Title

Learning Cross-linguistic Word Classes through Developmental DistributionalAnalysis

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors

Freudenthal, Daniel
Gobet, Fernand
Pine, Julian M.

Publication Date

2019

Peer reviewed

Learning Cross-linguistic Word Classes through Developmental Distributional Analysis

Daniel Freudenthal, Fernand Gobet, Julian M. Pine
Department of Psychological Sciences, University of Liverpool

Abstract

In this paper, we examine the success of developmental distributional analysis in English, German and Dutch. We embed the mechanism for distributional analysis within an existing model of language acquisition (MOSAIC) that encodes increasingly long utterances, and compare results against a measure of 'noun richness' in child speech. We show that, cross-linguistically, the mechanism's success in building an early noun class is inversely related to the complexity of the determiner and noun gender system, and that merging of determiners gives very similar results across languages. These results suggest that children may represent grammatical categories at multiple levels of abstraction that reflect both the larger category as well as its finer structure.

Keywords: language acquisition; cross-linguistic; distributional analysis.

Introduction

A major question in the study of language acquisition is how children acquire grammatical categories such as noun and verb. One source of information that children might draw on in this process is distributional information – nouns and verbs tend to occur in different lexical contexts (i.e. are preceded and followed by different sets of words). An influential approach to the learning of word classes through distributional analysis is that of Redington, Chater and Finch (1998) who show that it is possible to accurately cluster words into syntactic categories on the basis of the distribution of a small set of high frequency words that precede and follow them. The basic ideas behind distributional analysis have been employed and adapted by several other authors and, in some cases, applied to other languages (Frank et al., 2013; Mintz, 2003; Keibel, 2005).

However, a major weakness of many studies of distributional analysis is that, while they explore mechanisms that are thought to operate in language-learning children, they make limited contact with the developmental literature and child data. Thus, the focus tends to be on building large word classes with high accuracy. As a result, distributional analysis is often carried out on large corpora of complete utterances and hence ignores the developmental fact that most of children's early utterances are just one or two words long.

Freudenthal et al. (2016a, b) aimed to develop a more plausible mechanism by 1. gradually expanding the contexts available to the mechanism in a developmentally plausible way, and 2. simulating actual child data. Freudenthal et al. do this in the context of MOSAIC (Freudenthal et al. 2007, 2015), a computational model that has been used to simulate a range of phenomena in language acquisition. The key learning constraint in MOSAIC is an utterance-final bias: MOSAIC builds up the representation of the input it is trained on in a right-to-left manner. This feature interacts with the statistics and structure of the input language and is responsible for MOSAIC's successful simulation of (amongst others) cross-linguistic differences in the rates at which children produce Optional Infinitive errors.

As MOSAIC sees more input, it represents longer (utterance-final) phrases and thus has more contexts available for distributional analysis. Freudenthal et al. show that, in English, their developmental version of distributional analysis initially tends to link together nouns (which tend to occur in utterance-final position), a finding that is consistent with the claim that children acquiring English form a productive noun category earlier than they form a productive verb category (Akhtar & Tomasello, 1997; Olguin & Tomasello, 1999; Tomasello & Olguin, 1993).

Freudenthal et al. (2016b) also show that MOSAIC builds an initial noun class that is sufficiently large to simulate the rate of noun use in early child speech. Introducing a measure of *noun richness* – the ratio of the number of nouns over the number of nouns plus main verbs – they show that this ratio is considerably higher in early child speech than in child-directed speech. Simulations with MOSAIC show that roughly half of this difference can be explained through high noun richness in the utterance-final phrases in the model's output. Productive use (i.e. substitution) of distributionally similar words was sufficient to raise noun richness in MOSAIC's output to levels near those found in English-speaking children.

