The Ecclesiastes Principle in Language Change

Harald Baayen[ab], Fabian Tomaschek[a], Susanne Gahl[c], and Michael Ramscar[a]

[a] University of Tübingen, Germany
[b] University of Alberta, Canada
[c] University of Berkeley, USA

Running Head: The Ecclesiastes Principle
total word count: 8616

Corresponding author:
R. Harald Baayen
Seminar für Sprachwissenschaft
Eberhard Karls University Tübingen
Wilhelmstrasse 19
Tübingen
e-mail: harald.baayen@uni-tuebingen.de

1 Introduction

The history of mankind is characterized by constant change. One aspect of this change is the rise, spread, and demise in time and space of civilizations and religions. Another, perhaps more systematic, aspect of this constant change is that technological innovations, and thanks to these innovations, the amount of information available to agents in human societies has been increasing exponentially.

Societal changes lead to changes in language. Meibauer et al. (2004) and Scherer (2005) observed in a diachronic corpus of German newspapers that the use of multilexemic words increased over time, in response to increasing onomasiological needs for new technologies and a growing body of knowledge. Figure 1 illustrates for American English that over the last 200 years, the number of different words in use increased steadily. Counts were obtained by sampling, from the COHA corpus (Davies, 2010a,b), from each of the four registers distinguished in this corpus (news, fiction, non-fiction, and magazines), 1.5 million words and counting the number of different types. The numbers in Figure 1 were obtained by averaging the counts for 100 random samples taken for a given genre and decade. (For the first decade, the number of tokens was less than 1.5 million. Therefore, for the first decade of the 19th century, Figure 1 plots the total number of types in the corpus for the different genres as available for that first decade.) For each of the four registers, we observe a steady increase in the number of different words. (A regression model indicated that from 1860 onwards, the rate of increase was greater for magazines and news compared to non-fiction and fiction, and that the rates of increase were similar for magazines and news, and for non-fiction and fiction.)

One area where a growth is particularly visible is English personal names (Ramscar et al., 2013c). Names, both for people, natural species, artefacts and products, are part of the lexicon and co-determine the structure of the language. This is clearly documented for morphological structure, as pointed out for compounding by Scherer (2005) and for derivation by Lüdeling and Evert (2005). The accumulation of names has consequences that go beyond having a larger list of lexical items. For sound structure, the changing vocabulary comes with changes in lexical similarity and neighborhood density, which in turn has consequences for lexical processing (Gahl et al., 2012). Above the word level, the influx of new verbs such as *to ping* (in the sense of testing the availability of a computer on a network, which goes back to 1983) changes the quantitative structure of collostructural sets. Importantly, like common nouns, proper nouns have collocational structure, compare the names of three English renaissance composers, Thomas Knyght, Anthony Holborne, John Johnson, with the names of three species of trees, Ponderosa Pine, Narrowleaf Cottonwood, Rocky Mountain Juniper.

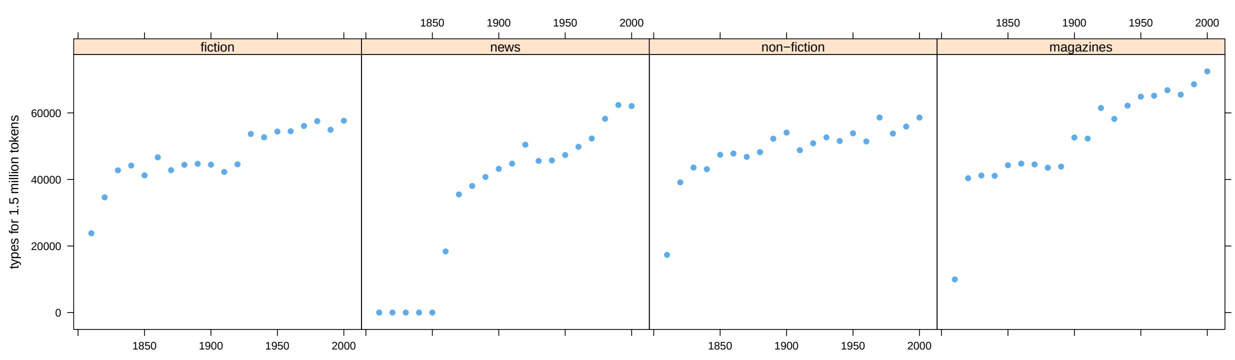

Figure 1: Number of different word types in 1.5 million tokens randomly sampled from the successive decades of American English in the COHA corpus, averaged across 100 samples.

Investigation of how different languages structure collocations for person reference shows that these follow specific conventions. Thus, the name grammar of pre-industrial England was more similar to the name grammar of modern Korean than the present-day name grammar of English (Ramscar et al., 2013c), which is changing rapidly. This can be seen in Figure 2, which presents US Social Security registration data on personal names. The horizontal axis represents the time period from 1880 to 2010. The vertical axis presents the entropy of the distribution of personal names, for males and females separately. The entropy measure quantifies the amount of uncertainty about a male or female name. Entropies are larger when there are more different names, and when these names have more similar probabilities. Thus, a greater entropy indicates, informally, that it is more difficult to guess a name and that it will take more time to retrieve a name from memory.

Figure 2 shows a sharp rise in name entropy that may have its onset around the end of the second world war. By the beginning of the eighties, there was a greater variety of names in use than ever before (female names especially), and by 2010, the entropy of names had skyrocketed. This means that, for an individual speaker born around the end of second world war, name vocabulary will have increased exponentially across her lifetime.

The goal of this study is to present an overview of some of our findings on the consequences for the grammar and lexical processing of having to accommodate the ever growing onomasiological needs of modern societies. Our main focus will be at the micro-level of changes over the lifetime of speakers and their processing consequences, but we will link our findings about aging and language to macro-changes over generations of speakers where possible.

