready for what Siskind (1996) calls "stage two": once the relevant conceptual symbols were discovered, a structured meaning is calculated for words that have more than one conceptual symbol. In the present model, instead, each sense starts with all possible valid subexpressions extracted from the utterance meaning as its D set. During the learning process, Rule 5 will remove all expressions that lack the necessary conceptual symbols of a sense. At the end of this process, only one subexpression should remain. This approach is simpler than the original calculations although it is not clear which one can be considered more plausible.

3 Simulations

Simulations were conducted for five corpora. Generation was controlled for the MLUw of each corpus (Parker and Brorson, 2005, for details) and for the distribution of types of utterances, as explained before.

3.1 Corpora

Table 4 summarizes the characteristics of the corpora used in the simulations. The MLUw measure takes the *mean number of words* instead of morphemes. The two measures are argued to be almost perfectly correlated (Parker and Brorson, 2005).

Corpora	Utter.	Words	MLUw	Lex.
Development	985	3065	3.11	52
"Head-final"	2071	10347	5.00	56
English	40863	245111	6.00	91
BP I	100000	575449	5.75	133
BP II	100000	577349	5.77	464

Table 4: Corpora used in the simulation.

Each corpus in the table above was conceived with a specific purpose. The "development" corpus was manually built in order to make learning easier and faster, such that the overall functioning of the model could be observed given a very favorable input. The vocabulary was smaller, as well as its MLUw, utterances were ordered from the simplest to the most complex and were also ordered to provide strong contrasts, making heuristics more effective.

The other corpora were all generated automatically. The "head-final" corpus also has a small vocabulary and had the intent of increasing the difficulty in the lexical acquisition task by eliminating the artificial simplicity and ordering of data of the development corpus. Finally, the English, the BP I and the BP II corpora, being much larger than the first two, had the goal of imposing a more substantial challenge to the learner. Given the richer morphology of Brazilian Portuguese, BP I and II corpora show the largest vocabularies and number of utterances, in order to ensure a sufficient exposition to all lexical items of BP.

3.2 General results

In this study, convergence means – as in Siskind (1996) – to acquire at least one meaning by word for 95% of the lexical items. For the development corpus, the learner fully converged to the target lexicon, without false positives. Convergence was also almost complete for the head-final corpus, but the learner's performance starts to fall down for the larger corpora. For these, the learner was successful in acquiring functional words in general (determiners, prepositions, etc.), nouns, adjectives, adverbs, copulas, auxiliaries and verbs in the imperative form. It also showed some success in acquiring passive verbs. However, in general, its performance was very poor for verbs either by converging to false positives or not converging at all. False positives were deviant cases where the meaning was partially correct, but not exactly. More specific details for each corpus and its respective simulation are provided in next subsection. Table 5 summarizes the learner's performance.

	Target	Acquired		
Corpus	Lex.	Lex.	False	Conv.
Development	52	52	0	100%
Head-final	56	54	0	96,4%
English	91	87	11	95,6%
BP I	133	70	2	52,63%
BP II	464	183	1	39,43%

Table 5: Summary of lexical acquisition for each corpus.

3.3 Specific results

As expected, the development corpus made it easy for the learner to converge. It consisted of 197 utterances which were iterated five times. Given the relative simplicity of the utterances (MLUw of 3.11), these iterations were meant to simulate multiple expositions to the same utterances while artificially excluding more complex utterances that could slower the learning process by creating too many concurrent senses for each word. Instead, this corpus favors higher contrasts between words thus leading to faster learning. The first iteration had a pre-specified order, starting with simple NP fragments, followed by NP with adjuncts, and finally clauses and yes/no questions. Consequently, almost all the target lexical items were acquired in the first iteration, the remainder being acquired in the second, as Figure 1 shows.

The head-final corpus is a small English corpus to which a strict head-final ordering was imposed. Although this property is not relevant for lexical acquisition, this corpus had the goal of removing the artificial restrictions of the development corpus. Thus, the behavior of the learner could be evaluated given a slightly more complex input (which also included Wh- questions). Results, shown in Figure 2, show that in fact the learner is able to converge in the face of random exposition to data. Because of its small size, the corpus was insufficient for the learner to converge for all words.