Taken together, these results show that it is possible to perform a developmentally plausible distributional analysis and use it to simulate actual child data, and thus greatly enhance the psychological plausibility of the approach. However, Freudenthal et al. (2016b) only apply their mechanism to English, a language that has a relatively fixed word order and is morphologically impoverished, two features that are likely to benefit distributional analysis.

The main aim of this paper is to extend this developmental distributional analysis to German and Dutch, two languages that have more variable word order, and are morphologically more complex (in particular, through their use of gender and case). Our main focus will be on how comparable the results of distributional analysis are, and how well they fit child noun richness scores in the three languages. In particular, we will focus on the complexity of the determiner and noun gender system. Incorporating the analyses within a computational model that learns progressively longer sequences also allows us to gradually expand the contexts available for distributional analysis and investigate how this interacts with the word orders of the three languages.

Typology of German and Dutch

German, Dutch and English differ in a number of ways that are relevant for the current analyses. Typologically, the main difference is that English is an SVO language, while German and Dutch are SOV/V2 languages where verb position is dependent on finiteness – finite forms take second position (see utterances 1a, 1b and 1c) whilst nonfinite forms take final position (see utterances 2a, 2b and 2c).

1a. I eat a cookie (E)
1b. Ich esse ein Keks (G - I eat a cookie)
1c. Ik eet een koekje (D - I eat a cookie)

2a. I want to eat a cookie.
2b. Ich moechte ein Keks essen (G - I want a cookie eat)
2c. Ik wil een koekje eten (D - I want a cookie eat)

3a. Do you want a cookie?
3b. Willst du ein Keks? (G – Want you a cookie)
3c. Wil je een koekje? (D – Want you a cookie)

English and German/Dutch also differ in terms of question formation (see utterances 3a, 3b and 3c). Where English forms (polar) interrogatives through the use of dummy modal *do*, German and Dutch use (main) verb inversion. These features mean that German and Dutch have a more variable word order, which may impact on the general success of distributional analysis. The verb-final feature may result in lower numbers of nouns occurring in utterance-final position. This in turn may affect the early construction of a noun class through distributional analysis. However, it also raises the possibility that German and Dutch children may show lower levels of noun richness than English children. A similar claim has been made for children learning languages such as Mandarin Chinese and Korean (Choi & Gopnik, 1995)

Table 1: Case marking in German

	Nom.	Gen.	Dat.	Acc.
Masc.	ein/der	eines/des	einem/dem	einen/den
Fem.	eine/die	einer/der	einer/der	eine/die
Neut.	ein/das	eines/des	einem/dem	ein/das
Plural	--/die	--/der	--/den	--/die

A second way in which the three languages differ is in their use of noun gender and case. English has neither gender nor case (except on personal pronouns). German has three genders and marks case on articles and adjectives. Dutch is like German in that it has gender, but is like English in that it does not mark case. Table 1 illustrates the Gender/Case system of German, for the definite and indefinite article. German gender extends to demonstratives, possessives and quantifiers.

Standard Dutch distinguishes two genders[1] (common and neuter), which take the same indefinite article (*een*), but differ in the definite article (*de*/*het*). Gender is marked on adjectives by the addition/omission of an *-e* suffix. This suffix is applied to all adjectives preceding common gender nouns. For neuter nouns it is applied to adjectives following the definite, but not the indefinite article. Dutch gender extends to demonstratives (but not possessives and quantifiers).

One of the consequences of the different case and gender systems of the three languages is that the degree of lexical variation in the position preceding nouns is largest for German, intermediate for Dutch and lowest for English, Construction of a noun category through distributional analysis is therefore likely to be least constrained in English and most constrained in German. However, while gender may hinder the learning of a noun class, it marks a distinction that children need to acquire, and, since it has very little (transparent) semantic or phonological basis[2], it is very likely to be one that has to be learned distributionally. We will examine how the complexity of the determiner system affects the learning of both the overall noun class as well as the finer gender classes. We will first perform a distributional analysis whilst differentiating between all determiners, and then compare the results with an analysis in which we conflate case and gender by merging the different forms of determiners. Keibel (2005) has previously shown that merging determiners in this way is beneficial for learning the German noun category.