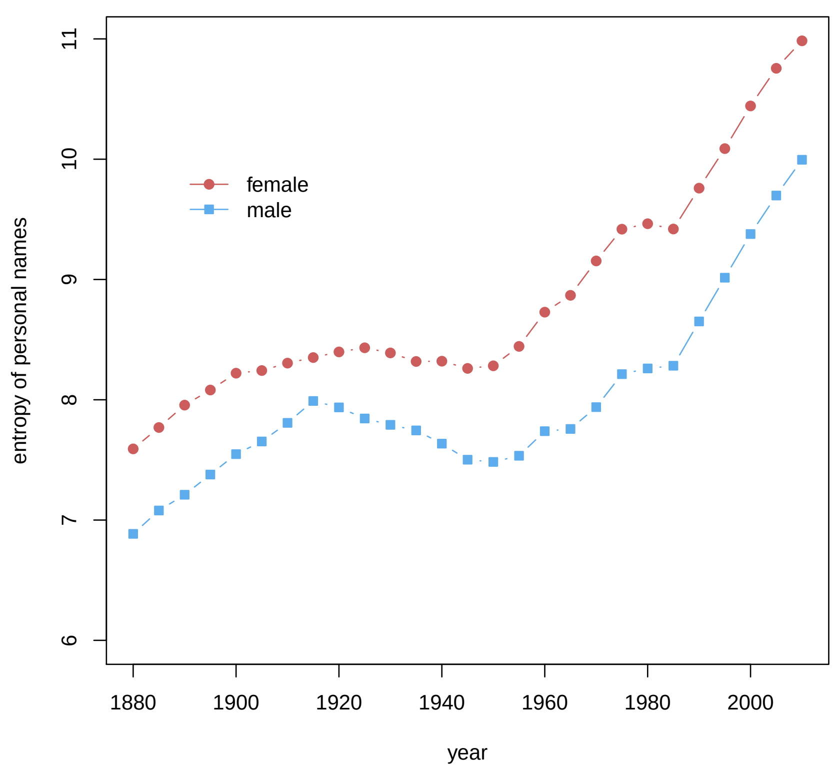

Figure 2: The increase in entropy of US given names 1880–2010.

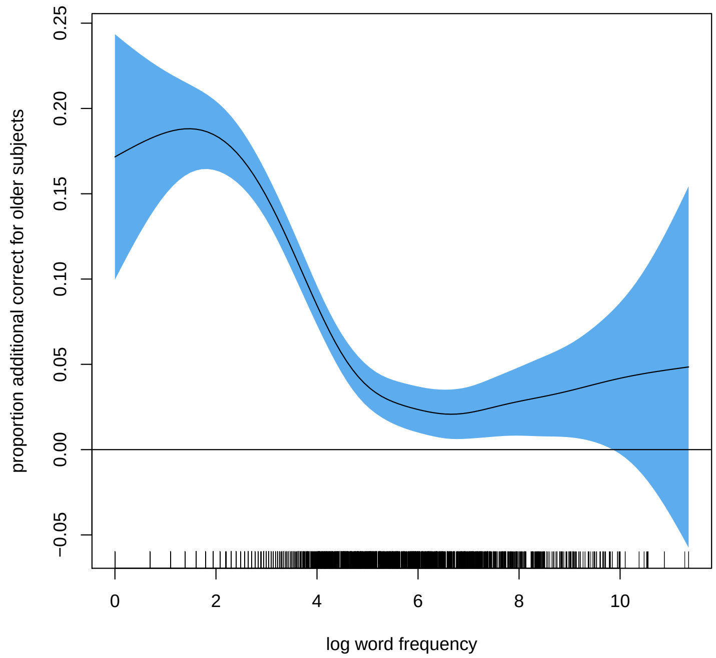

Figure 3: Difference in the proportion of additional correct lexical decisions that older adults have compared to young adults in the English Lexicon Project.

2 Language change over the lifetime

As speakers proceed through life, they do not only meet more and more people with different names, but they also encounter more street names, place names, brand names, and specialized vocabulary for domains in which they acquire expert knowledge. That older adults know more words is apparent from lexical decision studies. An analysis of the accuracy data in the English Lexicon Project (Balota et al., 2007) for monosyllabic monomorphemic words shows that the accuracy of older adults deciding whether a letter sequence presented to them is a word of their language is greater than the accuracy of younger adults. Figure 3 presents the difference curve for the accuracy of old versus young respondents. Across the full range of frequencies, older adults are more accurate, except for the highest-frequency words where estimates are uncertain due to data sparsity (the rug in Figure 3 thins out at the right-hand side of the horizontal axis). For words with a log frequency in the British National Corpus lower than 5, we see a dramatic increase in performance for older versus younger subjects. A recent crowd sourcing study (Keuleers et al., 2015) comfirmed the same pattern for a wide range of ages. Accuracy increases systematically over the lifetime, reaching a correctness score of about 75% by the age of 60. These results fit well with an analysis of the switchboard corpus (Meylan and Gahl, 2014), which revealed that as compared to younger speakers, older speakers exhibit greater vocabulary richness, and share fewer common words with their interlocutors in conversation.

What these studies show is that over the lifetime, we slowly but steadily increase our mastery of the vocabulary. This mastery, however, is not restricted to knowing more words, it extends to collocational knowledge and to articulatory fluency.

The increase in collocational knowledge becomes apparent when results obtained with the paired