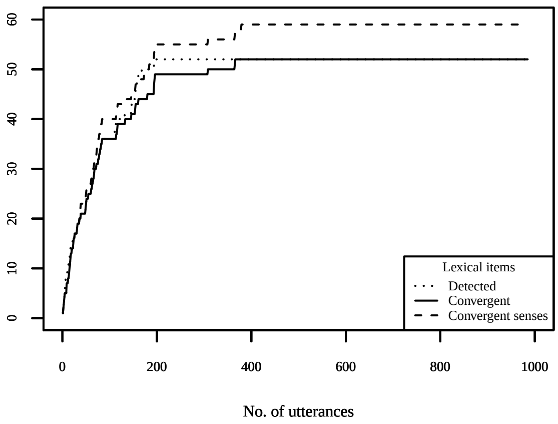

Figure 1: Lexical acquisition for the development corpus.

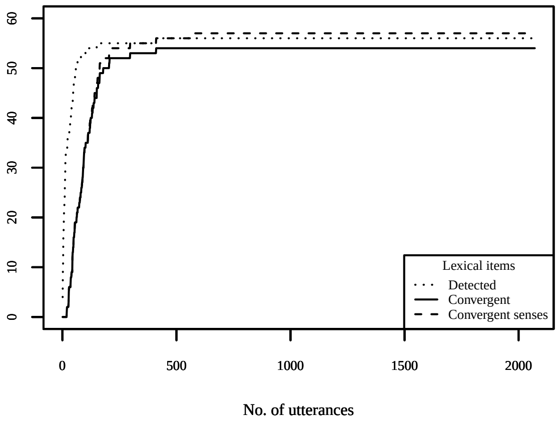

Figure 2: Lexical acquisition for the head-final corpus.

Starting with the English corpus, simulations tried to evaluate the performance of the learner given larger corpora with bigger vocabularies. The number of distinct verb stems was kept small with only two for each verb class (intransitive, unergative, etc.).

As Figure 3 shows, the learner converged almost fully, although it showed an interesting tendency of including definiteness as part of verb senses and excluding them from proper nouns. However, by inspecting the final lexical entries, it seemed possible that this tendency is temporarily

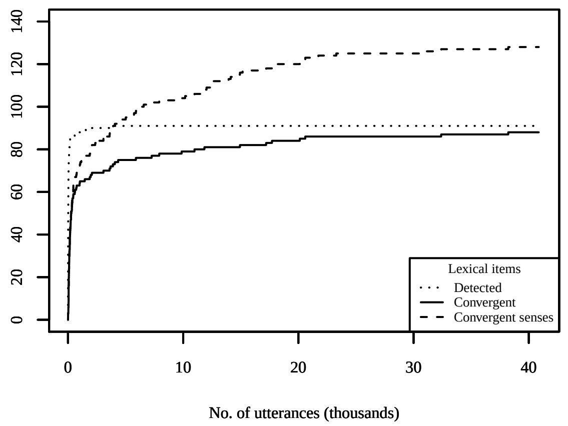

Figure 3: Lexical acquisition for the English corpus.

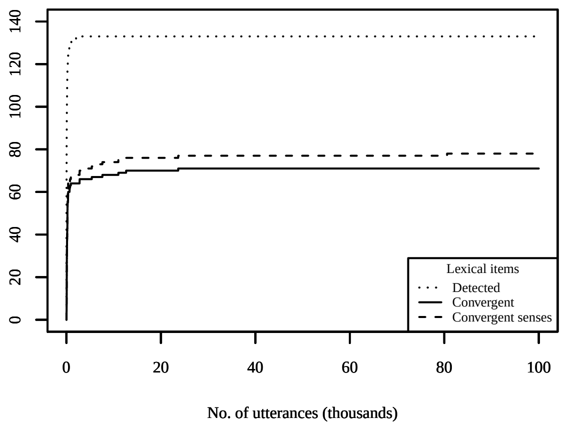

Figure 4: Lexical acquisition for the BP I corpus.

and could be overcome with more input data, as it did for some items. Related to this issue, we also see a strong tendency in this simulation for a high number of senses conjectured and converged to by the learner, as compared to other simulations. This is discussed in the next section.

