Evidence for a Morphological Acquisition Model from Development Data

Constantine Lignos[1], Erwin Chan[2], Charles Yang[1], Mitchell P. Marcus[1]
[1]University of Pennsylvania, [2]University of Arizona

1 Introduction

Work in morphology learning has thus far been primarily divided into two lines of research: cognitively-motivated models of morphology learning, which attempt to model human development and competency, and engineering-oriented models, which attempt to maximize application performance. In this paper we address the gap between these approaches by presenting results from applying the learning model presented in (Chan, 2008) to child-directed data and comparing its learning process to research in child language acquisition.

The first prominent computational model to address the learning of morphology in a manner aligned with research in child development was the connectionist model presented in (Rumelhart and McClelland, 1986). This model exhibits the same U-shaped learning patterns (Ervin and Miller, 1963) during training as children learning the English past tense. But while it replicates a development pattern, the model makes unpredictable errors that are unattested in child learning when trying to handle regular forms (Pinker and Prince, 1988). Rule-only theoretical models such as the Rules over Words model (Yang, 2002) avoid this problem by expressing both regulars and irregulars as rules, with irregular rules only applying to memorized lexical entries. A model such as Words and Rules (Pinker, 1999) presents a mixture of rule-based and associative models, for handling default rules and classes of irregular forms, respectively.

An important distinction to make in evaluating the appropriateness of learning models for language acquisition tasks is the level of supervision required. Unsupervised approaches to morphology learning use an unannotated corpus for input. Supervised approaches have focused on learning only a single inflection, such as the English past tense, using explicit word pairs as input such as *go/went, make/made, bake/baked*. To use a word pair-based approach in acquisition, a child learner would need to discover these words pairs from input, a process that is largely unexplored. While there is evidence that clustering-based approaches can identify sets of morphologically related words (Parkes et al., 1998; Wicentowski, 2002), little work has been done to attempt to extract the high-accuracy word pairs required by single-inflection learners. Thus while cognitively-motivated models have had success in replicating phenomena related to language acquisition, they have thus far not done so in an unsupervised setting.

Unsupervised learning of morphology is well-studied in engineering-oriented models. These models have focused on segmentation-based approaches, most commonly using simple transitional probability heuristics (Harris, 1955; Keshava and Pitler, 2006), or n-gram-based statistical models (most recently Spiegler et al., 2009, among many others). Often segmentation-based approaches rely on minimum description length-based approaches to guide the appropriate amount of segmentation (Brent et al., 1995), or organize the segmentations learned into paradigms (Goldsmith, 2001; Monson, 2008). While the use of paradigms creates what appears to be a useful organization of the learned rules, recent work questions the learnability of paradigms or item-based representations from realistic input due to the quantitative sparseness of data available to learners (Chan, 2008).

While these engineering-oriented models have been found to be capable of discovering morphological structure in a language, their relevance to language acquisition is questionable. The success of transitional-probability based models and experimental evidence in word segmentation that indicates humans are able to use transitional probabilities in artificial language learning (Saffran et al., 1996a,b) suggest that transitional probabilities could play a role in acquisition. But the load of computing transitional probabilities in an on-line fashion is high, and the performance of transitional probabilities when used alone is relatively poor (Gambell and Yang, 2006).

Bayesian approaches (Frank et al., 2008; Goldwater et al., 2006; Johnson et al., 2007; Naradowsky and Goldwater, 2009) of morphology learning retain the linguistic focus of cognitively-oriented models but differ in that they are trained in an unsupervised fashion, using a prior to guide the learner toward a desirable target. The computational load of such approaches is very high, requiring many iterations over the data as it gradually converges on the desired parameter set.

Models that rely on statistical optimization suggest that a learner would exhibit a broad, poor competence at first that varies in behavior as it converges to optimal parameters. This does not, however, align with development studies that suggest children acquire morphological inflections one at a time (Brown, 1973).

To address the gap between cognitive and computational models of morphological acquisition, we believe that a model must learn a psychologically plausible representation, demonstrate good performance on child-directed corpora, and learn in a fashion that parallels the developmental progression seen in children (see Yang, 2002). As the psychological plausibility of the model presented here has been extensively discussed in (Chan, 2008, chapter 4), in this work we demonstrate the model and algorithm's effectiveness at learning from child corpora and the correlations between its learning sequence and that of children.