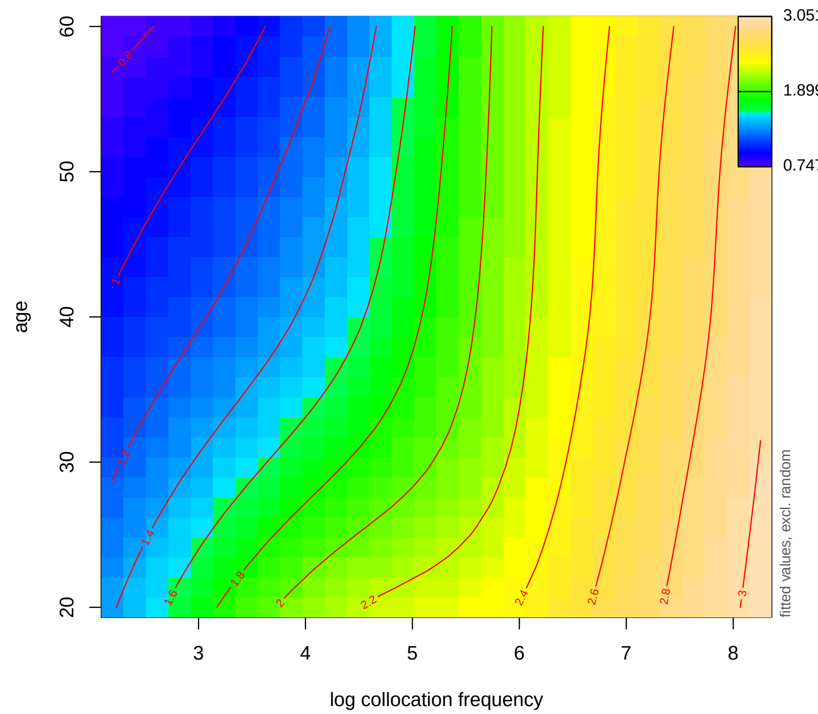

Figure 4: Correctness scores in the paired associate learning task as a function of log collocation frequency and age. Deeper shades of blue indicate lower correctness scores, warmer shades of yellow indicate higher scores.

associate learning task (PAL) are inspected (Ramscar et al., 2014, 2013b). In this task, subjects are first presented with pairs of words, some of which are associated to some extent (such as *baby* and *cries*), and some of which are arbitrary pairs (*obey* and *inch*). Subsequently, subjects are presented with the first word (*baby* or *obey*) and are asked to say out loud the second word (*cries* or *inch*). Unsurprisingly, subjects make more errors for hard pairs such as *obey* and *inch* than for easy pairs such as *baby* and *cries*.

What is at first blush surprising is that performance on this task decreases with age. In the psychological literature, the task has established itself as a diagnostic for cognitive decline over the lifetime. However, this body of work has never paid attention to the linguistic properties of the stimuli used. It turns out that collocation frequency (estimated here using counts of google hits for the word pairs) is a strong predictor of performance. Figure 4 presents the regression surface for PAL performance with log collocation frequency and age as predictors, using data reported in des Rosiers G. and Ivison (1986), fitted with a generalized additive mixed model (Wood, 2006; Baayen, 2014). Darker shades towards blue indicate lower scores, indicating less accurate performance. More yellow and reddish shades indicate more accurate performance. The contour lines connect points for which performance is identical. For a fixed value on the vertical axis, the contour lines represent the effect of collocation frequency. Across all ages, performance improves with frequency. When we fix collocation frequency, we find a strongly downwards sloping surface for low values, but hardly any effect for the highest frequencies. To see for which age the effect of collocation frequency is strongest, we count contour lines from left to right. For age 20, we cross 9 contour lines. For age 60, we count

11 contour lines. This indicates that the oldest subjects are most sensitive to collocation frequency, whereas the youngest subjects are the least sensitive. This finding is consistent with the fact that older subjects have had more experience with the language, and have as a consequence become more sensitive to lexical co-occurrence probabilities. In other words, with age, the knowledge of the language increases. However, it is also remarkable that the older subjects perform worse for the lower co-occurrence frequencies than the younger subjects. We will return to this issue below, where it will become clear that this is an inevitable consequence of learning.

The reasons why vocabulary increases across the lifespans are obvious. With age comes experience, and the more experiences of the world we learn to discriminate and encode in the language signal when communicating about these experiences, the more complex the code must become. To see this, consider a simple situation in which there are four experiences that constitute the universe of what we communicate about. In this case, four two-bit codes (10, 01, 00, 11) suffice to discriminate between these four experiences. When there are 10,000 experiences, we need binary codes with 14 bits to distinguish between all of them. Similarly, when onomasiological needs increase, a language will need to find ways to properly differentiate between what we seek to communicate. Of course, one could resort to phrasal circumlocution, but this is likely to be a wasteful solution energy-wise (cf. Zipf's law of abbreviation Zipf, 1949). Languages such as Chinese and Vietnamese started out with lexicons in which monosyllabic words were the norm (see, e.g., Arcodia, 2007). Given severe phonotactic constraints on what constitutes a usable monosyllabic word, the number of possible forms is soon exhausted. The code can allow for homonymic ambiguity thanks to contextually driven ambiguity resolution. But even this has its limits, and the majority of words in Chinese and Vietnamese are now bisyllabic compounds.

Instead of, or complementary to, resorting to compounding (or other less productive word formation processes), it is possible to implement discrimination by modulating the fine phonetic detail of articulation. 'Subphonemic' onomasiological discrimination is well established for so-called suprasegmentals such as tone (familiar from Chinese or Vietnamese), stress (familiar from English), and acoustic duration (such as the three durational contrasts in Estonian). However, other more subtle subphonemic contrasts have recently come to light for English and related Germanic languages such as Dutch. Gahl (2008) reported systematic differences in acoustic duration for English homophones such as *time* and *thyme*, Kemps et al. (2005a) and Kemps et al. (2005b) observed systematic differences in acoustic durations between stems in isolation and stems in inflected or derived words, and Plag et al. (2014) found systematic differences in acoustic duration for the English suffix *-s* that co-varied with its morphological function.

The subtlety of these findings underline how daunting the task is of mastering the motor control required for articulation. Here, as for any other motor skill, be it playing tennis or playing the violin, improvement comes with practice. We can see the effects on practice over the lifetime by investigating the fine details of articulatory trajectories for words of different frequency. The more frequent a word is, the more opportunities a speaker has had to practise her motor system on the articulation of that word, and the more likely it is that we can observe discriminative differentiation as a function of experience. Thus, by studying words of different frequency with respect to how articulatory gestures are executed, we can gain insight in changes that are likely to take place within a given word as a speaker becomes more practiced in uttering that word.