Corpora used

A challenge in cross-linguistic research involving corpora of child-directed speech (CDS) is that of ensuring comparability. The number of corpora available is limited and they differ in terms of size, recording situations, age range of the target children and availability of morphological information. We aimed to select from CHILDES a set of corpora for each language that were as comparable as possible in terms of their overall size. For English we selected the 6 largest sub-corpora from the Manchester corpus (Theakston et al., 2001). The Manchester corpus contains corpora for 12 individual children, and contains part-of-speech information for child and adult speech on the morphology (MOR:) tier. The selected corpora typically contained 30,000-35,000 utterances of child-directed speech

[1] Though vestiges of a third gender remain.

[2] There actually are a number of phonological, morphological, and semantic cues to German gender. MacWhinney et al. (1989) show that a neural network trained on 38 of these cues can correctly classify held out nouns. However, since Macwhinney et al.'s model learns in a supervised manner, gender information is actually available to the model. Since gender is essentially defined distributionally, lexical contexts appear a more potent cue to identifying a noun's gender.

per child. For German we selected the Rigol corpus, consisting of 4 children with roughly 45,000 child-directed utterances per child. After limited cleaning up of the corpus, we were able to run the CLAN mor facility, which was able to assign part-of-speech information to ~99% of all word tokens in the corpus. For Dutch, we selected the two children from the Van Kampen corpus. These corpora contain 65,000 and 25,000 maternal utterances. Since there is currently no functioning mor-grammar for Dutch, we assigned to the words in these corpora the most common part of speech derived from the Treetagger (Schmid, 1994).

Study 1: Child Noun Richness

The first analysis concerned children's cross-linguistic use of nouns and (main) verbs[3]. All the corpora used consist of multiple recordings (tapes) at different child ages. For all corpora we counted the number of nouns and verbs in child and adult speech on a tape-by-tape basis, and plotted noun richness (i. e. #nouns / (#nouns + #verbs)) relative to the child's Mean Length of Utterance (MLU) for the relevant tape. In line with current practice in MOSAIC, analysis was performed on utterance types. Figure 1 shows the trendlines for the scatterplot of English, Dutch and German child, and child-directed speech. For clarity, individual data points are not plotted. As can be seen, noun richness scores look remarkably similar across the three languages. While German child noun richness is (initially) lower than it is for Dutch and English, it is considerably higher than it is for adults, and thus suggests that, cross-linguistically, children are equally productive around nouns in the early stages.

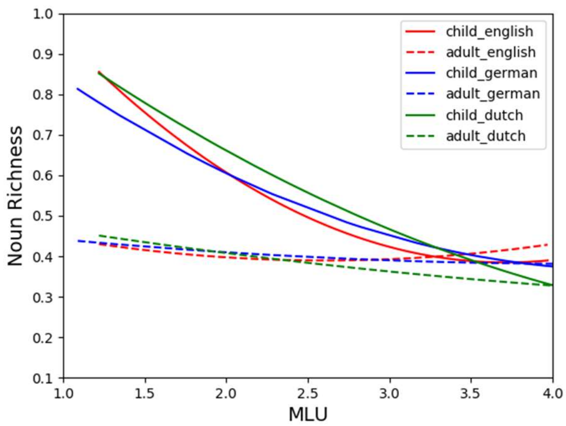

Fig. 1: Noun Richness in English, German and Dutch.

Study 2: Simulations with MOSAIC

Training MOSAIC models

MOSAIC learns from orthographically-transcribed child-directed speech and generates as output corpora of speech that can be directly compared to child speech. Learning in MOSAIC is slow, and takes place by feeding input through the model multiple times. With each exposure, MOSAIC represents more and longer (utterance-final) phrases and is thus capable of producing more and longer output, as is true of developing children. A detailed description of MOSAIC and how it is trained is provided in Freudenthal et al. (2015).

