Did, Made, Had, Said: Capturing Quasi-Regularity in Exceptions

Gary Lupyan (glupyan@cnbc.cmu.edu)
James L. McClelland (jlm@cnbc.cmu.edu)
Center for Neural Basis of Cognition
4400 5th Ave. Pittsburgh, PA 15207 USA

Abstract

The English past tense is a quasi-regular system, in that many of the irregular verbs share characteristics with regular items. Among high-frequency exceptions, in particular, several have the regular /d/ or /t/ ending but with either a reduction of the vowel (did, said) or a deletion of a stem consonant (had, made). Such forms suggest that many so-called irregular verbs reflect a joint influence of the systematic past-tense pattern captured in fully regular items together with a pressure to be short or simple. We adapt familiar neural network formalisms to show how such forms can arise if the phonological content of word forms are constrained (a) to support accurate communication of the word's meaning and (b) to be simple.

One of the defining and universal features of human languages is compositionality. At the phoneme level, this refers to the ability to combine phonemes into morphemes. At the word level, it refers to combining morphemes; for instance, combining the lexical-morpheme *cat* with the inflectional-morpheme *s* to produce a new token *cats*. Finally, at the sentence level, it is the ability to derive the meaning of a sentence based on the meanings of its constituents. Computational models such as those of Batali (1998) and Kirby (2000, 2002) demonstrate how compositional grammar emerges naturally in a population of agents that must establish a communicative system learnable by future populations—an obvious requirement of human languages. Given the natural emergence of compositionality, it is not surprising that a large proportion of the world's languages have come to represent features that are relatively orthogonal to other aspects of semantics, such as tense and plurality, through compositional means (Tyler, 1992; Bybee, 1985). By semantic orthogonality we mean that "past-ness" is largely independent of the meaning of any particular word, and so it is highly efficient to represent it through an invariant signal such as *-ed* (albeit conditioned slightly by the final phoneme of the word's stem). Some of the models leading to compositionality in word forms, however, predict there should be no stable irregularities within such systems (cf. Kirby, 2001). Perfectly regular systems such as the English progressive *-ing* do exist, but they are uncommon. Stable irregularities are the norm rather than the exception.

According to some, "irregular alternations are, by definition, functionless" (Greenberg, 1957, p. 65). Admittedly, not all structural complexities found in languages are functional (Bybee, Perkins, & Pagliuca, 1994), but the prevalence of irregularity and quasi-regularity begs for an explanation. The present work draws on the optimality theory approach of Burzio (2002) and the framework of minimum-description length (Rissanen & Ristad, 1994; Zemel & Hinton, 1994) to study how phonology changes over time to maintain regularity and produce morphological exceptions in response to various pressures on communication. Although many of our examples are from the English past-tense, our aim is more general: to elucidate the emergence of irregularity and quasi-regularity in compositional systems.

Burzio (2002) argues that the debate on the English past-tense has "focused on an artificially limited subdomain of English morphology" (pp. 159-160). He describes several levels within the lexicon, with the clearest division being between the so-called 'Level-1' and 'Level-2' formations. The former consists of derivations with Latinate affixes such as *-al*, *-ic*, *-ous*, and *-ive*, while the latter of those derived from by means of Germanic affixes such as *-ness*, *-less*, *-ful*, and *-hood*.