There are several ways in which the trajectories of articulators can be measured. Of a wide range of techniques, the simplest and most widely used are ultrasound (Gick, 2002) and electromagnetic articulography, EMA (Schönle et al., 1987). Tomaschek et al. (2014) used EMA to study how articulation covaries with frequency. They glued small metallic sensors to subject's tongue and lips, and placed the subject in an electromagnetic field generated by an EMA system. During speaking, the movements of the sensors give rise to changes in this field, which are registered and used to

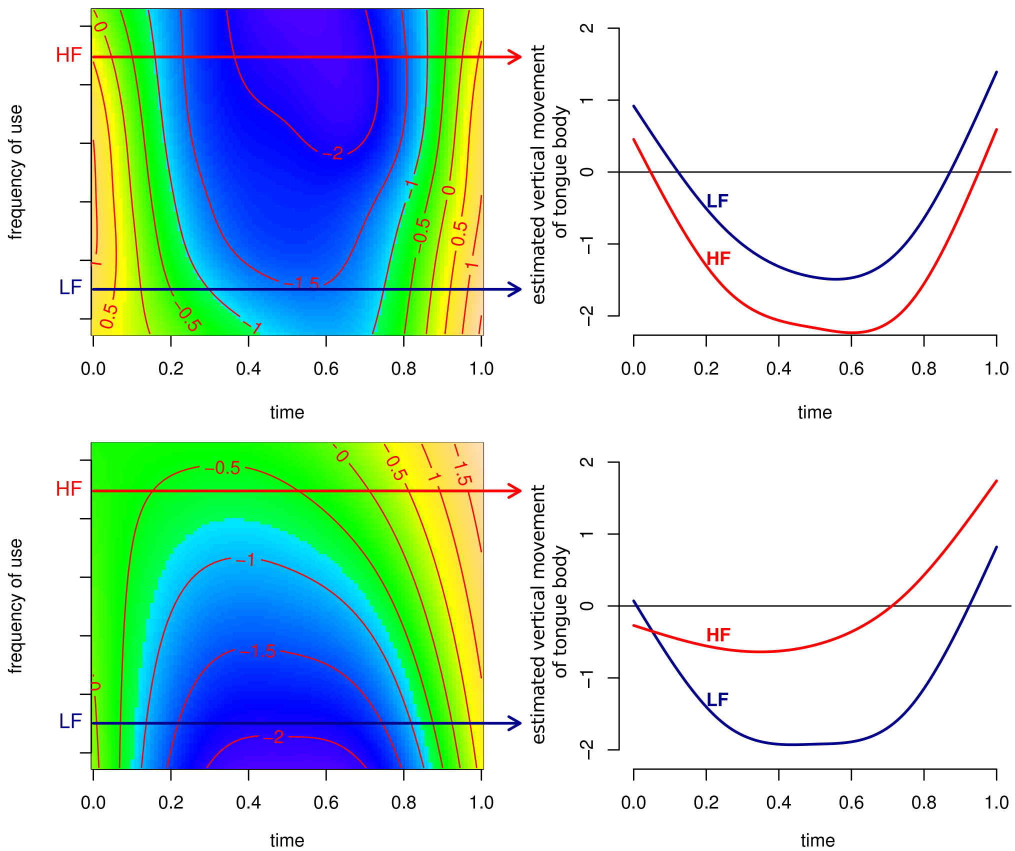

Figure 5: Vertical displacement of the tongue body sensor during the articulation of German *a* as a function of frequency of use and (normalized) time. Darker colors indicate lower positions of the tongue body sensor.

calculate time series of X, Y, and Z coordinates for each of the sensors. When time-locked with the acoustic signal, the part of the time series that corresponds to an event of interest, such as the articulation of a vowel, can be identified, and then subjected to statistical analysis.

Figure 5 presents the trajectories of the tongue body sensor during the articulation of the vowel /a:/ in German verbs and nouns which varied in their frequency of use. Horizontal axes represent time (normalized between 0 and 1). In the left panels, the vertical axis represents log-transformed frequency of use. In the right panels, the vertical axis represents vertical displacement of the same sensor. The left panels show regression surfaces estimated with the help of generalized additive mixed models. The contour lines connect points defined by time and frequency for which the vertical displacement of the tongue body sensor is estimated to be the same. The color coding indicates the direction of the displacement: darker colors indicate further down, lighter colors, further up. The curves presented in the right panels highlight two specific trajectors, one for a specific lower-frequency, and one for a specific higher frequency.

The upper panels show that the tongue body sensor moves further down when producing the vowel /a:/ when the word has a higher frequency of occurrence. In other words, for higher-frequency words in which the next syllable realizes the third person plural ending (an apical /n/), a higher frequency affords a more precise and distinctive articulation of this low vowel. The lower panels show the pattern in reverse, when a /t/ realizes the second person plural inflection. The more frequent this inflectional variant is, the earlier the tongue starts preparing for the articulation of the (laminal) /t/.[1]

An influential hypothesis, first advanced by Aylett and Turk (2004), holds that higher-frequency, and from the perspective of information theory less informative, words would be articulated with more centralized vowels and shorter durations (Aylett and Turk, 2006) in order to keep the rate at which information is transmitted relatively constant (smooth) in a channel with limited transmission capacity. However, we do not replicate these findings in the articulatory domain. When an apical inflection is following, speakers learn over time to realize a more distinctive *a*. When a laminal inflection is following, we see more coarticulation with the upcoming suffix as frequency increases. In other words, what experience makes possible is to speak in such a way that the fine phonetic detail of how a segment is realized becomes discriminative between higher and lower-frequency words (see Kuperman et al., 2006; Ferrer-i Cancho et al., 2013, for further empirical and also mathematical problems with the smooth signal redundancy hypothesis).

An increase in subphonemic discrimination is also visible in real time in the realization of English vowels over the lifetime. Figure 6 presents the changes in the first and second formants of English monophthongs as spoken by 11 speakers in the Up corpus. This corpus, which is described in Gahl et al. (2014), is based on five films from the film series known as the "Up" series of documentary by director Michael Apted. These films follow a set of individuals at seven year intervals over a period of 42 years. The most recent material included in the corpus shows the participants at age 49. The corpus is based on utterances of at least 20 seconds or more of uninterrupted speech. The corpus comprises 250 utterances for each of the eleven documentary participants, to a total of 21,328 word tokens representing 2463 unique word types. The changes in the formants graphed in Figure 6 are those predicted by a linear mixed-effects model that includes speaker and word as random-effect factors. Darker shades of gray indicate realizations later in life. What Figure 6 shows is that, on the whole, the vowel space expands with age when speaker and lexical variability is controlled for.

