

# The Dynamic Lexicon

Janet B. Pierrehumbert  
Department of Linguistics  
*and* Northwestern Institute on Complex Systems  
Northwestern University, Evanston, IL 60208, USA

October 16, 2010

## 1 Introduction

The lexicon is the central locus of association between form and meaning. The prior sections in this chapter focus on the lexicon as it figures in the cognitive systems of individuals. The lexicon can also be viewed at the level of language communities, as shared intellectual property that supports mechanisms of information transmission amongst individuals. This viewpoint is foreshadowed by Hawkins (this volume), and sketched for linguistic systems in general in Hruschka et al. (2009). Here, I consider the relationship between the lexical systems of individuals and lexical systems at the community level. The dynamics of these systems over time, rooted in their relationship to each other, can inform our understanding of the lexicon, and of the entries and relationships that comprise it. Tackling problems in lexical dynamics, in the light of experimental findings and synchronic statistics, provides laboratory phonology both with fresh lines of empirical evidence and with fresh arenas for theoretical prediction.

The lexicon is generally assumed to list any associations between form and meaning that are idiosyncratic and must be learned. Thus, it includes not only morphologically simple words, but also irregular or opaque complex

words, and collocations. Recently, it has been shown to include morphologically regular words as well (Alegre and Gordon, 1999; Baayen, Wurms, and Aycock, 2007). The following discussion emphasizes words (whether morphologically simple or complex), though frequent phrases also appear as a source of new words (Bybee, 2001). According to the phonological principle, forms of words (*wordforms*) are combinations of basic building blocks, which are characteristic of any individual language, meaningless in themselves, but meaningful in combination. Evidence has recently accumulated that, in addition to this abstract level of characterization, lexical entries also include density distributions over detailed phonetic or socio-indexical properties. I accordingly view wordforms as both detailed and abstract (Pierrehumbert, 2006a; contributions by Hawkins and Albright, this volume).

Do people use words? Or, do words use people? At the population level, words inhabit communities of speakers in rather the same way as species inhabit ecological niches (Altmann, Pierrehumbert, and Motter 2010). Although some words die out, just as some species go extinct, the addition of new words sustains the overall complexity of a lexical system. People both borrow words from other languages, and invent new words by generating names for new concepts (Munat, 2007). There is no corpus big enough to include all the words of a language; as a corpus expands to include more topics, more speakers, and longer time periods, new words are always found (Baayen, 2002; Manning and Schuetze, 1999). Even the most stable core vocabulary of Indo-European languages has been replaced as a rate of about 20 words per millenium (Sankoff, 1970; Pagel, Atkinson, and Meade, 2007). Berko's wugs paradigm demonstrated the ability of even small children to invent new words through productive use of morphology (Berko, 1958), and in adults, this ability is demonstrated both using the same experimental task (Albright and Hayes, 2003; Pierrehumbert, 2006b), and through the statistics of languages with highly productive morphology (Hankamer, 1989). Grammaticalization theory in turn reveals how morphologically complex forms can

provide a source of simpler forms on the historical scale (Bybee, 2001; Hopper and Traugott, 2003).

Like species, lexical innovations compete with preexisting forms to survive. Words are viable only insofar as they are successfully replicated. For species, biological reproduction is the mechanism for replication. For words, the mechanism is imitation. Children bring to the task of language acquisition fundamental drives to attend to and imitate speech patterns (Vihman, 1996), and to map word forms to word meanings on the basis of phonological and semantical contrast (Clark, 1987). Through iterated imitation, linguistic communities converge on shared names for objects and concepts (Steels, 1995, 1997) and on shared phonological inventories (de Boer, 2000). This population-dynamic view of the lexicon points to a nexus of cognitive and social factors in determining the long-term dynamics of the lexicon (Komarova and Nowak, 2001). I now review some general properties of words and lexicons that are critical for the understanding of this dynamics. I first consider the intrinsic nature of the coding system. Next, I discuss frequency as the reflex of a word's success and as a contributor to lexical dynamics. Finally, I discuss possible mechanisms for new words to overcome the disadvantage of their initially low frequency, and become widespread in the community.

## **2 The phonological code**

Words are replicated by being learned and then used later. The phonological representations of words supports highly accurate replication – if this were not so, then people would not be able to understand each other as well as they do. But there is also room in the lexicon for new words. These two characteristics can be understood by considering phonological representation as a error-correcting code.

Phonology is a code because it represents the speech stream using

sequences of elements from a finite alphabet. A simple illustration of this fact is that blends of two words, such as *celebrademic* from *celebrity* and *academic*, do not average the wordforms of the contributing lexemes, but rather sequence components from one lexeme with components from the other, or common to both (Lehrer, 2007). In classical linguistic theory, the alphabet was the set of phonemes of the language, defined as minimal units of lexical contrast (Hockett, 1961). Though this conceptualization of the phonological code has been updated by autosegmental-metrical theory, the central insight that the code concerns informative contrasts remains. In what follows, I will use the term *segment* as a theoretically neutral term for a phone or phoneme, without commitment to its minimality or abstractness.