A main characteristic of Level-2 morphology is that it is productive and applies to words in phonologically predictable patterns, e.g., *small-ness*, *help-ful*, *neighbor-hood*. On the other hand, Level 1 morphology is phonologically unpredictable (e.g., *correct+ion* ⇒ *correction*, but *expect+ion* ⇒ *expectAtion*) and of limited productivity (e.g., *parent-al*, but not *student-al*; *cavern-ous*, but not *tavern-ous*; Burzio, 2002, p. 161). Inasmuch as Level-1 morphology is productive, it behaves so at the expense of predictability, so we have *problem+ic* ⇒ *problemATic*, *habit+al* ⇒ *habitUal*, and *compel+ive* ⇒ *compUlSive*. These changes, however, are not entirely arbitrary—they result from a combination of factors, the first of which is a pressure to be relatively simple and consistent with the phonology of the language. An example is the shortening in *intervene* ⇒ *intervEntion* which prevents the syllable from being too long. The second factor driving the changes is compliance with existing morphological regularities. Often this pattern is a relatively minor one, showing up in smallish clusters of items such as *ring/rang*, *sing/sang*, *sink/sank*, etc. Other times, it may be one that is reflected in large numbers of items, such as the regular past tense alternation *like/liked, love/loved, hate/hated* etc. McClelland and Patterson (2002) noted that 59% of the irregular verbs partially reflect the regular pattern in that they end in /d/ or /t/ as regular past forms do. Among these are several subtypes, including some that delete a stem consonant (*had*, *made*), some that reduce the vowel (*did*, *said*, *kept*), and some, already ending in /d/ or /t/ in the stem, that undergo no change at all. These quasi-regular formations, in particular, clearly show the hallmarks of Burzio's Level 1 phonology: They exhibit partial consistency with a pattern found in many other words, while also reflecting simplicity and consistency with general phonological regularities.

From analyzing the types of phonological changes that occur in Level 1 words, Burzio concludes "lexical sectors that are morphologically irregular tend to be phonologically regular, and vice versa" (p. 160). This suggests a functionalist account of irregularity: irregulars arise to preserve phonotactics. So, we have the phonologically regular and reduced *made* instead of the phonologically irregular *maked*, *kept* instead of *keeped*, etc. Conversely, the phonological irregularity of *baked* is compensated by the pressure to be compositional. In our view, the pressure for compositionality can be *partially* overcome by frequent words like *make*, but not rarer words like *bake*. Even though *keep* and *make* violate compositionality by undergoing a stem-change in the past-tense, their past-tenses still end on /d/ or /t/—the phonemes characteristic of the regular inflection—making these items quasi-regular and partially compositional.

Overwhelmingly, irregular verbs are frequent verbs. It has been generally assumed the reason for this is that their frequency makes them easier to remember, and so more resistant to change (Pinker, 1991). So, while the relatively rare stem *plead* is often regularized to *pleaded* instead of its original past-tense *pled*, the much-more frequent *feed* retains its irregular inflection (Bybee, 1995). However, it is also possible that increased frequency subjects tokens to more, not less, change over time. By definition, production costs are greatest for frequent words—they are the ones for which the costs have to be paid most often. We formalized this notion by adopting the framework of minimum-description length. Given the goal of communicating a message from one individual to another through a sequence of phonemes, we can express the sending/receiving cost of various messages as the sum of two terms: $Phon + Error$ where $Phon$ is the cost paid to generate/perceive the phonology, and the error is the cost of reconstructing the message from the phonology. Compositionality minimizes the *Error* term by, for instance, reliably associating an affix with a semantic feature. Deviations from this pattern increase the *Error*, but can minimize *Phon*.

Given a pressure to shorten messages while limiting the processing necessary to decode them, it is most advantageous to minimize the length of the frequent tokens. Hence, in Morse Code the two most frequent letters in English, *e* and *t* are represented by a single dot and dash, respectively. It is then no coincidence that frequent words are both short and generally phonologically simple (Bybee et al., 1994). However, a pressure to be short cannot be allowed to act unconstrained by a need to convey the intended meaning. Therefore, in our simulations the pressure for phonological change is modulated so that its influence is weak unless the semantics is correctly activated. We should note here that shortness is a gross simplification of true phonological cost. While the number and length of phonemes plays a part in phonological cost, it is only one of many factors. The Manding *nba* may be shorter than *thank you*, but English speakers, unaccustomed to the /nb/consonant cluster, clearly pay a lower phonological cost for the latter. We hope to address the complexities of this cost in future simulations.

We are not the first to note that language reflects both a pressure to be brief and to achieve adequate communication. Simulations such as Kirby's (2001) explore the emergence of stable irregularity in response to a pressure for frequent words to be short; but, in Kirby's approach, words are either fully regular or suppletions like *be/was*, *go/went* which in fact are very rare (Bybee, 1985). We suggest that the use of graded activations and distributed representation provides a framework for capturing both pressures that would account for the existence of forms like *did, made, had,* and *said* that simultaneously exhibit sensitivity to both pressures.