This pattern of change is interesting in the light of the idea that changes in the speech of older adults are adaptive. Biological and physiological changes occur throughout the lifetime. Hearing loss, as well as changes such as atrophy of the vocal folds and calcification of the laryngeal cartilages may give rise to adaptive strategies that, in all, are remarkably successful (Hooper and Cralidis, 2009). The speakers at the latest point in time available in our sample, however, were in their forties, and it is therefore unlikely that the changes that we observe here are due to adaptation to hearing loss.[2] Rather, it is much more likely that this pattern reveals a genuine continuity of language development and change in healthy aging, such that over time, speakers become ever more proficient in producing increasingly discriminative speech signals.

What these examples show is that obtaining full mastery of the language as used in complex modern societies is a process that extends over the lifetime: As we grow older, we master more words, we become less prone to learn nonsense, our vowel space expands, and we articulate segments with greater skill. There is an interesting parallel in the culture of knowledge acquisition and knowledge extension. Jones (2005) reported that the greatest achievements in science are no longer the preserve of the young. Using data on Nobel Prize winners and great inventors, he observed that by the end of the 20th century, great inventions were made 8 years later on average than at the beginning of the century. Econometric modeling suggests that this is not just an effect of an aging population, but an effect due to productivity starting later in life — the accumulation of knowledge across generations

[1] These findings are of particular interest in the light of the scarcity of frequency effects in speech production for regular inflected words. The model of Levelt et al. (1999), for instance, rules out such frequency effects explicitly, and the evidence from both speech errors (Stemberger and MacWhinney, 1986) and behavioral paradigms (see,e.g., Bien et al., 2011; Tabak et al., 2010) is ambiguous.

[2] According to Walling and Dickson (2012), approximately one out of three adults in the age range from 61 to 70 years suffers from hearing loss. This proportion increases to 80% for those older than 85.

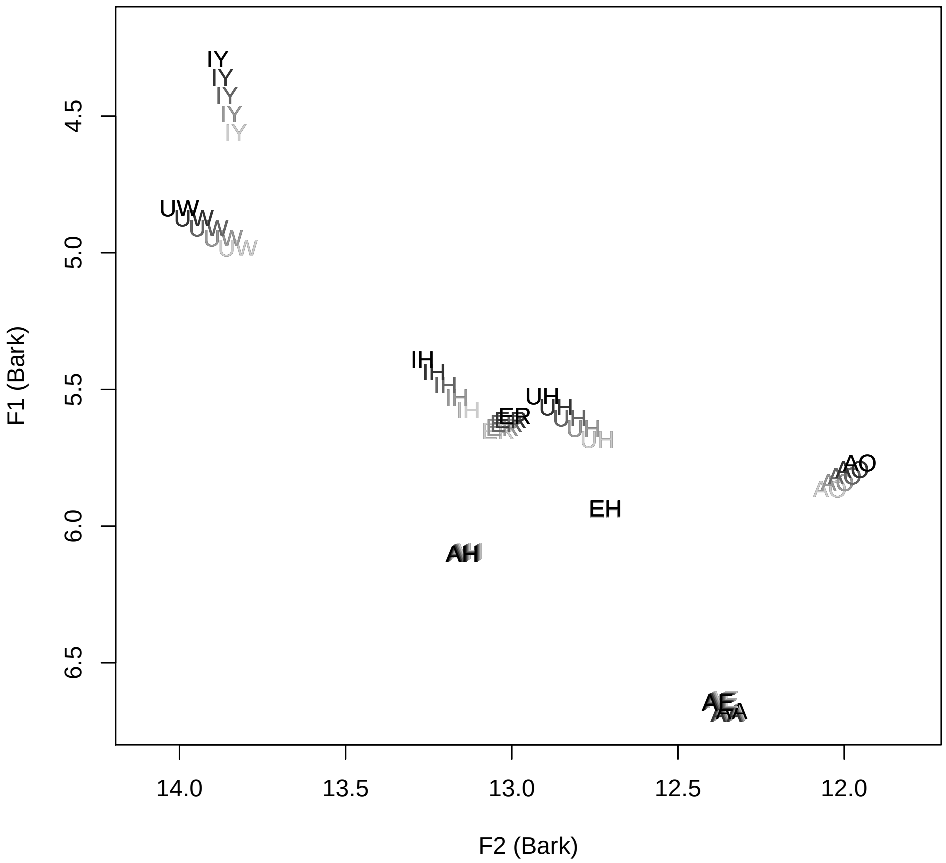

Figure 6: Expansion of the vowel space with age: The Up (Michael Apted) corpus. Darker shades of grey represent later points in time.

forces innovators to seek more education over time. Although the acquisition of the skill of speaking is not one marked by great landmark achievements, it seems likely that, compared to for instance early modern English, the lexis of present-day English is more complex and requires more time to master.

3 The Ecclesiastes Principle in language change

In the third century BC, the philosopher and wisdom teacher Qohelet wrote:

> For in much wisdom is much grief:
> and he that increaseth knowledge increaseth sorrow
> (Ecclesiastes 1:18, translation King James Version)

This characterization of the human condition applies straightforwardly to human learning. The accumulation of knowledge does not come for free. We refer to this as the Ecclesiastes Principle.

We noted above that name entropy has been increasing rapidly over the last 60 years (see Figure 2). This finding sheds new light on the well known difficulties we often experience with remembering people's names. As we age, and continue to sample from an ever increasing vocabulary of names, this problem is exacerbated. Several aspects of this phenomenon that are worth considering:

First, by the Ecclesiastes Principle, it is unavoidable that as we know more names, finding a specific name becomes more difficult. Looking up a name in a small telephone directory is faster than looking up a name in a large directory. Thus, as we go through life, and get to know more people, the entropy of the names we know increases, and as a consequence greater processing times are inevitable.

Second, the name finding problem is exacerbated by the way naming practices in English have developed over the second half of the twentieth century. Before the industrial revolution, *John*, *Thomas* and *William* were the most popular names for boys, accounting for some 50% of all different names in use (Galbi, 2002). For girls, *Mary, Elizabeth* and *Anne* represented the top ranks in the Zipfian name distribution. Further differentiation between the different Williams and Annes is achieved with discriminators for occupation (Smith), ancestry (Johnson), or place of origin (London). Ramscar et al. (2013c) point out that this results in a system that is very efficient for retrieving names.