From the earliest days of information theory, speech scientists sought to understand the information density of the phonological code, a literature reviewed in Boothroyd and Nitttrouer (1988) and Allen (1994). The basic unit of information is the bit, representing a choice or uncertainty between two equally likely alternatives. The smallest number of distinctive features proposed in any phonological theory is 12 (Mielke, 2008), which means that English would have an information density of at least 12 bits per segment if all feature values occurred equally and in all combinations. However, mathematical analysis of error patterns for speech perception in noise, with varying amounts of lexical and contextual information, reveals that well-formed English CVC words contain only 10.3 bits of information in total (representing a choice of one word out of 1260 alternatives) or 3.4 bits per segment on the average. Phonotactically well-formed monosyllables (considering words and nonwords together) have a greater information density than words, at 4.3 bits per segment. This reflects the existence of accidental gaps in the lexicon, which provide spaces for new words to be added. Since 4.3 is still far smaller than 12, it also reveals the great redundancy imposed by phonotactics and feature cooccurrence restrictions.

In the theory of information coding and transmission, redundancy is

useful for correcting errors. The redundancy in phonological representations reduces the likelihood that a sloppy, erroneous, or poorly heard production will be perceived as an unintended lexical meaning. Word error rates for human speech perception in good listening conditions are negligible, and perception in unfavorable listening conditions is surprisingly robust (Kalikow, Stevens, and Elliott, 1977). Many individual words can be uniquely identified even if one or more segments are missing. This is shown by phoneme restoration experiments, in which people fail to notice that a speech segment has been replaced by noise (Samuel, 1981), and by gating experiments, in which people prove able to progressively narrow the set of lexical choices as more and more of the word is provided, often achieving a unique identification before the end of the word (Grosjean, 1980). Eye-tracking experiments show that coarticulatory information is used as soon as possible (Dahan, Magnuson, and Tanenhaus, 2001). There is a strong lexical bias in speech perception (Ganong, 1980), so that phoneme category boundaries for well-formed non-words (such as *zill* and *woot*) are unconsciously shifted to perceive the most similar real words (*sill* and *wood*). Morphophonological alternations also have a strong tendency to operate within the discrete system of phonological representation (Kiparsky, 1985), a behavior that supports error-correcting perception and production for morphologically complex words (and not just simple ones). This functional pressure is so strong that it can cause phonetically conditioned alternations (such as the assimilation of consonants to neighboring vowels) to evolve over time to become more categorical, even at the expense of phonetic naturalness (Anderson, 1981).

Redundancy is a reciprocal informational dependency, as discussed in Broe (1993), Steriade (1995), and Frisch, Pierrehumbert, and Broe (2004). Elements are redundant to the extent that they can be predicted from each other. Predictions can ensue from either positive statistical correlations (known in phonological theory as *harmony* rules or constraints) or negative correlations (known as *OCP* or *Obligatory Contour Principle* constraints).

For example, in a language with coronal harmony (such as Chumash), the value of the feature [anterior] for any given strident is largely predictable from any other (Avery and Rice, 1989). A strong OCP effect on place of articulation is found in the Arabic verbal roots. The presence of a consonant at some given place in initial position strongly **disfavors** the occurrence of consonants with the same place in second position, **and vice versa**. Frisch and Zawaydeh (2001) demonstrate that such statistics are part of the implicit knowledge of native speakers. Lahiri (this volume) puts forward some examples of asymmetric informational dependencies relating to the featural makeup of segments. The interest of these examples lies in their contrast with the main thrust of the experiments just reviewed on speech perception in noise, phoneme restoration, gating, eye-tracking, and well-formedness. Overall, people make appear to make optimal use of available statistical information, including the correlations that cause great redundancy in the system. The primary source of informational asymmetry in speech processing is the flow of time in on-line tasks, which causes some information to be available sooner than other information.

In word phonology, redundancy is found at multiple time scales. At one extreme, consider the avoidance of long words. English has some 43 segment types, whose crossproduct would yield 1849 words with two segments, 79,507 words with three segments, in short  $43^n$  words of length  $n$ . But the overall distribution of word lengths is not exponentially increasing. Instead, it is approximately log-normal (Limpert, Stahel, and Abbt, 2001). Relatively few words are extremely short, but past the modal word length of 5 or 6 segments, the likelihood that a given phonological combination exists as a real word becomes vanishing small as length increases. This result can be derived by assuming that a cost function penalizes each additional coding unit (Mitzenbacher, 2004). The experiment on wordlikeness judgments by Frisch, Large, and Pisoni (2000) establishes the cognitive reality of this basic observation. Feature co-occurrence restrictions within segments provide an

example at the shortest time scale. For example, in Indic languages, stops contrast in both breathiness and voicing (2 bits of information), whereas in English these dimensions are conflated (providing only 1 bit taken together). The nondistinction between /r/ and /l/ in Japanese has been particularly well studied. The third formant is the primary cue for this contrast in English. Monolingual Japanese speakers have a poorer neural representation of the third formant than English speakers do, but the neural representation increases if they receive training in the distinction (Zhang et al., 2009). Such results indicate that phonological dimensions (not just phonological categories) are acquired by language learners in a manner that reflects how informative they are in the ambient language.