For the current analyses, we performed a distributional analysis at several points in the models' development. Here, we report results from selected runs between 36 and 50.[4] Over this range, the MLU of the utterances represented in MOSAIC (and hence of its output corpus) increases from roughly 2 to 5 words – and thus increasingly approximates corpus-wide statistics. The key consideration here is that, early in training, MOSAIC represents short utterance-final phrases that extend further to the left with increased training. This feature, which is responsible for MOSAIC's successful simulation of a number of phenomena in child speech, has the potential to interact with word order in shaping the cross-linguistic results of the distributional analysis.

The distributional analysis

The distributional analysis was carried out in the same manner for all languages. The target words were the 1,000 most frequent words for a given corpus, and the context words the 150 most frequent words. Utterance endings were also included as contextual elements. At each point in training, we searched the phrases represented in MOSAIC for the target words, and noted how often the context words occurred in the preceding and following position. Thus, for each target word, we generated two vectors that contained the counts for the context words in preceding and following position. Similarity between words was expressed as the similarity between these vectors, and two words were considered to be of the same class if their similarity exceeded a threshold value for both preceding and following position. For the current analyses we expressed similarity in both a non-parametric and a parametric way. We used a Spearman rank-order correlation, as well as cosine similarity based on the square root of the vector counts[5]. Freudenthal et al. (2013) have shown that (for English) a parametric measure is better for classifying nouns, while the rank order is better for classifying verbs. In the current analyses, the rank-order correlation gives better results when applied to English, while the parametric gives better results for German. This finding is in line with reports by Redington et al. (1998) for English and Keibel (2005) for German. Importantly, however, the two

[3] For Dutch and German, modal verbs were included as these can be used as main verbs. Copulas were excluded for all languages.

[4] Run number reflects number of exposures to the input.

[5] This is a departure from Freudenthal et al. (2016b) who used a distance measure that discarded frequency as well as counts from interrogative contexts.

measures give qualitatively similar results when used in isolation, but better quantitative results when combined.

Results

Unmerged determiners

Results for the distributional analysis are reported in Table 2, which shows the number of linked words, overall accuracy (proportion of same class links), noun richness (ratio of noun-noun to noun-noun plus verb-verb links) as well as numbers of links and accuracy for verbs and nouns. Two words were considered to be of the same word class if their rank-order correlation in preceding and following position exceeded 0.40, or their cosine similarity exceeded 0.65.

As can be seen in Table 2, the distributional analysis in English results in an early noun class, with verbs being classified later in development. This pattern is consistent with children showing early productivity around nouns and late emergence of a productive verb class (Akhtar & Tomasello, 1997; Olguin & Tomasello, 1993; Tomasello & Olguin, 1993).

It is also apparent from Table 2 that the mechanism is capable of classifying words with high accuracy, particularly for nouns, but also for verbs (in the later stages). Table 2 also shows that results for the Dutch distributional analysis are similar to those for English, though the mechanism is less successful in linking nouns, and is less accurate overall. German results mirror those from Dutch, but the distributional analysis is even less successful at building a noun class. Thus, the models never exceed 1000 noun links, even in the later stages. Across runs, the German noun class is approximately a quarter of the size of the English noun class.

The results from Table 2 thus suggest that the less constrained word order in Dutch and German leads to lower overall accuracy, but also that the size of the noun class is inversely related to the complexity of the determiner system. This pattern is not surprising, but it appears to be in conflict with the child noun-richness data from Fig. 1, which suggest that children from all three languages are equally productive around nouns. It also suggests that German and (to a lesser extent) Dutch MOSAIC models may struggle to simulate early child noun richness scores[6].