Consider, by way of example, a situation in which 900 people need to be distinguished, and assume, for simplicity, that all names are equiprobable. If each individual has a unique name, name uncertainty (gauged by Shannon's entropy) is maximal, and equal to $\log_2(900) = 9.8$. If we have 30 first names and 30 family names, which also allows us to discriminate between 900 people, the entropy of retrieving the first name is halved, $\log_2(30) = 4.9$, and the same holds for the entropy of retrieving the family name. Since the industrial revolution, English has been trending towards the latter situation. As a consequence, the name finding problem has been increasing. In the light of the census data shown in Figure 2, a 70-year old in 2010 is faced with a much higher variety in names compared to when this same person was a 20-year old in 1960. Ramscar et al. (2014) report simulation studies suggesting that the joint consequence of encountering more names, and more diverse names, over a period of 50 years results in a processing delay of no less than some 150 ms. This example illustrates that, independently of the Ecclesiastes Principle, societal changes can modulate language change in a way that is dysfunctional for its speakers. Ramscar et al. (2013c) note that naming systems need not be dysfunctional in this way, and mention Korean as an example of a language with an efficient name grammar.

The Ecclesiastes Principle has more subtle, but no less far reaching consequences when we consider the details of lexical learning. Above, we examined the performance on the paired associate learning task as a function of collocation frequency and age. We observed that older speakers reveal greater sensitivity to collocation frequency, which fits with our hypothesis that language proficiency increases as experience accumulates over the lifetime. What we have not yet discussed is why it is that older speakers perform less well than younger speakers on the pairs with lower collocation frequencies.

To see why this happens, we consider a highly simplified example of collocational learning in which a Rescorla-Wagner network (Rescorla, 1988; Ramscar et al., 2010; Baayen et al., 2011; Baayen and Ramscar, 2015) is required to learn six word pairs, as shown in Table 1. (Below, we provide further details on this computational model.) During the first training phase, we present the model with 10 examples each of the pairs *baby - window* and *obey - inch*. The model is instructed to learn to predict the second word given the first word. Following this training phase, the model is presented with further word pairs in which *baby* and *obey* are the first words. The word pair *baby - cries* has a frequency of 80, *baby - sleeps* has a frequency of 40, and the two pairs with obey have a frequency of 60 each. Thus, the total number of occurrences of *baby* as first word is equal to that of *obey* as first word. The order in which pairs were presented was chosen randomly. Each presentation of a word pair constitutes a learning event at which the model is given the first word as input cue, and is

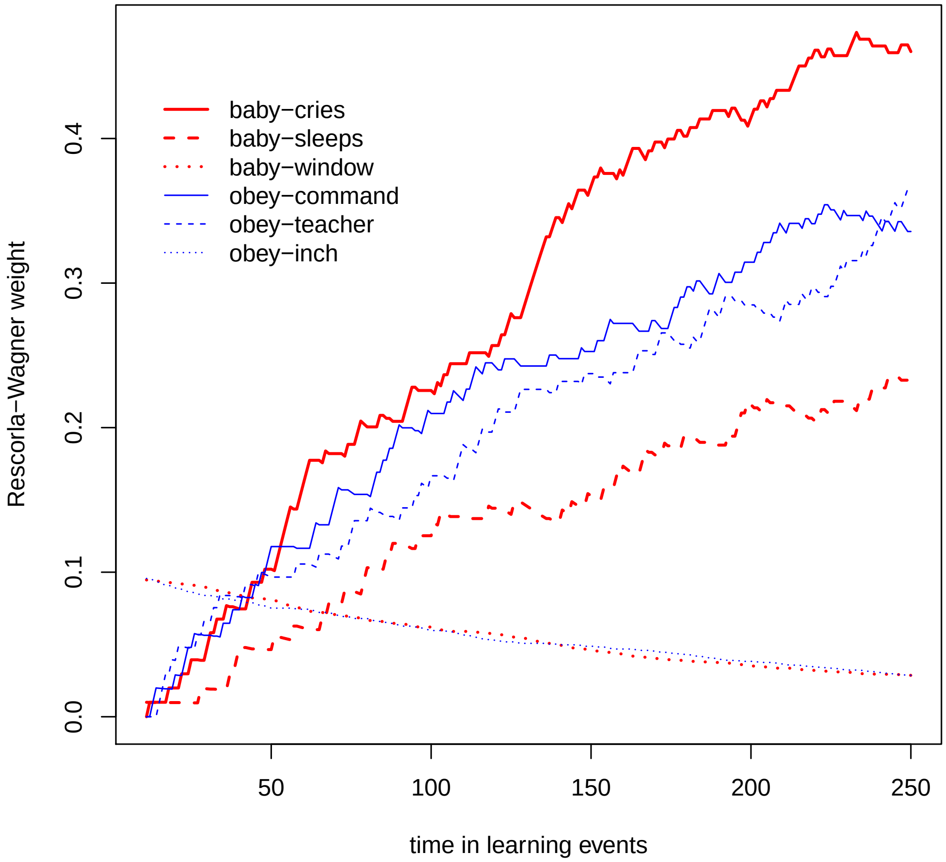

Figure 7: Development in a Rescorla-Wagner network of the weights from the first word to the second word for the 6 word pairs in Table 1.

asked to predict the second word as the output outcome. If the prediction was correct, the weight on the connection from the first to the second word is strengthened, otherwise, it is weakened.

	Cues	Outcomes	Frequency	Training Phase
1	baby	window	10	A
2	obey	inch	10	A
3	baby	cries	80	B
4	baby	sleeps	40	B
5	obey	teacher	60	B
6	obey	command	60	B

Table 1: Data for the learning simulation with six collocations.