The learner's performance drops drastically when exposed to Brazilian Portuguese data. The BP I corpus was also controlled for the number of verb stems, 1, by class of verbs. However, given the possible inflected verb forms of Brazilian Portuguese, the final vocabulary of BP I contained 42 more items when compared to the English corpus. As we can see in Figure 4, although the learner received more than twice the number of input utterances available in the English corpus, it acquired less words, consisting mostly of functional items, nouns, verbs in the imperative and passive forms, adjectives and adverbs. For almost all of the other inflected forms, the learner could not converge.

In the final simulation, with the BP II corpus, the learner followed the same tendency, with even lower proportional results (Figure 5). In this corpus, differently, there were more verb stems – up to 8 – per verb class.

3.4 Discussion

It was mentioned above a strong tendency, by the learner, of conjecturing and converging to a higher number of senses in the simulation for English. This is, in part, a direct consequence of the higher number of contexts in which the same word form appears with subtle meaning differences in English. However, for another part, the learner had a tendency of converging to senses close to but divergent from the target ones. For instance, the learner had a tendency of converging to verb senses which included the definiteness feature, thus, showing at least two alternative senses for the same entry, one for definite and another for non-definite contexts. Sometimes it also included another sense along with these, now without the definiteness feature, thus closer to the target. This is a curious tendency and it is not yet clear whether it is temporarily and could be overcome by more data or if it may result from some inconsistency in the input data.

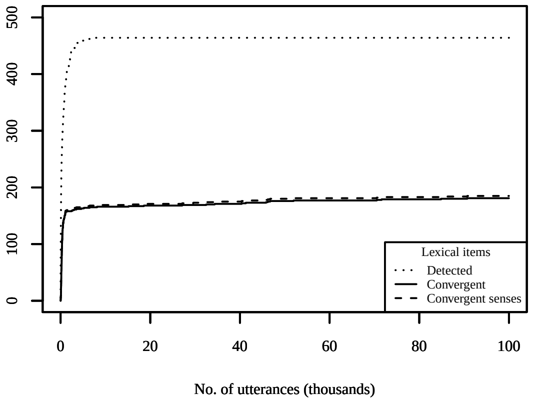

Figure 5: Lexical acquisition for the BP II corpus.

Apart from that, two main reasons seem to be involved in the learner's performances, in particular, for the lower performances for BP I and BP II corpora. First, it is possible that the learner could converge for Brazilian Portuguese if more data were available, given the sensitiveness of the heuristics to homonymy. As mentioned before, the polysemic nature of the input data in this model makes it likely that a corpus of up to a million words could be necessary for convergence. Unfortunately, technical issues prevented the learner to be exposed to such amounts of data. Thus, this can be taken as a first explanation for the learner's low performance for the BP corpora.

A second factor relates to morphological properties of the input language. Siskind's (1996) heuristics were only evaluated against English data, for which the present learner was also similarly successful. Thus, it is likely that the richer morphology of BP is causing problems to the heuristics as it leads to higher sparsity of data. It turns out that words show much lower frequencies in the BP corpora, when compared to the English corpora, as we can see in Figures 6 and 7.

As we see, there is a significant difference between frequencies for English and Brazilian Portuguese. Although they all lie below 10%, for BP the majority of the frequencies are close to zero. Consequently, occurrences of the correspondent lexical items will be dispersed through the corpus, probably distant from each other in terms of the number of utterances between them. This fact will not only make learning slower for these words, but will also lead the "garbage collection" procedure to discard non-convergent senses for these words before they have the chance to converge.

Conceived by Siskind both to discard wrong sense assignments caused by referential uncertainty and to keep the number of PSAs as low as possible (thus, increasing efficiency), the garbage collection, in cycles of 500 utterances, removes all "non-frozen" senses, that is, non-convergent ones or convergent ones that were not used successfully a predefined number of times. It is a way of having the learner "forgetting" unproductive senses. The problem is that for the BP data, given its sparseness, unfrequent words are reset again and again.

For this reason, in the simulations another strategy for garbage collection was also evaluated: instead of a cycle of 500 utterances, it assumed a cycle of 50 expositions to a given word. If a sense did not converge during the cycle, it was then discarded. However, this change did not have the desired effect. This indicates that, along with other adjustments, such simple garbage collection routines are not adequate. It is important to have in mind, nonetheless, that this model does not decompose words into morphemes. And this could be a way of overcoming the learning difficulty, since word stems would have higher frequencies and its affixes would fall into the category of functional words, for which the learner shows much better performance.