The nature and interaction of dependencies at different scales provides the motivation for autosegmental-metrical theory as an advance over classic phonemic theory. An autosegmental-metrical constraint amounts to a claim about a statistical dependency at the scale of the constraint. As reviewed in Goldsmith (1990), autosegmental-metrical representations are directed acyclic graphs that encapsulate these dependencies. The leading idea is that dependencies prove to be local if the proper abstract units are defined. Locality is defined in two ways. Metrical units, such as the syllable, the foot, and the prosodic word provide the underpinnings for constraints that involve a head-dependency structure. Tiers provide the underpinning for constraints that pertain to a span without regard to headedness.

The cognitive reality of autosegmental-metrical constraints is demonstrated by a variety of experimental paradigms, including speech segmentation, well-formedness judgments, error patterns, and memory effects. Suomi, McQueen, and Cutler (1997) show that vowel harmony in Finnish is exploited to segment the speech stream into words. Cutler and Butterfield (1992) show that the typical trochaic stress pattern of English words is used in the same way. Lee and Goldrick (2008), and Kapatsinski (2009) provide recent best-practice examples of an immense literature on syllable structure. Both bring

together multiple strands of evidence to compare the syllable rhyme and the body (defined as the onset plus nucleus) as cognitively relevant units of prosodic structure.

Accidental gaps in the lexicon are words that do not exist, but are perfectly possible. Autosegmental-metrical theory posits constraints on words in general; these constraints are gradient insofar as the theory is statistically fleshed out. In between the accidental gaps and the general theory lie a set of phenomena that have recently provided critical evidence about the cognitive representations. These are the lexical neighborhood and lexical gang effects.

The lexical neighborhood of a word is the set of words that are minimally different from it (see Frisch, this volume). Though the size of a word's lexical neighborhood is correlated with its overall phonological likelihood, careful experiments have identified dissociations that provide an important argument for a cognitive system with multiple levels of representation, including both an encoding level and a lexical level (Vitevich and Luce, 1998; Luce and Large, 2001; Thorn and Frankish, 2005). Lexical gangs are sets of words with shared phonological and semantic properties that influence morphological productivity. An example is the set of monosyllabic degree adjectives ending in obstruents that accept the suffix *-en*, such as *black+en*, *white+en* but not *\*green+en*, *\*abstract+en* (Alegre and Gordon, 1999). Gang behavior can also be identified for groups of words with shared phonological and semantic components that do not share morphemes in the standard sense, such as *glimmer*, *gleam*, *glint*. (Bergen, 2004; Tamariz, 2009).

Experimental results on lexical gangs and neighborhoods show that subsets of the full lexicon, defined as clusters of words that are particularly similar amongst themselves, have pervasive force. The results support a picture of the lexicon in which words are organized in a network, where the links represent shared phonological and semantic properties (McClelland and Elman, 1986; Bybee, 2001; Hay and Baayen, 2005). The same network is explanatory both for speech processing, and for phonological abstraction and

productivity. In processing, activation and inhibition of nodes over time explains perception and production as they unfold in time. Abstractions over groups of word provide the foundation for constraints and for the creation of well-formed new words. Can arbitrary groups of nodes provide the grist for abstraction and generalization? Clearly not. All successful approaches share the insight that the cognitive system forms abstractions from coherent or natural sets of words. A central goal of the network representation is to define the link structure in a way that makes natural groups appear as connected sub-networks of the entire network. Evidence is accumulating that the dimensions of similarity and comparison that define the links are shaped by functional factors at all levels from the perceptual and articulatory periphery to general principles of cognition. For example Lindblom and Maddieson (1988) and Lindblom et al. (1995) present typological data indicating that the consonant inventories reflect a tradeoff of perceptual distinctiveness and articulatory complexity. The results of Albright and Hayes (2003) imply that phonological material temporally adjacent to an affix is more relevant to the productivity of the affix than material in more remote parts of the word. Hudson-Kam and Newport (2009) adduce a cognitive bias towards categorization of frequencies, e.g interpreting experience frequencies as more extreme than they really are.

Though these functional factors are reminiscent of innate knowledge in the classic sense of generative phonology, there are also important differences. The differences arise because of the way that functional biases interact with the replication dynamics for the language system. Slight biases can have large effects in structuring the system, because their effects cumulate over time (Reali and Griffiths, 2009). Under strong simplifying assumptions, the system is even guaranteed to converge to the prior biases that the learner brings to the learning task (Griffiths and Kalish, 2007); but as these authors note, the prior biases may either be innate to the cognitive system, or be rooted in external factors. Under more realistic assumptions,

social subgroups can prevent shared norms from emerging (Lu, Korniss, and Szymanski, 2007) and oscillations and chaotic variation in the system over time can also arise (Mitchener, 2003; Mitchener and Nowak, 2004). I return to the challenges raised by these findings in the last section.

### 3 Frequency

Statistical learning is central to the picture of lexical dynamics presented thus far. Word types survive to the extent they can replicate themselves through the learner’s experience of word tokens (Nowak, 2000) and the abstract generalizations that govern lexical productivity are also statistical in nature (Pierrehumbert, 2003). Let us therefore consider word frequency more carefully.