Figure 7 illustrates how the weights on the connections from the first to the second words develop during the second training phase, from learning event to learning event. The solid and dashed lines show the development of the weights for the four frequent pairs. The pair *baby* - *cries* occurs twice as frequently as the pair *baby - sleeps*, which explains why the curve for the first pair is located above that of the second pair. The two blue curves represent pairs with identical frequencies midway between those of *baby - cries* and *baby - sleeps*. Unsurprisingly, we find them in between the two red curves. The dotted lines represent the two pairs that were presented only 10 times during the initial training phase, after which they were never presented again. These pairs show the weakest weights. Thus, after about a hundred trials and continueing up to the end of learning, the model reflects the collocation frequency effect that we observed for the paired associate learning task. The model also captures the effect of age. After 100 trials, the model represents a young adult. After 250 trials, the model reflects more advanced learning. For the young adult, the effect of collocation frequency, although already present, is not as differentiated yet to the extent that becomes visible with further learning.

What is important is that the weights on the connections from *obey* to *inch* and from *baby* to *window* decrease as the other pairs are learned (see also Ramscar et al., 2013b). This is a straightforward consequence of discrimination learning. Everytime that the word *obey* occurs in a learning event, the model considers all possible outcomes, and adjusts weights upward if the outcome is present (which will happen for *obey - command* and *obey - teacher*), but adjusts them downward when the outcome is not present (*obey - inch*). In other words, the model learns that it can expect *command* or *teacher* given *obey*, and it also learns that it doesn't make sense to expect *inch*. Because both *obey - inch* and *baby - window* are downgraded the same number of times, as both first words occur in the same number of other collocations, their curves are the same, modulo how the random numbers turned out in the simulation.

Thus, the Ecclesiastes Principle manifests itself in the context of learning as a force prohibiting the learning of novel knowledge *if and only if* that novel knowledge does not make sense given prior experience.[3] We think this same force may serve to speed the demise of words that are in the process of becoming obsolete. It is not only that the contexts in which words such as *telegraph* or *walkman* were once used will become increasingly rare, but the lexical collocates that were once predictive of these words will increasingly loose this predictivity where they continue to be encountered in other contexts.

We discussed earlier how the changing onomasiological demands of increasingly complex modern societies give rise to continuously increasing name and word finding difficulties. One possible adaptation that offers a way of sidestepping this problem is to revert to favoring pronouns instead of names. Two predictions follow. First, over time, the use of pronouns should increase in the speech community. Second, over the lifespan, the use of pronouns should increase likewise.

The first prediction is well supported by an inspection of the frequencies of use of the pronouns *he, she, they* in the samples in the COHA corpus (Davies, 2010b) over two centuries of American English. Figure 8 graphs their by-decade relative frequencies. The left panel shows the pronoun frequencies divided by the number of tokens in COHA for the respective decades. The right panel presents the pronoun counts, but now normed against the corresponding frequencies of the definite article *the*. The right-hand plot is more informative as it specifically compares pronominal definite reference with non-pronominal definite reference.[4] For each pronoun in the right hand panel of

[3] In other words, as knowledge accumulates over the lifetime, it becomes increasingly difficult to learn nonsense, in the form of word pairs that simply don't make any sense: For such pairs, connection strengths have been driven towards zero during learning, or may even have become negative. At the same time, as the number of words that a speaker knows increases, such a speaker will perform with greater accuracy in the lexical decision task, but due to the greater complexity of her lexicon, this speaker will necessarily respond more slowly. The crucial point here is that response speed and accuracy in paired associate learning should not be considered in isolation, but rather in the context of the language system as a whole.

[4] By-decade corpus size and by-decade frequency of *the* are highly correlated ($r = 0.987$), but a regression analysis with a quadratic polynomial supports a slight leveling off of the counts for *the* for the more recent decades.

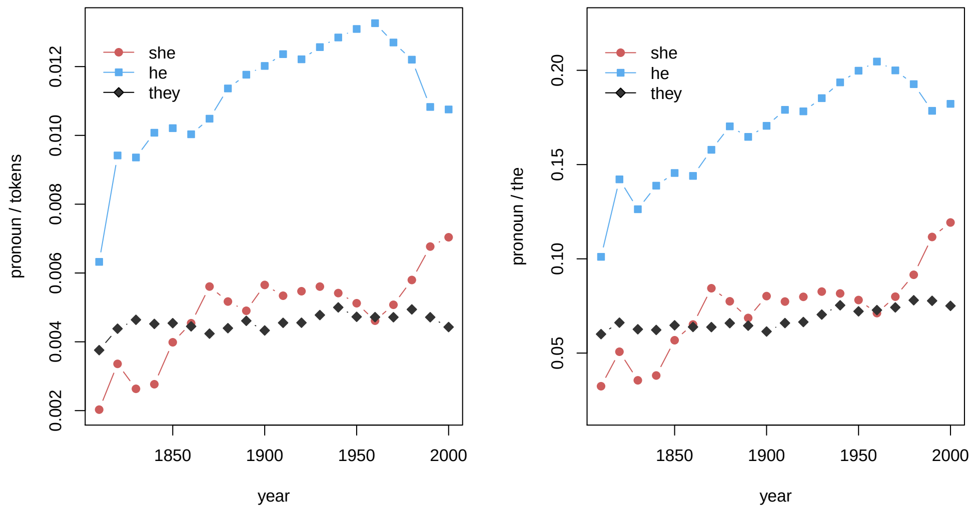

Figure 8: Development of the use of third person pronouns *he, she, they* in American English as gauged by the COHA corpus. Left: counts normed by tokens; right: counts baselined against the frequency of *the*.

Figure 8, we see an increase over time, which is small but well-supported overall for *they* (r = 0.89, t(18) = 8.27, p < 0.0001) and more pronounced for *she* (r = 0.867, t(18) = 7.39, p < 0.0001) and *he* (r = 0.884, t(18) = 8.04, p < 0.0001). In the second half of the twentieth century, we may be observing the generic use of *he* for both genders giving way to more gender-specific use for males, which allows frequency to accrue to *she*. The global pattern is one in which writers are increasingly favoring the use of anaphora instead of definite referring expressions. It seems likely that the more rapid increase in the use of *he* and *she* compared to *they* reflects the exponential increase in the number of different personal names as compared to the more modest increase in types in other parts of the vocabulary.