2 A Model and Algorithm for the Unsupervised Learning of Rules

In this section we present an algorithm to learn the morphological rules of a language in a fully unsupervised fashion from input. We first present a model of

1. Place all words in the corpus in the Unmodeled set and create empty Base and Derived sets.
2. Until a stopping condition is met:
    (a) Count suffixes in words of the Base ∪ Unmodeled set and the Unmodeled set.
    (b) Hypothesize transforms from words in Base ∪ Unmodeled to words in Unmodeled.
    (c) Select the best transform.
    (d) Move the bases used in the transform to the Base set and the derived forms used by the transform to the Derived set.

Figure 1: The learning algorithm

rule-based morphology and then an algorithm to train the model from a corpus. Both the model and algorithm were originally presented in (Chan, 2008).

2.1 The Base and Transforms Model

The Base and Transforms model allows for the representation of morphologically related words in a generative fashion by defining a base set of words and a set of transforms that can change the base into its derived forms. Transforms are acquired in a frequency-driven learning process, but the model itself is a discrete, non-probabilistic representation.

A set of morphologically related words can be represented as a single base form and one or more derived forms that can be created from the base through transforms. For example, the word *bake* will be the base form for the derived forms *baked*, *baking*, and *bakes*. A transform is a rewrite rule applied a base to create a derived form. A transform is defined as two affixes *(s1, s2)*, where *s1* is removed from the base before concatenating *s2*. We represent a null affix as *$*.

A transform also has a corresponding word set, which is the set of base-derived pairs that the transform accounts for. The word set of each transform gives pairs of morphologically related words, but unlike in supervised models the algorithm discovers these pairs from unstructured input rather than receiving them as training data.

2.2 An Algorithm for Discovering Transforms

We now present an unsupervised algorithm to discover the suffixal morphology underlying the words of a corpus. It takes an unstructured corpus as input and returns a representation of the morphological grammar of the input corpus as represented by the Base and Transforms Model.

The algorithm is an iterative, greedy procedure. In each iteration, it selects the transform that models as many word types as possible and meets constraints for the selection of a transform. All word types in the corpus are placed in the Unmodeled set at the start of the algorithm's execution. As the algorithm acquires transforms, it moves words to the Base or Derived word sets based on the function they serve in the learned transforms. The algorithm primarily uses the number of types that suffixes and transforms represent; token frequencies of words are only used to break ties.

The operation of the algorithm requires several numerical parameters, including the maximum length of a suffix, the minimum size of a word after a suffix is removed, and thresholds for the minimum number of word types a transform can represent. Values for these parameters in the experiments reported below were set by hand using a development corpus.

An overview of the algorithm is given in Figure 1. Each word in the corpus belongs to one of three word sets at any point in execution: Base, Derived, or Unmodeled. All words begin in the Unmodeled set and are moved into Base or Derived as transforms are learned. The Base set contains the words that are used as bases of learned transforms, and the Derived set contains words that are derived forms of learned transforms. When proposing transforms, the algorithm creates word pairs whose bases are in the Base or Unmodeled sets and whose derived forms are in Unmodeled. This results in a bootstrapping mechanism that encourages the reuse of existing bases for new transforms.

A key feature of the algorithm is how it exploits Zipfian distributions of morphology, which allow the directionality of base-derived relationships to be inferred using only the frequencies of the forms in the word set of each transform. Cross-linguistic evidence for Zipfian distributions of morphology and evidence demonstrating the algorithm's correct identification of base forms are given in (Chan, 2008, chapters 4, 5).

We now present the learning loop of the algorithm in detail.

Count Suffixes. Iterate over the words in two sets, the Base ∪ Unmodeled set and the Unmodeled set, and count all of the suffixes of length 0-5 contained in each word, maintaining a separate count for the suffixes in each set. For example, the word *hopelessness* (transcribed as *HH.OW.P.L.AH.S.N.AH.S*), contains the suffixes (*-$*, *-S*, *-AH.S*, *-N.AH.S*, *-S.N.AH.S*, *-AH.S.N.AH.S*), and if it is only in the Base set those affixes would be counted toward the Base ∪ Unmodeled set's affix counts, not the Unmodeled set's. A suffix is only counted if removing it leaves a sufficiently long stem, in this case three phonemes. This length limitation exists to prevent the modeling of extremely short words that are likely closed-class or morphologically irregular words. Affixes are ranked by the number of types they appear in.