Word frequency effects are among the most robust effects known in psycholinguistics. Less-frequent words are recognized more slowly and less reliably than more-frequent words. They are more vulnerable under unfavorable listening conditions (Kalikow, Stevens, and Elliott, 1977). They are also more vulnerable to replacement on historical time scales (Bybee, 2001; Lieberman et al., 2007). This last effect arises not only because they are less well learned, but also because they are less likely to be learned at all. A rare word may simply fail to occur by chance in the experience of a learner, and in that case it will not be learned and reproduced for future learners. In the aggregate, statistical sampling considerations mean that the frequencies of individual words are subject to *random walk* effects over generations, and that any word whose frequency happens to become too low will be irretrievably lost. The random walk of frequencies can create morphological gaps (Daland, Pierrehumbert, and Sims, 2007). It entails that the total number of distinct words in the community lexicon would decrease over time, if new words were not continually added (Fontanari and Perlovsky, 2004).

Word frequencies can vary by orders of magnitude across contexts

(Altmann, Pierrehumbert, and Motter, 2009), and the context for early word learning – the daily lives of small children – is different from the context for later word learning. Later words are only learned in competition with earlier ones, obeying general principles of contrastiveness in form and meaning (Clark, 1987). A new word will be learned only if the powerful error-correcting mechanisms of speech recognition and lexical access do not cause it to be recognized as a pre-existing word. It is initially encoded with the phonological resources that the child commands at that time. Werker and Stager (2000) find that 11 to 12-month olds require multiple points of phonological contrast to successfully map new words onto new referent. A fascinating series of studies by Storkel (2002, 2004) indicates that phonotactics and similarity neighborhoods are dynamically redefined as the lexicon emerges. This dynamics for word learning also predicts individual differences in acceptability of nonwords as new words. Frisch et al. (2001) indeed report that individuals with large vocabularies are more accepting of statistically marginal nonwords. This might occur because unusual phonological components of the nonwords are more likely to already occur in their vocabularies. It might occur because phonological generosity is what permitted them to learn so many words in the first place. These two possibilities can be integrated into a more general and abstract picture, in which a positive feedback loop relating vocabulary size and phonological encoding provides the explanatory dynamics for vocabulary growth; see Munson et al. (this volume) for further discussion.

Frequency effects play a large role in grammaticalization theory, which documents a connection between synchronic statistics on frequency and word length, and typical patterns of historical evolution (Bybee, 2001 and 2007; Hopper and Traugott, 2003). Synchronically, more frequent words tend to be both shorter than less frequent words and less subject to analogical pressure. Diachronically, words and phrases that become more frequent through semantic bleaching (loss of semantic concreteness in connection with usage

as grammatical markers) also become shorter. A typical example is the rise of *gonna* as a future (from the expression *going to* (Poplack and Tagliamonte, 1999; Cacoullos and Walker, 2009). Now, frequent words are more expectable than infrequent words. An optimal coding system is obtained if high frequency words have logarithmically shorter labels than more surprising lower frequency words (Shannon, 1948; van der Helm, 2000). Thus, the lexicon is shaped by functional pressures towards uniform information density, a functional pressure that is thought to be relevant for the linguistic system at all levels. (Zipf, 1949; Goldsmith, 2002; Aylett and Turk, 2004; Levy and Jaeger, 2007; Frank and Jaeger, 2008). Shortening words that become frequent is desirable because it helps to optimize the transmission of information. It is possible because frequent words are perceived faster and more reliably even if degraded. It is implemented through articulatory reduction of wordforms that are accessed more easily through their frequency, contextual predictability, and lack of close competitors (Bell et al. 2009).

The loss of internal word boundaries during grammaticalization can further be interpreted within probabilistic models of morphology (reviewed in Hay and Baayen, 2005). According to these models, lexical items with meaningful subparts may be accessed either directly as wholes, or indirectly through the subparts. This approach makes nuanced predictions about the decomposibility of words and the productivity of affixes (Hay 2002, 2003). In relation to grammaticalization, the line of prediction is that the complex form will lose word structure as a function of three factors: if its frequency runs ahead of the frequencies of the parts, if the meaning is unpredictable from the parts, and if hypo-articulation induces the loss of phonotactic cues to the boundary. *Gonna* exemplifies this pattern through loss of the motion component of *going*, loss of the velar nasal as cue to a word boundary, and its rise in frequency as it becomes a generic future. Overall, given that a wordform rises in frequency, the observed phonological and morphological trajectories documented in grammaticalization theory are predicted.

But what might cause a word's frequency to rise in the first place? Words compete with each other in production, perception and learning, and the results presented thus far all favor high-frequency competitors over low-frequency competitors. A more frequent form appears more reliably in any finite sample of linguistic experiences used in learning. It is more likely to be learned earlier, interfering with later learning of lower frequency forms. It is more reliably encoded and decoded. The first factor alone already predicts that the lexicon will be simplified over time, and the other factors would only serve to accelerate this trend. To sustain the overall complexity of the lexicon over time, there must be a mechanism for newly invented – and therefore infrequent – words to climb the frequency gradient and come into widespread use.

## 4 Heterogeneity

In research on population biology and opinion dynamics, heterogeneity has proved key to understanding innovation and diversity over time. Heterogeneity is the opposite of uniformity. For words, we need to consider both lack of uniformity in the context and lack of uniformity amongst the speakers.