Simulation 1: Effect of token frequency on the emergence of quasi-regularity in a random corpus

To illustrate how this can occur in a simplified framework, we trained networks on a small, abstract, fully-compositional corpus of phonological and semantic representations. We hypothesized the pressure for phonological reduction simultaneously balanced by the pressure to learn the mapping, would act to reduce the phonology of the most frequent words, thus making them quasi-regular, while preserving the full regularity of the less-frequent tokens.

Method

Networks All simulations used three-layer feedforward networks. Being interested in the change to phonology, we needed a way for representations to change over time. We achieved this by extending backpropagation to the phonology input layer. The back-propagated δ terms, normally used to change weights, can also be used to specify how unit activations should be changed directly (Miikkulainen & Dyer, 1991). The learning rate for change to the representation was small (.01 times the connection weight learning rate), and was scaled by e^{-E}, where E is the error on the current example; this is the previously mentioned tradeoff between output accuracy and susceptibility to phonological change. The inputs were stochastically-generated binary values, e.g., an input of 0.75 corresponded to a unit being on at 1 75% of the time, and 0 otherwise. The error signal that ultimately affected the representations was generated from this binary value even though the change occurred to the real-valued stochastic values. This corresponds to the idea that a phoneme is either produced or not, and is therefore binary. At the same time, the probability of a phoneme occurring in a particular word is real-valued. The pressure for phonological reduction (the *Phon* part of the total cost) was implemented by an exponential function, $S(1 - e^{-cu})$ u being the output of the focal input unit; c was a constant set to 4, S was a strength constant set to 10). All simulations used this cost function. The backpropagation-to-representation algorithm thus acted to raise or lower the probability of phonemes being on in ways to both decrease the error (learn to map phonology to semantics), as well as to satisfy the cost function by simplifying the phonology. The learning rate for all simulations was set to 0.01 and momentum to 0.9. Varying the parameters used in these simulations affected the time-course of the learning, but did not significantly change the pattern of the reported results. The current simulation used 28-input, 60-hidden, and 20-output units. All simulations were run in a modified version of LENS (Rohde, 1999).

Materials Each input consisted of a stem "word" made up of 10-unit onset, 8-unit vowel, and 10-unit coda slots. At the start only one phoneme was active in each slot. The lexicon combined 100 stems with 100 corresponding past-tense tokens which contained an additional inflection phoneme. Words were mapped to randomly generated semantics represented by 20-bit binary vectors of which 5 units were set to 1 for each stem, plus an additional inflection unit for the past tense.

Procedure The networks were trained for 7500 epochs. During training, the tokens were sampled probabilistically according to a Zipfian power-law function, $P(r) = \frac{C}{r^{\alpha}}$, where rank-$r$ token had the frequency $P(r)$. We used $C = 0.1$ and $\alpha = 1.01$—the values for natural languages (e.g. Li, 1992). For each weight-update, 5 binary patterns were generated from their stochastic representations. The results are averaged over three runs using different random corpora. The network weights were randomly initialized on each run between (-1,1).

Results As predicted, the past-tenses of the most frequently-presented tokens were reduced by the greatest amount (Fig. 1). The most frequent token was reduced from a starting length of 4, to just the inflection unit. In all cases, past-tense tokens maintained the inflection unit due to the perfect correlation between it and the inflection unit in the output layer. The next simulation shows this does not need to be the case—networks trained on a more natural corpus can in addition to regulars produce both suppletions (e.g.,

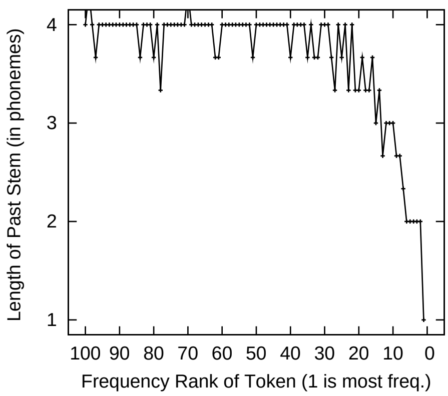

Figure 1: Effect of presentation frequency on the reduction of the past-tense form. All tokens began with a length of 4 phonemes.

go ⇒ *went*) and ablauts (e.g., *swim* ⇒ *swam*). Since the most frequent words were the first ones learned, they were the first ones free to change due to their low errors. However, the network was considerably over-trained, having learned the corpus to perfection after only 700 epochs. Compositional irregularity was therefore a stable property of this network.