Merged determiners

We examined whether German and Dutch gender and case hamper the construction of a noun category by merging determiners into one lexical item, and adding their respective counts. For German, this meant that all 6 forms of the definite article were merged, as were all 6 forms of the indefinite article (thus maintaining the distinction between the definite and indefinite article). For Dutch, we merged both forms of the definite article. Since there is only one form of the indefinite article, this cannot be merged. Results for the distributional analysis with merged determiners are shown in Table 3.

Table 2: Results of Distributional Analysis for English, Dutch and German

Run	Links	Overall accuracy	Noun-richness	Nouns	Verbs	Noun-accuracy	Verb-accuracy
English							
36	1,641	0.80	0.94	1,218	70	0.83	0.42
38	2,215	0.80	0.91	1,553	153	0.83	0.52
40	3,037	0.83	0.89	2,230	237	0.85	0.63
44	4,144	0.90	0.86	3,164	437	0.91	0.81
50	4,576	0.91	0.83	3,375	615	0.92	0.87
Dutch							
36	1,140	0.73	0.95	774	34	0.77	0.23
38	2,030	0.78	0.96	1,467	62	0.80	0.38
40	2,995	0.81	0.96	2,260	90	0.82	0.43
44	3,496	0.85	0.91	2,582	256	0.85	0.75
50	3,310	0.84	0.80	2,122	502	0.84	0.86
German							
36	841	0.52	0.93	282	20	0.54	0.27
38	935	0.61	0.89	383	43	0.64	0.47
40	1,227	0.71	0.87	581	86	0.71	0.61
44	1,985	0.78	0.78	905	253	0.80	0.84
50	2,563	0.79	0.52	754	697	0.83	0.89

[6] Note, though, that since (unlike Freudenthal et al., 2016b) we do not currently generate output from MOSAIC, we cannot directly relate the size of the noun class to child noun richness scores.

Table 3: Results of Distributional Analysis with merged determiners for Dutch and German.

Run	Links	Overall accuracy	Noun-richness	Nouns	Verbs	Noun-accuracy	Verb-accuracy
Dutch							
36	1,515	0.70	0.96	997	37	0.73	0.17
38	2,749	0.76	0.96	1,955	70	0.78	0.30
40	4,140	0.80	0.96	3,134	104	0.81	0.36
44	5,151	0.84	0.93	3,940	292	0.86	0.65
50	4,788	0.84	0.84	3,290	573	0.85	0.80
German							
36	2,091	0.49	0.97	836	27	0.51	0.15
38	2,399	0.56	0.95	1,095	58	0.57	0.26
40	3,543	0.65	0.94	1,914	131	0.65	0.40
44	5,992	0.73	0.91	3,540	364	0.74	0.73
50	6,287	0.76	0.80	3,226	816	0.77	0.84

It is evident from Table 3 that the merging of determiners results in an increase in the number of nouns that get linked for both languages, but that this increase is considerably larger for German (by a factor of 4) than it is for Dutch (by a factor of 0.4). It is also obvious that the overall results for Dutch and German are now quite similar to the results of the English analysis (though overall accuracy scores are still lower for Dutch and German), and more in line with the cross-linguistic child noun richness scores (see Fig. 1).

Taken together, these results suggest that gender and case are detrimental to learning a noun category through distributional analysis. However, if children are able to ignore the identity of determiners, distributional analysis yields remarkably similar results across the three languages, despite their differences in word order.

Learning gender subclasses

The fact that gender (and case) hamper the learning of a noun category is not surprising since gender divides the noun category into a number of subcategories that differ in their distributional characteristics. A relevant question therefore is to what extent maintaining the distinction between the different determiners allows the mechanism to distinguish (and hence children to learn) the different noun genders. This was investigated by taking the noun-noun links from Table 2, and determining to what extent these involved nouns from the same and different genders. Results (confusion matrices) from run 50 are shown in Tables 4 (German) and 5 (Dutch).