Hypothesize Transforms. Hypothesize transforms from all combinations of the top 50 affixes as *s1* and *s2*. For example, from the common English suffixes *-$*, *-Z*, and *-IH.NG* the transforms *($, Z)*, *(Z, $)*, *($, IH.NG)*, *(IH.NG, $)*, *(IH.NG, Z)*, *(Z, IH.NG)* are hypothesized. For each hypothesized transform, iterate over every word in the Base $\cup$ Unmodeled set containing $s1$ and check if the word that is the result of applying the transform to that word is in the Unmodeled set. If it is, add that base-derived pair to the word set of this transform. Transforms are ranked by the number of word pairs they account for, without regard to the frequency of the words in those pairs.

Select a Transform. The highest ranked transform is selected, provided it meets the criteria for an acceptable transform. A transform should be rejected if it appears to be modeling a relationship between two forms that should both be derived forms rather than a relationship between a base and a derived form. This relationship can be indicated by an *overlap ratio*. A transform's overlap ratio is calculated as the ratio of its *stem overlap* to *base overlap*. The base overlap is the number of base forms in the proposed transform that are base forms in the current grammar. The stem overlap is the number of base forms' stems (computed as the first four phonemes) in the proposed transform that are also stems of words in the Base set. The stem overlap is an approximation of the lexical similarity between two sets of words.

An acceptable transform must have a sufficiently low overlap ratio. A high overlap ratio implies that the bases in the transform's word set are very similar to words in the Base set, but not members of it. The likely cause is that the bases used in the transform are derived forms of words in the Base set, and thus accepting the transform would cause the Base set to include multiple inflections of the same lexical category. This is undesirable as it results in inconsistent base forms.

If the first and second ranked transforms account for the same number of types and are symmetric pairs, such as the English transforms *($, Z)* and *(Z, $)*, a tie-breaking procedure is invoked that selects the transforms whose bases are more frequent. This procedure is given in more detail in (Chan, 2008, chapter 5).

Stopping conditions. If there are no possible transforms remaining that account for five or more base/derived pairs, learning stops, as selection of any remaining transforms would only result in over-fitting the input corpus.

3 Results

To evaluate the algorithm's effectiveness as a model of language acquisition, we tested the algorithm on CHILDES corpora of English child-directed speech transcriptions (MacWhinney, 2000). Six children's corpora were used: Adam, Eve, Naomi, Nina, Peter, and Sarah. The corpora were pre-processed, removing any annotations and child utterances. Pronunciation data for all words in the

Corpus	Types	Tokens	Transforms Learned
Combined	7,174	730,328	23
Sarah	4,407	182,030	14
Adam	3,437	117,022	14
Nina	3,123	184,042	13
Peter	2,829	136,714	13
Naomi	2,511	52,760	9
Eve	1,935	57,760	9

Table 1: Type and token counts for CHILDES corpora used

corpus were obtained from CMUDICT 0.6, (Weide, 1998) the Carnegie Mellon University Pronouncing Dictionary. If multiple pronunciations were found for a word, the first pronunciation was selected. Words for which no pronunciation could be found were removed from the corpus. The token and type counts of the CHILDES corpora used are given in Table 1. As shown in Table 1, the number of transforms learned in a corpus is directly proportional to the number of types present in the corpus. We present the algorithm's output when run on a corpus that combines all children's data and when run on the corpus for each child.

3.1 Results on the combined corpus

The transforms learned when run on the combined corpus are given in Table 2 with example word pairs (in orthography), annotations for the most common morphological function, and type and token statistics for each transform. Each transform is given using ARPABET transcriptions.