The niche of a word – analogizing to the niche of a species – may be viewed as the thematic and social contexts in which it is used. In population biology, the viability of a species is strongly correlated with the size of its niche (Jablonski, 2005; Foote, 2008). An analogy can be drawn to the viability of words by considering that a linguistic community explores an abstract conceptual space through its discourse over time, and that a word's viability depends on establishing a sufficiently large niche (Altmann et al. 2010). Cattuto et al. (2009), analyzing the lexicon of tags on Internet social networking sites, show that the typical growth rate for the number of word types as a function of text length can be derived from a few simple assumptions: Each word has few semantic associates (relative to the total size of the

lexicon), and the conceptual exploration by the community takes the form of a random walk. In this picture, global frequency is a chimera and what matters to learning and imitation by individuals is frequency in context. Word types that are very infrequent in general (averaging over time, space, and social context) can be very frequent and predictable in particular contexts (Church and Gale, 1995; Altmann et al., 2009), accruing in that context all the advantages of high frequency.

Just as a genetic mutation can create a species with a fitness advantage, a new word can have a fitness advantage deriving from the value and importance of its referent. In studies of opinion dynamics, this type of fitness is called an exogenous factor (in contrast to endogenous factors, which are internal to the system being studied). Studies of recommendation networks for YouTube (Crane and Sornette, 2008) and memes (popular phrases) on the internet (Leskovic, Backstrom, and Kleinberg, 2009) indicate that exogenous factors – such as new inventions, the occurrence of a concert, or the timetable for an election – can cause surges of popularity in the expressions used to discuss them on a scale of weeks or even days. When the value of a product increases with the number of people who have already adopted it, a small minority of users may define a tipping point for universal adoption. Mitchener (2003) develops this line of analysis for language by analyzing the replicator dynamics equations with a fitness function that increases as the number of speakers sharing a given linguistic pattern increases.

Most challenging is the case of endogenous change, in which a new expression gains traction without any real novelty in meaning or functional advantage (as argued for *gonna* in Cacoullos and Walker, 2009). This case can be analyzed from the point of view of the speakers, as the diffusion of a rare expression through a social network. The links in the network represent social affinity, regions of the network relate to subcommunities of the linguistic community, and adopting a new expression is similar to adopting a new opinion. Mathematical methods similar to those used to study

epidemics and catastrophic failures can then be used to explore the likelihood of an *information cascade* (a term introduced in Bikhchandani et al., 1998).

All current models of opinion dynamics that can generate cascades from a small minority of innovators, in the absence of a fitness advantage, depend on heterogeneity in the social network to do so. Baxter et al. (2009) show that a neutral model of social interaction cannot explain convergence to the current New Zealand norm with any realistic choices of parameters. Watts (2002) and Watts and Dodds (2007) generate opinion cascades by positing heterogeneity in the decision thresholds for adopting the new opinion; their *early adopters* can be understood in the present context as people who will use a rare new form because of its association with people that they particularly wish to emulate. Nettle (1999) demonstrated that linguistic cascading can be obtained by assuming that some highly connected individuals are much more influential than other people. A more sophisticated model by Fagyal et al. (2010) also generates cascading of initially rare innovations by assigning disproportionate importance to input received from speakers who are themselves socially well-connected.

Much work remains to be done in this area, because it is far from clear that innovative forms typically originate from or socially close to well-connected high status people. Indeed, the sociolinguistic literature shows that linguistic change typically originates from lower status speakers (Labov, 1994). However, the models provide clear support for the idea that individual words are associated with indexical information in people's minds. This is necessary because people use words later – sometimes much later – than they last heard them. Preferential adoption of words learned from certain people, or characteristic of certain groups or situations, depends on long-term encoding of these social factors. Indeed, experimental results demonstrate that indexical properties, including speaker identity, are encoded and remembered. (Palmeri, Goldinger, and Pisoni, 1993; Church and Schacter, 1994; review in Nygaard, 2005) The long-term dynamics of the lexicon provides

independent motivation for the conclusions of these studies.

## 5 Conclusion

The lexicon is a locus of creativity in language. When invented, novel forms reuse in novel combinations the discrete elements of the system, whether phonological or morphological. To be learned and adopted, novel forms must compete successfully against pre-existing forms in the replicator dynamics, a process of learning and imitation that is generally error-correcting, but also exhibits a systematic bias towards optimal encoding in the relationship of word length to word frequency.

Frequency effects, both in acquisition and in processing, predict the steady attrition of infrequent forms and the steady rise of frequent forms. Research on grammaticalization attests to this trajectory, including the predicted correlation of frequency with shortening. In the absence of additional factors, the lexicon would simplify over time, but the creation of new forms maintains its complexity. A new form must swim against the tide of frequency effects, and it can do so by several mechanisms. It may be intrinsically extremely fit because of exogeneous factors related to its meaning. It may cascade through the population on the strength of social factors. Mathematical models of cascading in related cases of opinion dynamics indicate that cascading of a rare innovative word can occur if the social network is heterogeneous, indexical properties are encoded with words, and these properties play a role in decisions to produce the word.

## 6 References

Allen, J.B. (1994) How do humans process and recognize speech? *IEEE Transactions on Speech and Audio Processing* 2(4), 567-577.