Comparison of Tables 4 and 5 reveals that the distributional analysis is remarkably good at distinguishing the German gender subcategories, at least for the singular genders. At one level, this is not surprising since merging the determiners increases the size of the German noun class four-fold. However, inspection of the actual forms of the German determiners (see Table 1) shows that 6 different forms of each determiner are used in a paradigm containing 16 cells. Most determiners therefore occur with nouns of different genders, suggesting that the German genders are quite confusable.

Table 4: German Gender Confusion Matrix (run 50)

	Masc.	Fem.	Neut.	Pl.
Masc.	216	15	39	5
Fem.	15	198	0	18
Neut.	39	0	203	2
Pl.	5	18	2	25

Table 5: Dutch Gender Confusion Matrix (run 50)

	Common	Neuter	Plural
Common	1415	187	102
Neuter	187	249	10
Plural	102	10	17

Table 4 shows that the distributional analysis is far less successful in Dutch, with many neuter and plural nouns being linked to common gender nouns. This is caused by the fact that Dutch gender is marked on the definite, but not on the indefinite article. The Dutch noun genders are thus distributionally more similar, and far more confusable than the German noun genders. Since there are few cues to grammatical gender other than distributional information, these results suggest that acquisition of gender may be more challenging for Dutch- than for German-learning children.

Conclusions

The main conclusion to be drawn from the analyses reported here is that they provide strong support for the viability of distributional analysis. Thus, we show that it is possible to obtain plausible (and very similar) results across three different languages that differ in their word order as well as the detail of their gender and case system. Importantly, we do so using a fixed set of parameters, and in the context of a computational model that gradually expands the contexts available to the mechanism – allowing us to investigate how the increasing length of utterances that children represent may affect their word class learning. Moreover, by comparing the results to actual child data (noun richness), we were able to evaluate the relative size of the (early) noun class across the three languages.

However, it is also clear that the successful construction of a noun category depends critically on the complexity of the determiner system, and hence on how determiners are treated. If the identity of the German determiner is maintained, distributional analysis results in a noun class that is very small compared to Dutch and English, but that distinguishes between the different genders quite successfully. Merging the determiners brings the size of the verb class more in line with English and Dutch, but necessarily conflates the different genders. This effect is less pronounced in Dutch. However, the finer-grained structure of Dutch gender is distributionally less well-defined, and thus suggests that it may be more difficult to acquire for language-learning children.

The German (and Dutch) results thus suggest that grammatical categories need to be represented at different levels of abstraction that reflect both their more general properties as well as their finer-grained structure. The suggestion that children may represent both 'merged' and 'unmerged' determiners may seem surprising since one of the key characteristics of children's early speech is the fact that it lacks closed-class items like determiners. However, there is actually considerable evidence that children represent more of the closed class items than they produce.

Lew-Williams and Fernald (2007) show that Spanish three-year-olds in a looking-while-listening task can use the identity of (gendered) determiners to orient towards a target of the relevant gender. Similar findings have been reported for 24-month-old children in French (van Heugten & Shi, 2007), a language where, like Spanish, the determiner is fully predictive of the gender of the noun. Interestingly, children learning Dutch appear delayed relative to French children in this task (van Heugten & Johnson, 2011), thus providing support for the notion that the relatively poor separation of Dutch gender found in the current analyses may make it particularly hard to acquire. Studies on German (Hőhle et al., 2004) also show that children as young as 16 months (but not 12 months), can distinguish between novel words used in a nominal vs. verbal context after being habituated with a determiner-novel word sequence, but not after a pronoun-novel word sequence. These results suggest that children can use determiners to classify nouns from a very young age, but equally that they can use gender information in the on-line processing of speech, at least in languages where determiners reliably predict gender.

Taken together, the results also highlight the strengths of our approach. By embedding distributional analysis within an existing model of language acquisition that simulates children's increasing MLU, applying it to three different languages, and comparing it to actual child data, we were able to investigate how word order and the complexity of the determiner system affect the formation of an early noun class, as well as the potential implications this has for children's representations of closed class items.