Simulation 2a: From *doed* to *did*: Emergence of quasi-regularity in an English corpus

This simulation extends the framework of Simulation 1 to a more natural corpus of English verbs. We hypothesized the competing pressures for accurate communication and phonological simplicity would produce patterns of quasi-regularity similar to those evidenced in English. The simulation was initialized with all past tenses in the training set fully regular; a later simulation explores the effects of including other types of exceptions in the initial corpus.

Method

Networks The networks were identical to those used in Simulation 1 except their size was changed to accommodate a much larger corpus of actual phonology and semantics. There were 60 input units (see Table 1), 200 hidden units, and 80 output units.

Materials Like the first simulation, the input layer contained units for onset, vowel, and coda phonemes (see Table 1). The groupings indicate mutually exclusive phonemes—only one can be active per group in a word. The groups are ordered left-to-right in accordance with ordering constraints for monosyllabic English words. So, while /r/ can precede /l/, e.g., *curl*, the reverse is not true. The representations differ slightly from those used by Plaut, McClelland, Seidenberg, and Patterson (1996). In particular, the phonotactic constraints have been strengthened for our corpus. Also, the vowel representations have been modified to convey complexity or markedness by means

Table 1: Phonological Representations Used in Simulations 2a-c

onset	s S C z Z j f v T D p b t d k g m n h l r w y
vowel	a e i o u A^ U E
coda	r l m n N p k t* f v T D S Z s z C j
	p b k g d t d t (z s)**

/a/ in sprawl, /e/ in help, /i/ in pick, /o/ in pop, /u/ in put, /A/ in catch, /^/ in was, /aE/ in like, /eE/ in wake, /iE/ in leer, /oU/ in hope, /uU/ in moo, /^U/ in saw, /^UE/ in hoist, /N/ in ring, /S/ in she, /C/ in chin, /Z/ in beige, /T/ in thin, /D/ in this. All other phonemes are represented in the conventional way.

* These are only used in combinations with /s/ such as /ps/ in lapse /ks/ in fix, /ts/ in cats. The separate set to disambiguate between /ps/ and /sp/, /ks/ and /sk/, /ts/ and /st/, which are all allowed in English.

** Used for the second d/t sound when the stem already ends on d/t (e.g., hate ⇒ hated). The /z/ and /s/ units correspond to third-person inflections (/s/ in lapses, /z/ in graces) which are not used in this corpus.

of an extra active unit—diphthongs and other long or complex vowels have two units active, while simple or short vowels have only one active unit.

The semantics of the output layer were obtained by training an auto-encoder network to compress the 645-feature semantics generated by Harm (2002) into semantics distributed over 80 units. This compression was performed both to reduce training time and to produce quasi-compositionality in the semantic representation of the past-tense—that is, the difference between stem and past semantics was a pattern that was similar, but not identical, from one verb to another. The corpus, adapted from Harm (2002), contained 739 monosyllabic verb stems (1478 tokens) of which 124 were exceptions. Each word was assigned a frequency derived from the American Heritage Dictionary (AHD) corpus (Carrol, Davies, & Richman, 1971). The AHD provides more robust frequency estimates than the more commonly used Brown corpus without the idiosyncrasies of the Wall Street Journal corpus. The frequency was transformed into a probability of presentation through a moderate square-root compression (see Plaut et al., 1996, for further discussion). The resulting frequency difference between the very low-frequency words such as *thrash*, *stow*, *foist*, and *meld* and the very high-frequency words such as *be*, *say*, and *do* was about 200:1. This is a considerably smaller frequency difference than is encountered in real language which in many cases can be greater than 20,000:1. At the same time, because our corpus contained only monosyllabic verbs, and almost all irregulars are monosyllabic, irregulars were over-represented. Training with a compressed frequency range compensates for this over-representation because irregulars tend to be of higher frequency than regulars, and so are disproportionately affected by the compression (Plaut et al., 1996).

Procedure Training details were the same as in Simulation 1 except that the networks were trained for only 4500 epochs. The results combine the average of three runs.