Albright, A. and B. Hayes (2003). Rules versus analogy in English past tenses: A computational-experimental study. *Cognition* 90, 119-161.

Alegre, M., and P. Gordon (1999) Rule-Based versus Associate Processes in Derivational Morphology. *Brain and Language* 68, 347-354.

Altmann, E.G, J. B. Pierrehumbert, and A. E. Motter (2009). Beyond Word Frequency: Bursts, Lulls, and Scaling in the Temporal Distribution of Words. *PLoS ONE*. PLoS One 4(11), e7678. doi:10.1371/journal.pone.0007678

Altmann, Eduardo G., Janet B. Pierrehumbert and Adilson E. Motter (2010) Niche as a determinant of word fate in online groups. arXiv:1009.3321v1.

Anderson, S. R. (1981). Why phonology isn't natural. *Linguistic Inquiry* 12, 493-539.

Avery, P, and K. Rice (1989) Segment structure and coronal underspecification. *Phonology* 6, 179-200.

Aylett, M. and A. Turk (2004) The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language & Speech* 47, 31-56.

Baayen, R. H. (2002) *Word Frequency Distributions*, Springer, Berlin.

Baayen, R. H., L. H. Worm, and J. Aycock (2007) Lexical Dynamics for Low-Frequency Complex Words: A Regression Study Across Tasks and Modalities *The Mental Lexicon* 2, 419-463.

Baxter, G. J, R. A. Blythe, W. Croft, and A. J. McKane (2009) Modeling

language change: An Evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change* 21, 257-296.

Bell, A., Brenier, J., Gregory, M., Girand, C., and D. Jurafsky (2009) Predictability effects on durations of content and function words in conversational English, *Journal of Memory and Language* 60, 92-111.

Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language* 80, 290-311.

Berko, J. (1958) The child's learning of English morphology. Reprinted 2004 in B. Lust and C. Foley (eds). *First language acquisition: the essential readings*. Blackwell Publishing, Oxford, 253-273.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1998), Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades, *Journal of Economic Perspectives*, 12(3), pp. 151-170.

Boothroyd, A. and S. Nittrover (1988) Mathematical treatment of context effects in phoneme and word recognition. *J. Acoust. Soc. Am.* 84(1), 101-114.

Broe, M. (1993) Specification theory: the treatment of redundancy in generative phonology Unpublished Ph. D. dissertation, University of Edinburgh.

Bybee, J. L. (2001) *Phonology and Language Use*. Cambridge University Press.

Bybee, J. L. (2007) *Frequency of Use and the Organization of Language*. Oxford University Press.

Cacoullos, R. T. and J. A. Walker (2009). The present of the English future: Grammatical Variation and collocations in discourse. *Language* 85, 321-354.

Cattuto C., A. Barrat, A. Baldassarri, and G. Schehr (2009) Collective dynamics of social annotation. *Proc Natl Acad Sci USA* 106:10511–10515.

Church, B. A. and D. L. Schacter (1994) Perceptual specificity of auditory priming – implicit memory for voice, intonation, and fundamental frequency. *J. Exp. Psych.: Learning, Memory, and Cognition* 20, 521-33.

Church, K. W. and W. A. Gale (1995) Poisson Mixtures. *Natural Language Engineering* 2, 15-59.

Clark, E.V. (1987) The principle of contrast: A constraint on language acquisition. In B. MacWhinney (ed), *Mechanisms of language acquisition*, Lawrence Erlbaum, Hillsdale, 1-34.

Crane, R. and D. Sornette (2008) Robust dynamic classes revealed by measuring the response function of a social system, *Proc. Nat. Acad. Sci.* 105, 15649-15653.

Cutler, A., and S. Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception, *J. Mem. Lang.* 31, 218-236.

Dahan, D, J.S. Magnuson, M.K. Tanenhaus (2001). Time Course of Frequency Effects in Spoken-Word Recognition: Evidence from Eye Movements *Cognitive Psychology* 42 (4), 317-367.

Daland, R., J. B. Pierrehumbert and A. D. Sims (2007) Much ado about nothing: a social network model of Russian paradigmatic gaps. *Proceedings*

of the 45th Annual Meeting of the Association for Computational Linguistics, 936-943.

de Boer, B. (2000) Self organization in vowel systems, *Journal of Phonetics* 28 (4), 441-465.

Fontanari, J. F. and L. I Perlovsky (2004) Solvable null model for the distribution of word frequencies, *Phys Rev E* 70 (4), 042901.

Foote M., J.S. Crampton, A.G Beu, R.A Cooper (2008) On the bidirectional relationship between geographic range and taxonomic duration. *Paleobiology* 34:421–433.

Frank, A. F. and T.F Jaeger (2008) Speaking Rationally: Uniform Information Density as an Optimal Strategy for Language Production, *Cogsci*, Washington DC: CogSci.

Frisch, S. A., N.R. Large, and D. B. Pisoni (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42:481-496.

Frisch, S.A., Large, N. R., Zawaydeh, B., and Pisoni D.B. (2001) Emergent phonological generalizations in English and Arabic. in J. L Bybee and P. Hopper (Eds.) *Frequency and the Emergence of Linguistic Structure*. John Benjamins, Amsterdam, 159-180.