Acknowledgements

Daniel Freudenthal, Julian Pine, and Fernand Gobet are members of the International Centre for Language and Communicative Development (LuCiD) at the University of Liverpool, for which the support of the Economic and Social Research Council [ES/L008955/1] is gratefully acknowledged.

References

Akhtar, N., & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. *Developmental Psychology, 33*, 952-965.

Choi, S. & Gopnik, A. (1995). Early acquisition of verbs in Korean, a cross-linguistic study. *Journal of Child Language*, *22*, 497-529.

Frank, S., Goldwater, S. & Keller, F. (2013). Adding sentence types to a model of syntactic category acquisition. *TopiCS in Cognitive Science* 5 (3), pp. 495–521.

Freudenthal, D., Pine, J. M., Aguado-Orea, J. & Gobet, F. (2007). Modelling the developmental patterning of finiteness marking in English, Dutch, German and Spanish using MOSAIC. *Cognitive Science*, *31*, 311-341.

Freudenthal, D., Pine, J.M., Jones, G. & Gobet, F. (2013): Frequent frames, flexible frames and the noun-verb asymmetry. In: M. Knauf, M. Pauen, N. Sebanz E I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the Cognitive Science Society*. (pp. 2327-2332). Austin, TX: Cognitive Science Society.

Freudenthal, D., Pine, J.M., Jones, G. & Gobet. F. (2015). Simulating the cross-linguistic pattern of Optional Infinitive errors in children's declaratives and Wh-questions. *Cognition*, *143*, 61-76.

Freudenthal, D., Pine, J. M., Jones, G., & Gobet, F. (2016a). Developmentally plausible learning of word categories from distributional statistics. In 38th *Annual Conference of the Cognitive Science Society* (pp. 674-679). Philadelphia.

Freudenthal, D., Pine, J. M., Jones, G., & Gobet, F. (2016b). Simulating Developmental Changes in Noun Richness through Performance-limited Distributional Analysis. In *38th Annual Conference of the Cognitive Science* Society (pp. 602-607). Philadelphia.

Hőhle, B., Weissenborn, J. Kiefer, D., Schultz, A. & Schmitz, M. (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy,* 5, 341-353.

Keibel, J.H. (2005). Distributional patterns in German child-directed speech and their usefulness for acquiring lexical categories – A case study. Unpublished Doctoral Dissertation. Freiburg, Germany.

Lew-Williams, C. & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical Gender in spoken word recognition. *Psychological Science*, *18*, 193-198.

MacWhinney, B. (2000). *The CHILDES project: Tools for analysing talk (3rd Edition)*. Mahwah, NJ: Erlbaum.

MacWhinney, B, Leinbach, J., Taraban, R., & McDonald, J. (1989). Language Learning: Cues or Rules? *Journal of Memory and Language*, *28*m 255-277.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90,* 91-117.

Olguin, R., & Tomasello, M. (1993). Twenty-five-month-old children do not have a grammatical category of verb. *Cognitive Development*, 8, 245-272.

Redington, M., Chater, N. & Finch, S. (1998). Distributional Information: A powerful cue for acquiring syntactic structures. *Cognitive Science*, 22, 425-469.

Schmid, H. (1995). Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland.

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.

Tomasello, M., & Olguin, R. (1993). Twenty-three-month-old children have a grammatical category of noun. *Cognitive Development*, *8*, 451-464.

van Heugten, M. & Johnson, E.K. (2011). Gender-marked determiners help Dutch learners' word recognition when gender information itself does not. *Journal of Child Language*, *38*, 87-100.

van Heugten, M. & Shi, R. (2009). French-learning toddlers use Gender information on determiners during word recognition. *Developmental Science*, *12*, 419-425.

van Kampen, J. (2009). The non-biological evolution of grammar: Wh-question formation in Germanic. *Biolinguistics*, 2-3, 154-185.