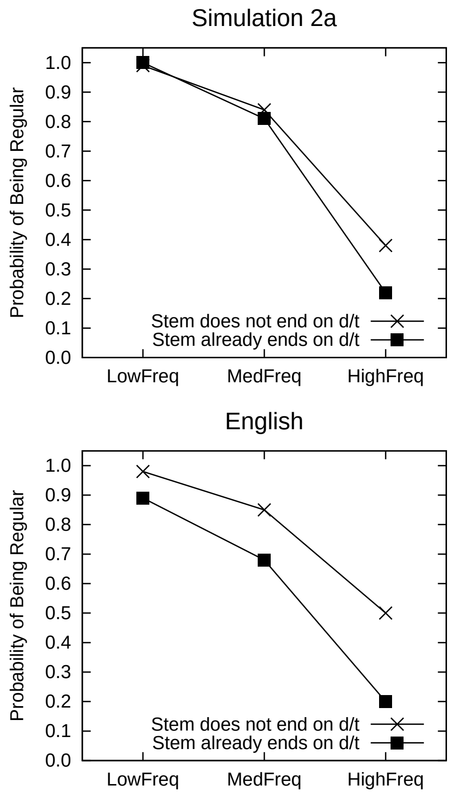

Figure 2: Comparison of networks trained on a fully-regular English corpus with data from a real English corpus. Probability of being regular refers to the proportion of words that fully retained their stem, and ended on /d/ or /t/ in the past-tense.

Results The pattern of quasi-regularity produced by the networks matched closely to that of English (Figure 2). To simplify analysis, we divided the corpus into three frequency categories: the highest-frequency contained the most frequent 10% of the words, and the mid- and low-frequency categories each contained 45% of the remaining tokens. The general trend is clear: the more frequent a word is, the less likely it is to maintain compositionality by being fully regular. We also see an effect of stem type (whether the stem ends on d/t or another phoneme) and the probability of being regular. The explanation is intuitive: if a word already ends on /d/ or /t/, it already contains the phonological signal characteristic of the past-tense. If this word is a high-frequency word, it is under a particularly strong pressure to maintain its brevity, which can be retained through a stem-change. At the same time, the presence of d/t ending maintains the word's compositionality. The pressure for shortness is further illustrated in Figure 3. Since the low-frequency words are almost exclusively regular, there is an average of 1 phoneme added to the past-tense form. In contrast high-frequency words often maintain their length in the past-tense, and so the average number

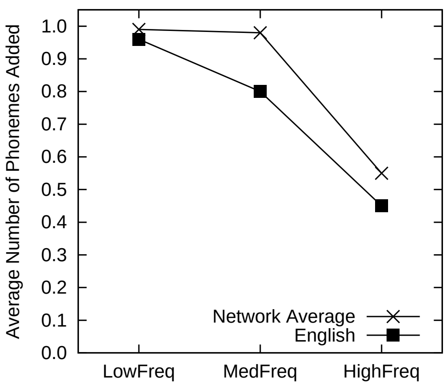

Figure 3: Average number of phonemes added to the past-tense as a function of frequency.

of phonemes added is closer to 0.5.

The patterns produced by the networks deviated in several ways from English. In English, of the 48 irregulars whose stems end on /d/ or /t/, none take an additional /d/ or /t/ in the past-tense. In contrast, the networks add an additional /d/ or /t/ to about half of such irregulars producing patterns like *builded* and *standed*. Similar over-regularization can be observed in low-frequency irregulars. For instance, *cast* is often regularized to *casted*, and *wed* to *wedded*—the network duly matched such tendencies by over-regularizing all such words. Allowing for such over-regularization, the low-frequency d/t-stem datapoint in Figure 2 would match the network's value of 1. Another difference between English and our simulation results is that the latter did not produce no-change verbs (e.g., *bid*). This is to be expected since the networks' task is to learn to map phonology to semantics—having the same word for both present and past-tenses would increase this error. Such lack of change is possible in a natural language due to contextual and pragmatic cues, which are not available in our simulations.