4 Conclusions

The study presented above had the goal of contributing to the understanding of lexical acquisition by children, by imposing conditions that, by assumption, can be considered as closer approximations to the ultimate complexities of the data available to the learner. As a consequence, Siskind's (1996) algorithm had to be adapted to be able to handle such input data. Two main aspects of the input are different. Informationally, more conceptual symbols are involved both to account for the meaning of functional words and to types of utterances. As a consequence, polysemy is added to the data. Morphologically, the input data shows higher sparsity – that is, words occur less frequently – caused by the various verb inflections and agreement morphology of Brazilian Portuguese.

Results indicate that both changes impose difficulties to the learning heuristics, although it is an open question whether the learner could overcome the challenge posed by polysemy if exposed to much more data. Nevertheless, sparseness seems to be more crucial to the learner's performance and it may demand a change in the "garbage collection" conceived in Siskind (1996). Another possibility, is to have the model being capable of decomposing words into stems and affixes, what by hypothesis could eliminate the problem of sparsity both by guaranteeing frequent expositions to the stems and by assigning affixes to the category of functional words for which the learner in the present study showed satisfactory performance.

Still, there are some more open issues to consider. First, although this study claims to be evaluating Siskind's (1996) heuristics, it is important to also guarantee that the implementation is at least equivalent to the original.[2] Therefore, a future

[2] In a recent work by Yu & Siskind (2013), the authors investigate a distinct approach, based on probabilistic methods, for learning "representations for word meanings from short clips paired with sentences". Given its perceptual grounding (on video clips), it covers only a toy grammar for some spacial relations and interactions. That particular study and the present one can be seen as complementary: as far as the probabilistic approach is able to model the cross-situational learning strategy successfully, studies like the present one provide knowledge about the kind of robustness the learner must have

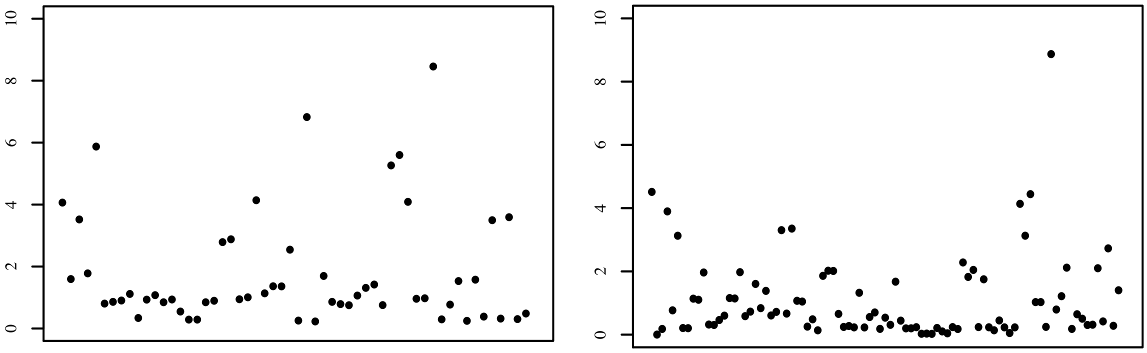

Figure 6: Word frequencies for the head-final and the English corpora.