The majority of the transforms correspond to common morphological rules in English, as shown by the annotations in the "Morpheme" column. Because the algorithm operates at a phonemic level, allomorphs for each morpheme such as *Z/S/AH.Z* for the noun plural are learned in multiple transforms. Also, in cases where multiple morphemes have the same phonemic representation, such as the plural and third person singular, a single transform may represent multiple morphemes. The most common regular verb inflections (plural, present progressive, past tense) are represented by seven of the first nine transforms learned.

In general, initially acquired transforms are more likely to represent linguistically reasonable morphological rules. Low type-frequency transforms are more likely to be spurious. For example, the transform *($, K)* is marked "spurious" because none of its base-derived pairs *(stay/steak, core/cork, stung/stunk, ming/mink, poor/pork)* contain morphologically related words.

Some transforms connect morphologically-related words, but do so by forming relationships between two forms that would ideally each be modeled as derived words. The transform *(T.IY, TH)*'s base-derived pairs *(forty/fourth, fifty/fifth,*

Iter.	Transform	Tokens	Types	Example	Morpheme
1	($, Z)	116591	518	trouble/troubles	Noun pl., Poss., 3P Sg.
2	($, IH.NG)	75830	284	land/landing	Present progressive
3	($, S)	105930	195	ant/ants	Noun pl., Poss., 3P Sg.
4	($, IY)	21588	100	noise/noisy	Adj. derivation, Dimin.
5	($, D)	24151	95	open/opened	Past tense
6	($, T)	25720	89	step/stepped	Past tense
7	($, ER)	45854	76	sing/singer	Agentive, Comparative
8	($, AH.Z)	11501	58	fix/fixes	Noun pl., Poss., 3P Sg.
9	($, AH.D)	34326	29	lift/lifted	Past tense
10	($, L.IY)	2836	20	bad/badly	Adverb derivation
11	($, AH.N)	6091	19	hid/hidden	Past participle
12	($, N)	1161	14	tore/torn	Past participle
13	($, AH.L)	44171	13	what/what'll	Contraction of "will"
14	($, AH)	3485	12	floor/flora	Spurious
15	(AH.N, $)	4325	8	garden/guard	Spurious
16	($, AH.S)	1116	7	fame/famous	Adj. derivation
17	($, AH.N.T)	18750	7	could/couldn't	Contraction of "not"
18	($, AH.T)	202	6	wall/wallet	Spurious
19	(AH.L, L.IY)	250	6	passable/passably	Adverb derivation
20	($, K)	8618	5	stay/steak	Spurious
21	(IY, $)	1474	5	daddy/dad	Adj. derivation, Dimin.
22	(AH.L, $)	2741	5	wiggle/wig	Spurious
23	(T.IY, TH)	95	5	ninety/ninth	Ordinal derivation

Table 2: Rules learned on English CHILDES data combined from six children

sixty/sixth, seventy/seventh, ninety/ninth) connect morphologically related words but connect two forms that should each be derived from a common base. It would be more desirable to have transforms *($, T.IY)* to represent *nine/ninety* and *($, TH)* for *nine/ninth*. A similar phenomenon occurs for the derivational rule *(AH.L, L.IY)*, where we would prefer two derivational rules: *($, AH.B.AH.L) (pass/passable)*, and *($, AH.B.L.IY) (pass/passably)*. We attribute the absence of these preferred rules to composition of the vocabulary of the small corpus.

The algorithm can also learn transforms that represent a base-derived relationship in a direction opposite than expected. The transform *(IY, $)* is learned after the more desirable *($, IY)* because its base words *(lady, monkey, daddy, lucky, puppy, Jenny)* have been placed in the Base set by other transforms (*lady/ladies, Jenny/Jenny's*) and thus cannot be derived by *(IY, $)*. This problem can be avoided by allowing words to move between the Base and Derived sets as the algorithm learns, a technique discussed in detail in (Lignos et al., 2009).