There is also a significant discrepancy between the mid-frequency points in Figure 3. We can understand its basis by analyzing the breakdown of irregulars. It turns out that of the 45 irregulars of mid-frequency whose stems do not end on /d/ or /t/ the networks formed 90% of the past-tenses by over-applying patterns like *keep* ⇒ *kept*, presetving the regular ending. This compares to only 24% in English—much of the remainder take the form of ablauts (e.g., *swim*) whose length in the past is the same as in the present. Simulation 2b addresses this discrepancy.

Simulation 2b: Inclusion of Sub-Regularities

Simulation 2a started out with a fully-regular corpus—all words ended on /d/ or /t/. This consistency created a very strong tendency to maintain those endings and satisfy the pressure for shortness through other means, such as compressing the stem. There is nothing to suggest, however, that English, or any language, *began* as a fully-regular system. Old-English, for instance, contained seven classes of strong verbs, each undergoing a change that was systematic within that class. In our framework, the frequent members of each class would act to support the particular idiosynrasies of other members of their class. The goal of this simulation was to observe whether items in such classes can maintain their inflections even while there is a pressure to shorten regular items.

Procedure As a first attempt to address these issues, we used corpora containing a mixture of regularized and standard past-tense forms. Five corpora were constructed, each containing a randomly-selected portion (20% without replacement) of word whose past tenses were initially fully regular. The remaining 80% of each corpus was left in its original form, so that most of the irregular verbs appear as they do in contemporary English.

Results For the regularized forms, the results were not significantly different from those of Simulation 2a—the networks still over-applied the d/t ending. However, the ablaut verbs (such as *sing* ⇒ *sang* whose past tenses were not irregularized initially *did* resist the pressure for adding d/t. Considering that the English ablauts originated from the Old-English vowel-changing classes of strong verbs, and so never had a d/t suffix to begin with, the networks' results are not at odds with English diachronics. It appears the moderate-frequency of these verbs, together with neighborhood effects prevented them from being subsumed by the pressure to add d/t.

Simulation 2c: Systematizing Semantics

Unlike Simulation 1 which assumed the past-tense to be fully orthogonal to semantics, Simulations 2a-b used a distributed notion of "past-ness." Evidence from language acquisition supports the view that, at least for young children, "past-ness" is not abstract, and so is not fully independent of semantics. Shirai and Anderson (1995) report that when the past-tense first appears in children's speech, it is largely restricted to punctate events that have endpoints and produce results (e.g., 'I dropped it'); it then gradually spreads to cases in which one of the typical properties (is punctate, has endpoint, produces results) is violated (McClelland & Patterson, 2002). This simulation aimed to evaluate differences produced by using past-tense units that are fully orthogonal to the stem semantics.

Procedure All training details were identical to Simulation 2a except for the addition of three extra units to the output layer. These units were set to 0 for present-tense tokens. Past-tense tokens used the semantics of the present-tense, and had the three "pastness" units set to 1.

Results Networks trained on the systematized corpus maintained a stronger pressure to regularize, regularizing 60% of high-frequency words whose stems end on /d/ or /t/, as opposed to networks in Simulation 2a which regularized only 22%. The better-fit of Simulation 2a (22%) to English data (20%) suggests semantic influence on inflectional features such as tense, plurality, and aspect may contribute to morphophonological change. Namely, high-frequency forms of words can take a life of their own, often being used in highly idiosyncratic ways (Bybee, 1985) and in doing so, become partially freed from compositionality constraints.

Conclusion

We have shown how feedforward neural-networks can produce patterns of stable, semi-compositional irregularity in both an abstract corpus and a corpus of English verb stems. The networks provide a framework employing distributed representations and graded constraints, which may provide a useful basis for investigating language structure and language change. Our simulations do not take into account many historical factors which are surely influential in the evolution of any particular language. Also, since the networks are strictly feedforward and the words are presented one at a time, it does not have the benefit of inferring the past-tense from context, something that humans readily do (e.g. Bybee, et al., 1994). The lack of recurrence also means the networks "hear" a word all at once instead of sequentially, so the phonological reduction is just as likely to happen to the word-onset instead of the middle or final word-segments, as it typically does in natural languages (Burzio, 2002). Test runs of a simple-recurrent network on a corpus similar to the one used in Simulation 1 were encouraging. The networks reliably dropped the vowel units, corresponding to reductions like *say* ⇒ *said* (/seE/ ⇒ /sed/) or offset units, e.g., *make* ⇒ *made* (/meEk/ ⇒ /meEd/). In spite of the mentioned limitations, the emergent patterns of quasi-regularity as produced by the networks capture aspects of the patterns found in an actual language.