Frisch, S. A., Pierrehumbert, J. B. and Broe, M. (2004) Similarity avoidance and the OCP, *Natural Language and Linguistic Theory* 22, 179-228.

Frisch, S A. and B. Zawaydeh (2001). The psychological reality of OCP-place

in Arabic. *Language* 77: 91-106.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110-125.

Goldsmith, J. A. (1990) *Autosegmental and Metrical Phonology*. Basil Blackwell, Oxford.

Goldsmith, J. (2002) Probabilistic Models of Grammar: Phonology as Information Minimization. *Phonological Studies* 5, 21-46.

Griffiths, T. L , Michael L. Kalish (2007) Language Evolution by Iterated Learning With Bayesian Agents. *Cognitive Science* 31, 441-480.

Grosjean F. (1980) Spoken word recognition processes and the gating paradigm. *Percept. Psychophys.* 28(4):267-83.

Hankamer, J. (1989) Morphological parsing and the lexicon in *Lexical representation and process*. MIT Press, 392 - 408.

Hay, J. B. (2002) From Speech Perception to Morphology: Affix-ordering Revisited. *Language* 78 (3), 527-555.

Hay, J. B. (2003) *Causes and Consequences of Word Structure*. Routledge, New York and London.

Hay, J. B. and R. H. Baayen (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences*, 9(7): 342-348.

- Hockett, C.F. (1961) Linguistic elements and their relations. *Language*. 37 (1), 29-53.
- Hopper, P.J., and Traugott, E.C. (2003) *Grammaticalization*. Cambridge University Press.
- Hruschka, D.J., M. H. Christiansen, R. Blythe, W. Croft, P. Heggarty, S. S. Mufwene, J. B. Pierrehumbert, S. Poplack. (2009), Building Social Cognitive Models of Language Change, *Trends in Cognitive Sciences*. 13:464–469.
- Hudson-Kam C.L., and E. L. Newport (2009) Getting it right by getting it wrong: When learners change languages. *Cognit. Psychol.* 59, 30-66.
- Jablonski, D (2005) Mass Extinctions and macroevolution. *Paleobiology* 31:192-210.
- Kalikow, D. N., Stevens, K. N., and Elliott, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J. Acoust. Soc. Am.* 61, 1337-1351.
- Kapatsinski, V. (2009) Testing Theories of Linguistic Constituency with Configurational Learning: The Case of the English Syllable Language. *Language*, 85:2, 248-277.
- Kiparsky, P. (1985) Some consequences of Lexical Phonology. *Phonology Yearbook* 2, 85-138.
- Komarova, N. L. and M. Nowak (2001) The Evolutionary Dynamics of the Lexical Matrix. *Bulletin of Mathematical Biology* 63, 451-484.

Labov, W. (1994) *Principles of Linguistic Change: Internal Factors* Blackwell, Oxford.

Lee, Y., and Goldrick, M. (2008). The emergence of sub-syllabic representations. *Journal of Memory and Language*, 59, 155-168.

Lehrer, Adrienne (2007) Blendalicious. In Munat, J. (Ed.) *Lexical creativity, texts and contexts*. John Benjamins, Amsterdam. 115-136.

Leskovec, J, L. Backstrom, and J. Kleinberg (2009) *The Dynamics of the News Cycle*, Proc. 15th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining.

Levy, R. and T.F Jaeger (2007) Speakers optimize information density through syntactic reduction. *Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems*.

Lieberman, E, J.-B. Michel, J. Jackson, T. Tang, M. A. Nowak (2007) Quantifying the evolutionary dynamics of language, *Nature* 449, 713.

E. Limpert, W. A. Stahel, and M. Abbt, Log-normal distributions across the sciences: Keys and clues. *Bioscience* 51, 341352 (2001).

Lindblom, B. (1990) Explaining phonetic variation: a sketch of of the H&H theory. in W. J. Hardcastle & A. Marchal ( eds.) *Speech production and speech modelling*. Kluwer. 403-440.

Lindblom, B., and I. Maddieson (1988) Phonetic Universals in Consonant Systems. In L.M. Hyman and C. N. Li (Eds.) *Language, speech, and mind: Studies in honor of Victoria A. Fromkin*, Routledge, London, 62–80.

- Lindblom, B., S. Guion, S. Hura, SJ Moon, (1995) Is sound change adaptive? *Riv Ling Ital J Ling*, 5-37.
- Lu Q., G. Korniss, and B.K. Szymanski(2009) The Naming Game in social networks: community formation and consensus engineering. *J Econ Interact Coord* 4:221-235.
- Luce,P.and N. R. Large (2001).Phonotactics,density,andentropyinspokenwordrecognition. *LanguageandCognitiveProcesses*,16,565581.
- Manning, C. and H. Schutze (1999) *Foundations of Statistical Natural Language Processing*, MIT Press. Cambridge, MA.
- McClelland, J. L. and J. L. Elman (1986). The TRACE model of speech perception. *Cognitive Psychology* 18: 1-86.
- Mielke, J. (2008) *The Emergence of Distinctive Features* (Oxford Studies in Typology and Linguistic Theory). Oxford Univ. Press.
- Mitchener, W. G (2003) Bifurcation analysis of the fully symmetric language dynamical equation, *Journal of Mathematical Biology* 46(3), 265-285.
- Mitchener, W. G. and M. A. Nowak (2003) Chaos and Language. *Proc. Roy. Soc. B*, 271, 701704.
- M. Mitzenmacher (2004) A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Mathematics* 1 (2), 226-251.
- Munat, J. (Ed.) (2007) *Lexical creativity, texts and contexts*. John Ben-

jamins, Amsterdam. 115-136.