Transform	Adam	Eve	Naomi	Nina	Peter	Sarah	Mean	σ
($, Z)	1	1	1	1	1	1	1	0.00
($, IH.NG)	2	2	2	2	2	2	2	0.00
($, S)	3	3	3	3	3	3	3	0.00
($, T)	4	5	4	4	4	5	4.33	0.47
($, IY)	6	4	5	7	6	6	5.67	0.94
($, D)	7	8	6	5	5	4	5.83	1.34
($, ER)	5	6	7	8	7	7	6.67	0.94
($, AH.Z)	8	7	8	NL	8	8	7.8	0.40
($, AH.D)	9	NL	9	10	9	10	9.4	0.49
(AH.N, $)	NL	NL	NL	9	10	NL	9.5	N/A
($, AH.L)	12	9	NL	NL	11	9	10.25	1.30
($, N)	10	NL	NL	11	NL	NL	10.5	N/A
($, AH.N)	11	NL	NL	NL	NL	11	11	N/A
($, L.IY)	14	NL	NL	12	NL	12	12.67	0.94
(AH.L, $)	NL	NL	NL	13	NL	NL	13	N/A
($, AH.N.T)	13	NL	NL	NL	NL	13	13	N/A
($, K)	NL	NL	NL	NL	NL	14	14	N/A

Table 3: Order of rules on individual children's data sets in CHILDES

3.2 Results on individual children's corpora

Table 3 shows the transforms learned when the algorithm was tested on corpora for individual children. A transform is listed if it was learned from any of the six children's corpora. The order in which the transform was acquired is given for each corpus with a mean and standard deviation for the transform across all corpora. If a rule was not learned from a particular corpus, it is marked "NL." If a rule was not learned from at least half the corpora, its standard deviation is not given.

Whether a particular transform is learned from a particular corpus depends primarily on the number of word pairs that the transform can be applied to, which is largely determined by the number of word types in the corpus but is also affected by the bootstrapping effects of previous rules. Because of varying sizes of the children's corpora, the number of transforms learned before stopping varied from 9 to 14 rules, with the number of types in the corpus predicting the number of transforms learned.

Because of the small size of these corpora, not all transforms are learned across all corpora. In an acquisition scenario, it is likely there is a threshold below which morphological patterns are indistinguishable from noise in the data. Thus in the early stages of acquisition, when the learner has only been exposed to a relatively small amount of data, only a few rules can be learned. As more words are observed, rarer morphological patterns can rise above the noise and be learned, as shown in the larger number of rules learned in the combined corpus.

Morpheme	Brown Average Rank	Corresponding Transforms	Mean Transform Rank
Present progressive	2.33	($, IH.NG)	2.00
Plural	3.00	($, Z/S/AH.Z)	3.93
Possessive	6.33	($, Z/S/AH.Z)	3.93
Past regular	9.00	($, D/T/AH.D	6.52
Third person regular	9.66	($, Z/S/AH.Z)	3.93
Contractible copula	12.66	($, Z/S)	2.0
Contractible auxiliary	14.00	($, D/AH.L)	8.04

Table 4: Brown (1973) English morpheme acquisition order and corresponding transforms

In addition to learning morphological rules in an iterative manner, an accurate model of morphological acquisition would learn rules in an order similar to that of children. We compare our results to those of Brown (1973), who manually analyzed child-directed speech transcripts for Adam, Eve, and Sarah. In Table 4, we present the English acquisition sequence for regular, suffixal morphemes, as analyzed in Brown (1973). For each morpheme, we list the corresponding transforms that the algorithm learns. Because of allomorphy in many English morphemes, multiple transforms are needed to represent a single morpheme, such as the past tense, and in some cases each transform also can represent the phonemic realization of multiple morphemes, such as the transform *($, Z)*. For each transform, we give a ranking corresponding to the order in which the transform was learned across the six children's corpora. When a morpheme corresponds to a single transform, the mean rank of the transform as given in Table 3 is used. When multiple transforms map to a particular morpheme, the mean of their mean ranks is given.

Although it is difficult to perform a direct comparison, it can be seen by that using the mean ranks of transforms, the sequential order of acquisition of Present progressive, Plural, Possessive, and Past Regular is correctly predicted by the algorithm. The main inconsistency between the algorithm's and children's orders of acquisition is in the Third person regular and Contractible copula. These are acquired in two separate morphological rules by children, each of which has multiple surface phonological forms. The algorithm, however, acquires them in three separate phonological transforms, one for each allomorph.