Acknowledgments

The first author wishes to thank Punitha Manavalan for fruitful discussions of many of the ideas contained in this paper, Doug Rohde for valuable programming tips, Mike Harm for kindly providing semantic representations, Luigi Burzio for helpful commentary, and Erin Ingvalson for comments on an earlier version of the draft.

References

Batali, J. (1998). Computational simulations of the emergence of grammar. In Hurford, J. R., Studdert-Kennedy, M. and Knight C., editors, *Approaches to the Evolution of Language - Social and Cognitive Bases.* Cambridge: Cambridge University Press.

Bybee, J. (1985). *Morphology: A study of the relations between meaning and form.* John Benjamins Publishing Company, Phildalphia.

Bybee, J. (1995). Regular Morphology and the Lexicon. *Language and Cognitive Processes.* 10(5): 425-455

Bybee, J., Perkins, R., Pagliuca, W. (1994). *The Evolution of Grammar: tense, aspect, and modality in the languges of the world.* The University of Chicago Press.

Burzio, L. (2002). Missing players: Phonology and the past-tense debate. *Lingua* 112: 157-199.

Carroll, J. B., Davies, P., Richman, B. (1971). *The American Heritage Word Frequency Book.* Boston: Houghton-Mifflin.

Greenberg, J.H. (1957). *Essays in Linguistics.* The University of Chicago Press.

Harm, M.W. (2002). Building Large Scale Distributed Semantic Feature Sets With WordNet. *CNBC Tech Report* PDP.CNS.02.1

Kirby, S. (2000). Syntax without Natural Selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, editor, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*: 303-323. Cambridge University Press.

Kirby, S. (2001). Spontaneous Evolution of Linguistic Structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation* 5(2):102-110.

Kirby, S. (2002). Learning, Bottlenecks and the Evolution of Recursive Syntax. In Ted Briscoe, ed., *Linguistic Evolution through Language Acquisition: Formal and Computational Models.* Cambridge University Press.

Li, W. (1992). Random Texts Exhibit Zipf's-Law-Like Word Frequency Distributions. *IEEE Transactions on Information Theory*, 38(6): 1842-1845

McClelland, J.L., Patterson, K. (2002). Rules or Connections in Past-Tense Inflections: What does the Evidence Rule Out? *Trends in Cognitive Science* 6(11): 465-472.

Miikkulainen, R., Dyer, M.G. (1991). Natural Language Processing with Modular PDP Networks and Distributed Lexicon. *Cognitive Science* 15:343-399.

Pinker, S. (1991) Rules of language. *Science*, 253, 530-535.

Plaut, D.C., McClelland, J.L., Seidenberg, M.S., Patterson, K. (1996). Understanding Normal and Impaired Word Reading: Computational Principles in Quasi-Regular Domains. *Psychological Review*, 103, 56-115

Rissanen, J., Ristad, E.S. (1994). Language Acquisition in the MDL Framework. In Eric Sven Ristad, editor, *Language Computation.* American Mathemtatical Society, Philedelphia.

Rohde, D.L.T. (1999). LENS: The light, efficient network simulator. *Technical Report* CMU-CS-99-164, Carnegie Mellon University, Department of Computer Science, Pittsburgh, PA.

Shirai, Y., Anderson, R.W. (1995) The acquisition of tense-aspect morphology: a prototype account. *Language* 71, 743762

Smith, M. A., Cottrell, G. W., Anderson, K. L. (in press). The early word catches the weights. In: *Advances in Neural Information Processing Systems* 12. The MIT Press, Cambridge, MA

Tyler, L.K. *Spoken Language Comprehension: An experimental approach to disordered and normal processing.* The MIT Press.

Zemel, R.S. and Hinton, G. E. (1994). Developing population codes by minimizing description length. *Advances in Neural Information Processing Systems* 6. J. D. Cowan, G. Tesauro and J. Alspector (Eds.), Morgan Kaufmann: San Mateo, CA.