Nettle, D., 1999, Using Social Impact Theory to Simulate Language Change. *Lingua* 108, 95.

Nowak, M. (2000). The Basic Reproductive Ratio of a Word, the Maximum Size of a Lexicon. *J. Theor. Bio.* 203, doi:10.1006/jtbi.2000.1085

Nygaard, L. C. (2005) Perceptual Integration of Linguistic and Nonlinguistic Properties of Speech. in D.B. Pisoni and R. E. Remez (Eds.) *The Handbook of Speech Perception*. Blackwell Publishing. 390-414.

Palmeri, T. J., S. D. Goldinger, and D. B. Pisoni (1993) Episodic encoding of voice attributes and recognition memory for spoken words. *J. Exp. Psych.: Learning, Memory and Cognition* 19, 309-28.

Pagel, M., Q. D. Atkinson and A. Meade (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449, 717-721.

Pierrehumbert, J. B. (2003). Probabilistic Phonology: Discrimination and Robustness. In R. Bod, J. Hay and S. Jannedy (Eds.) *Probability Theory in Linguistics*. The MIT Press, Cambridge MA, 177-228.

Pierrehumbert, J.B (2006a) The next toolkit. *J. Phonetics* 34(6), 516-530.

Pierrehumbert, J. B. (2006b) The Statistical Basis of an Unnatural Alternation, in L. Goldstein, D.H. Whalen, and C. Best (eds), *Laboratory Phonology VIII, Varieties of Phonological Competence*. Mouton de Gruyter, Berlin, 81-107.

Poplack, S., and Tagliamonte, S. (1999) The grammaticization of going to in (African American) English. *Language Variation and Change* 11, 315-342.

Reali, F., and Griffiths, T.L. (2009) The evolution of linguistic frequency distribution: relating regularization to inductive biases through iterated learning. *Cognition* 111, 317-328.

Samuel, A.G. (1981) Phonemic restoration: insights from a new methodology. *J. Exp. Psychol. Gen.* 110(4), 474-94.

Sankoff, D. (1970) On the Rate of Replacement of Word-Meaning Relationships. *Language* 47, 564-569.

Shannon, C.E. (1948) A mathematical theory of communication. *Bell System Tech J* 27:379-423.

Steriade, D. (1995). Underspecification and markedness. In J. Goldsmith (Ed.) *The handbook of phonological theory*. Oxford: Basil Blackwell.

Steels, L., (1995) A Self-Organizing Spatial Vocabulary. *Artif. Life* 2(3), 319-331.

Steels, L. (1997) The synthetic modeling of language origins, *Evolution of Communication* 1, 134.

Storkel, H. L. (2002) Restructuring of similarity neighbourhoods in the developing mental lexicon, *Journal of Child Language*. 29, 251-274.

Storkel, H.L. (2004). The emerging lexicon of children with phonological

delays: Phonotactic constraints and probability in acquisition. *Journal of Speech, Language, and Hearing Research* 47, 1194-1212.

Suomi, K, J. M. McQueen and A. Cutler (1997) Vowel Harmony and Speech Segmentation in Finnish. *Journal of Memory and Language* 36(3), 422-444.

Tamariz, M. (2008) Exploring Systematicity between Phonological and Context-Cooccurrence Representations of the Mental Lexicon. *Mental Lexicon* 3 (2), 259-278.

Thorn, A. S. C. and C. R. Frankish (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *J. Exp. Psych.: Learning, Memory, and Cognition* 31, 729-735.

van der Helm, P. A. (2000) Simplicity Versus Likelihood in Visual Perception: From Surprisals to Precisals. *Psychological Bulletin*, 126 (5) 770-800.

Vihman, M.M. (1996). *Phonological development: the origins of language in the child* Blackwell, Cambridge (1996).

Vitevitch, M.S. and P.A. Luce (1998). When words compete: Levels of Processing in spoken word perception. *Psychological Science*, 9, 325-329.

Watts, D. J. (2002) A Simple Model of Global Cascades. *Proc. Nat. Acad. Sci.* 99, 9 5766-5771.

D. Watts and P. Dodds (2007). *Influentials, networks, and public opinion formation*, *Journal of Consumer Research* 34, 441-458.

Werker, J. and C. Stager (2000). Developmental Changes in Infant Speech

Perception and Early Word Learning: Is there a link? in M. Broe and J. Pierrehumbert (eds.) *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge University Press. 181-194.

Zhang, Y, P.K Kuhl, I. Toshiaki, P. Iverson, J. Pruitt, E. B. Stevens M. Kawakatsu, Y. Tohkura, and I. Nemoto (2009), Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *NeuroImage* 46, 226-240.

Zipf, G.K. (1949) *Human Behavior and the Principle of Least Effort* Addison-Wesley.