The transforms for */Z/*, */S/*, and */AH.Z/* are learned relatively early by the algorithm (mean rank 3.93), whereas children acquire their corresponding inflections (Plural, Possessive, Past regular, and Contractible Copula) somewhat later. There is a principled reason for this discrepancy. In addition to identifying surface forms of morphemes, children are also acquiring the syntactic uses of morphemes and determining allophones. It is known in the acquisition literature (Slobin, 1973, 1986) that acquisition of homophonous morphemes is delayed since a child must sort out the different syntactic functions of morphemes, whereas acquisition of unambiguous morphemes is faster. The English allophone *Z/S/AH.Z* is three-ways ambiguous with respect to underlying function, which is why it is not acquired earlier as predicted by the algorithm. Since the algorithm only looks at the frequencies of surface phonological forms, we should not expect the order of morpheme acquisition to be the same as children for ambiguous morphemes.

Unlike children, the algorithm has no competency-performance gap and does not take the difficulties associated with identifying syntactic context into account. We would thus expect variation the algorithm's learning order and the competency demonstrated by children even if both were using the same method for rule selection. Reevaluating the algorithm with extensions to identify and merge allomorphs and generalize the applicability of rules would help provide a better estimation of the difficulty of learning individual rules, but it would not address the lack of a competency-performance gap that all discrete-representation computational models face.

4 Discussion

The algorithm's success in learning the most common rules and the correlation between its learning order and the acquisition order of children suggest that type-frequency is a good indicator of rule quality and the primary indicator of how quickly children acquire rules. This supports the accepted notion that frequencies in the input are important to learners but extends it by demonstrating that the frequency of hypotheses that explain the input, not just the frequency of symbols in the input, can effectively be used to guide learning, as suggested in (Chan, 2008; Yang, 2002). For a child learner to operate in this way, she must not only have an innate concept of rules, but also an innate metric or set of constraints to select these rules. The invariance across children suggests that these metrics may be built into the learner.

The alignment between the algorithm's and children's learning orders suggests the child learner adopts hypotheses that explain the greatest number of types in the input, progressively hypothesizing new rules to cover unmodeled data after each rule. Experiments have shown that children are capable of performing type-based computations (Gerken, 2006; Gerken and Bollt, 2008), although laboratory experiments cannot answer how such information can be used in the acquisition of an entire language.

The quality of the rules learned on such small corpora suggests that a simple, greedy approach may be all that is needed to learn the regular rules of the language. The process of hypothesizing and selecting the morphological rules of a language may be a more computationally simple process than statistical models would imply. The presented algorithm does not require tracking and reestimating a large number of parameters or probability estimates as transitional-probability based models do, instead it simply performs type-based counting. Unlike Bayesian or other statistical optimization approaches, the algorithm does not require large numbers of iterations over the data to converge on parameters; it incrementally acquires hypotheses that are interpretable at every step of learning.

While the algorithm uses statistical patterns of the input to select rules, the rules themselves are discretely represented and discretely applied. A rule applies to a particular set of words, but within that set it always applies. This stands in contrast to approaches which may learn a discrete set of rules but apply them probabilistically.

5 Conclusions

The algorithm and model presented demonstrate that a simple, intuitive algorithm can succeed at an acquisition task by taking advantage of the statistical characteristics of the expected input and using the frequency of hypotheses, not only input data, to select which hypotheses to adopt. This concept can be extended to a higher level than the simple affixal transform evaluated in this work; to learn many of the desired rules of language, more abstract rules are required. A more evolved learner would build on this by taking the rules learned from seen data and estimating their productivity such that they can be appropriately generalized to new data, as suggested in (Yang, 2005).

References

Michael R. Brent, Sreerama K. Murthy, and Andrew Lundberg. Discovering Morphemic Suffixes: A Case Study in Minimum Description Length Induction. In *Proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics*, 1995.

Roger Brown. *A First Language: The Early Stages.* Harvard University Press, Cambridge, MA, 1973.

Erwin Chan. *Structures and Distributions in Morphology Learning*. PhD thesis, University of Pennsylvania, 2008.

Susan M. Ervin and Wick R. Miller. Language Development. *Child Psychology*, pages 108–143, 1963.

Michael C. Frank, Denise Ichinco, and Joshua B. Tenenbaum. Principles of Generalization for Learning Sequential Structure in Language. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, 2008.