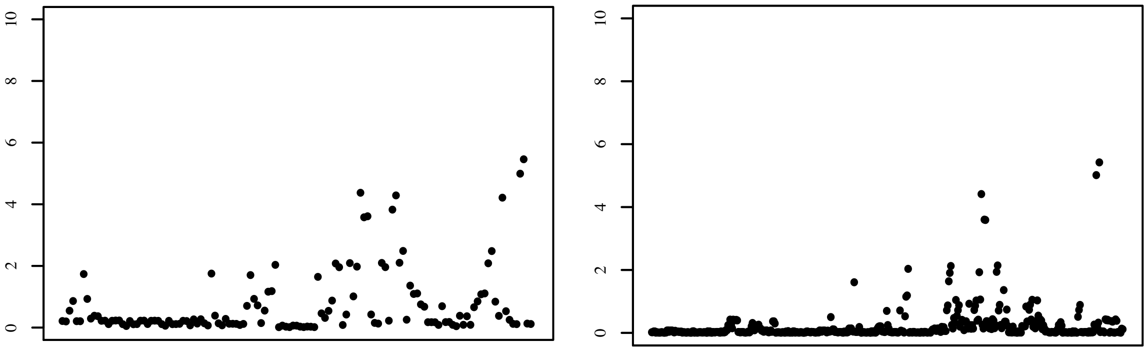

Figure 7: Word frequencies for the BP I and BP II corpora.

goal is to fully replicate Siskind's results, for all parameters (vocabulary size, rate of homonymy, etc.) involved. Such replication will not only add support to the results presented but will also make it possible to evaluate the same parameters for the kind of input data assumed here.

Apart from that, it is important to face the challenge of dealing with omitted words in utterances, such as argument omission (subject, object, etc.) and ellipsis phenomena. The present algorithm is a step in that direction as it is able to handle conceptual symbols – for instance, for utterance type – that lack morphological realization both in English and in Brazilian Portuguese. But the changes made to the original algorithm are probably not sufficient and have to be improved.

In somewhat the opposite direction, agreement morphology in languages cause the input to have two or more morphemes that share the same information. Thus, how is the algorithm to handle such cases? Certainly, it will have to allow some constrained meaning overlapping between morphemes in an utterance. However, the actual nature of the constraints needed in this case is still not clear. Adding Brazilian Portuguese to this simulation is a small but important step towards cross-linguistic coverage in this regard. Given that BP is from the family of Romance languages, being able to deal well with it makes it likely that the model will also be able to handle other languages of this family. Of course, it is important to keep adding languages from other families, specially those that show greater differences from English and BP.

Finally, although this model may be taken as reasonably plausible as a psychological model, it demands empirical support for the nature of the semantic-conceptual representation, as well as the learning heuristics, properties of the processor, etc. For all of these, it is necessary to state their empirical predictions and find ways of assessing them experimentally.

Acknowledgments

Thanks to Sao Paulo Research Foundation – FAPESP – for funding this research through grant no. 09/17172-3 and also thanks to the reviewers for their useful comments and suggestions.

in order to succeed in the face of distinct languages and more realistic grammars.

References

Robert C. Berwick. 1985. *The Acquisition of Syntactic Knowledge*. The MIT Press, Massachusetts.

Thea Cameron-Faulkner, Elena Lieven, and Michael Tomasello. 2003. A construction based analysis of child directed speech. *Cognitive Science*, 27:843—873.

Pablo Faria. 2013. *Um modelo computacional de aquisição de primeira língua*. Phd dissertation, University of Campinas (UNICAMP), Campinas, SP, Brasil, November.

Cynthia Fisher, D. Geoffrey Hall, Susan Rakowitz, and Lila Gleitman. 1994. When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, 92:333–375.

Helen L. Gaylard. 1995. *Phrase Structure in a Computational Model of Child Language Acquisition*. Ph.D. thesis, University of Birmingham, March.

Erika Hoff-Ginsberg. 1986. Function and structure in maternal speech: Their relation to the child's development of syntax. *Developmental Psychology*, 22(2):155–163.

Matthew D. Parker and Kent Brorson. 2005. A comparative study between mean length of utterance in morphemes (mlum) and mean length of utterance in words (mluw). *First Language*, 25(3):365–376.

Lisa Pearl. 2010. Using computational modeling in language acquisition research. In E. Blom and S. Unsworth, editors, *Experimental Methods in Language Acquisition Research*. John Benjamins.

Steven Pinker. 1989. *Learnability and cognition: The acquisition of argument structure*. MIT press, Cambridge, Massachusetts.

Jeffrey M. Siskind. 1996. A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61(1):39–91.

Aline Villavicencio. 2002. *The acquisition of a unification-based generalised categorial grammar*. Doctoral dissertation, University of Cambridge.

Charles Yang. 2011. Computational models of syntactic acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*.

Haonan Yu and Jeffrey M. Siskind. 2013. Grounded language learning from video described with sentences. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 53–63.