Timothy Gambell and Charles D. Yang. Word Segmentation: Quick but not Dirty. MS, 2006. URL `http://www.ling.upenn.edu/ycharles/papers/quick.pdf`.

LouAnn Gerken. Decisions, Decisions: Infant Language Learning when Multiple Generalizations are Possible. *Cognition*, 98(3):67–74, 2006.

LouAnn Gerken and Alex Bollt. Three Exemplars Allow at Least Some Linguistic Generalizations: Implications for Generalization Mechanisms and Constraints. *Language Learning and Development*, 4(3):228–248, 2008.

John Goldsmith. Unsupervised Learning of the Morphology of a Natural Language. *Computational Linguistics*, 27(2):153–198, 2001.

Sharon Goldwater, Thomas Griffiths, and Mark Johnson. Interpolating Between Types and Tokens by Estimating Power-Law Generators. *Advances in Neural Information Processing Systems*, 18, 2006.

Zellig S. Harris. From Phoneme to Morpheme. *Language*, pages 190–222, 1955.

Mark Johnson, Thomas Griffiths, and Sharon Goldwater. Bayesian Inference for PCFGs via Markov Chain Monte Carlo. In *Proceedings of NAACL HLT*, pages 139–146, 2007.

Samarth Keshava and Emily Pitler. A Simpler, Intuitive Approach to Morpheme Induction. In *Proceedings of 2nd Pascal Challenges Workshop*, pages 31–35, 2006.

Constantine Lignos, Erwin Chan, Mitchell P. Marcus, and Charles Yang. A Rule-Based Unsupervised Morphology Learning Framework. In *Working Notes of the 10th Workshop of the Cross-Language Evaluation Forum*, Corfu, Greece, September 30–October 2 2009. CLEF2009.

Brian MacWhinney. *The CHILDES Project: Tools for Analyzing Talk*. Lawrence Erlbaum Associates, 2000.

Christian Monson. *ParaMor: From Paradigm Structure to Natural Language Morphology Induction*. PhD thesis, Carnegie Mellon University, 2008.

Jason Naradowsky and Sharon Goldwater. Improving Morphology Induction by Learning Spelling Rules. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2009.

Cornelia H. Parkes, Alexander M. Malek, and Mitchell P. Marcus. Towards Unsupervised Extraction of Verb Paradigms from Large Corpora. In *Proceedings of the Sixth Workshop on Very Large Corpora*, Montreal, Quebec, Canda, August 15–16 1998.

Steven Pinker. *Words and Rules: The Ingredients of Language*. Harper Collins, New York, 1999.

Steven Pinker and Alan Prince. On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition. *Connections and Symbols*, pages 73–193, 1988.

David E. Rumelhart and James L. McClelland. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 2: Psychological and Biological Models*. MIT Press, Cambridge, MA, 1986.

Jenny R. Saffran, Richard N. Aslin, and Elissa L. Newport. Statistical Learning by 8-month-old Infants. *Science*, 274(5294):1926, 1996a.

Jenny R. Saffran, Elissa L. Newport, and Richard N. Aslin. Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, 35(4):606–621, 1996b.

Dan I. Slobin. Cognitive Prerequisites for the Acquisition of Grammar. *Studies of Child Language Development*, pages 173–208, 1973.

Dan I. Slobin. *The Crosslinguistic Study of Language Acquisition*. Lawrence Erlbaum, 1986.

Sebastian Spiegler, Bruno Golénia, and Peter Flach. Promodes: A Probabilistic Generative Model for Word Decomposition. In *Working Notes of the 10th Workshop of the Cross-Language Evaluation Forum*, Corfu, Greece, September 30–October 2 2009. CLEF2009.

Robert L. Weide. The Carnegie Mellon Pronouncing Dictionary [cmudict. 0.6], 1998.

Richard Wicentowski. Modeling and Learning Multilingual Inflectional Morphology in a Minimally Supervised Framework. *Johns Hopkins University Ph. D. thesis*, 2002.

Charles D. Yang. *Knowledge and Learning in Natural Language*. Oxford University Press, USA, 2002.

Charles D. Yang. On Productivity. *Linguistic Variation Yearbook*, 5(1):265–302, 2005.