The Ecclesiastes Principle predicts that a similar trend should be visible across the lifespan of individual speakers. Interestingly, Hendriks et al. (2008) report that older adults use more pronouns than younger adults, exactly as expected. Their interpretation of this finding, however, is very different from ours. Hendriks et al. (2008) interpret their results as indicating that older speakers' cognitive capacities are in decline. By contrast, just as the increase in use of pronouns in American English is not a symptom that the language is somehow terminally ill, the increased use of pronouns in older speakers is a sensible adaptive strategy to manage the increased knowledge that is inevitably accrued in healthy aging. The view of aging as an inevitable process of cognitive decline is incompatible with the present findings, which consistently indicate that with age adults become more proficient speakers of their language. Surprisingly, the myth of cognitive decline (Ramscar et al., 2014) is propagated in the psychological literature in total ignorance of the Ecclesiastes Principle, reflecting a general attitude to the elderly that in the gerontological literature has been dubbed *ageism* (in parallel to racism and sexism, see Palmore and Manton, 1973).[5]

[5] Ageism is also apparent in the claim that vocabulary size decreases with age, starting around the age of 40 (Singh-Manoux et al., 2012), which, as we have demonstrated above, is incorrect. Increased use of pronouns has also been taken to provide evidence of loss of cognitive skills (Hendriks et al., 2008). Below, we offer a very different perspective on this finding. The longer response times of the elderly in chronomtric tasks would at first blush also

4 Entrenchment and the Ecclesiastes Principle

Frequency of occurrence is widely used as a measure of entrenchment in memory (see Hilpert, this volume). However, simple frequency counts do not take into account the effects of co-learning and the costs that accrue with the accumulation of knowledge. Statistical measures as used in studies of lexicogrammatical attraction (Allan, 1980; Stefanowitsch and Gries, 2003; Ellis, 2006; Schmid and Küchenhoff, 2013) take into consideration that words are used as part of a system (see also Ellis, this volume). However, the 2×2 contingency tables on which these measures are calculated require simplified binary contrasts that do not do full justice to the complexity of the language system. The same holds for behavioral profiles (see (Gries and Divjak., 2009) and Hilpert, this volume), which call attention to the many paradigmatic relations in which a construction may participate, but that will tend not do consider such relations across a multiplicity of constructions. Furthermore, the burstiness of words and their non-uniform dispersion across documents potentially have consequences for entrenchment that go beyond what can be gauged with simple frequency counts.

This leaves the analyst with two options. One option is to complement frequency counts with a wide range of other measures, such as burstiness, dispersion, age of acquisition, conditional probabilities given preceding or following words, and multiword probabilities (Bannard and Matthews, 2008). Baayen (2011b) showed, using multiple regression, that when a wide range of variables correlated with word frequency is taken into account, there is very little variance left for word frequency to explain. Simple counts isolate units such as words from the system of which they are part. The more measures that probe the system are taken into account, the less useful bare frequency counts become.

Another option, which we pursue here, is to use measures that reflect the consequences of the accumulation of knowledge under the constraints of discrimination learning. Baayen (2011a) proposed to use the activation measure of the naive discriminative reader model presented by Baayen et al. (2011), and discussed similarities and differences with other measures such as ΔP (Allan, 1980) and distinctive collexeme strength (Gries and Stefanowitsch, 2004). Recent developments in naive discriminative learning theory (Milin et al., 2015; Baayen et al., 2015b; Shaoul et al., 2015) offer a new measure that not only is highly predictive for lexical processing as gauged with the visual lexical decision task, but that we believe is also particularly promising as a measure of lexical entrenchment.

However, let us first address the question why discrimination learning might provide us with quantitative measures that are more informative than straightforward counts of frequency of occurrence. To answer this question, consider a simple priming experiment in which pictures precede words that subjects have to read out loud. Marsolek (2008) showed that when a picture of a grand piano is the prime for the word `table`, subjects are slower compared to a control condition with a picture that does not share visual features (such as having a large horizontal flat surface, and having legs) with the word's denotatum (e.g., a picture of a bowl with oranges). Marsolek named this phenomenon *anti-priming*, and explains it with the principles of discriminative learning. Upon seeing a grand piano, the strength of the link between the feature of having a large flat surface is strengthened to the musical instrument, and at the same time associations to other objects, including tables, are weakened. Because oranges do not share many features with tables, they do not 'negatively prime' the target `table` to the same extent as does the grand piano. The consequences of this constant recalibration of the strengths of connections between features (henceforth cues) and

suggest deteriorating performance parallelling the hardships of the failing body in old age, until it is realized that their accuracy is so much higher. It is worth noting that declines in cognitive faculties as a consequence of neurodegenerative disease is correlated with age of retirement (Dufouil et al., 2014): The earlier the age of retirement is, the earlier the onset of dementia, consistent with the hypotheses of 'use it or lose it'. Under normal conditions of cognitive and social stimulation, there is no reason to suppose that mental capacities decline with age.

outcome classes (tables, oranges, pianos) are profound (see Ramscar et al., 2010, 2013a, for detailed discussion). For corpus linguistics, the implication is that simple counts are not precise enough.

For example, the word *great* can be followed by many other words (*care, deal, story, about, for, if, on, used, ...*). Whenever *great* is followed by *story*, the link between *great* and *story* is strengthened, while the link between *great* and all other words that have been encountered following it are weakened. As a consequence, simply counting frequencies and co-occurrence frequencies will not do justice to the constantly ongoing recalibration of the language system with respect to the words that could have appeared, but didn't. Collostructional analysis and distinctive collexeme analysis (Stefanowitsch and Gries, 2003; Gries and Stefanowitsch, 2004) as well as the Delta-P measure (Allan, 1980) advocated by Ellis (2006) take aspects of this constant recalibration into account, but do so only for cross-tabulations of constructions and lexical items (e.g., double-object versus prepositional object constructions, and *give* versus all other verbs used with datives). The recalibration that takes place across all lexical items and across all potential constructions in which these lexical items participate is not taken into account (see also Baayen et al. (2015a) for a critical discussion of frequency of occurrence as a predictor for language processing).

This is where discriminative learning becomes useful. It allows the researcher to bite the bullet, and to systematically work through all positive and negative adjustments of lexical associations given a corpus. The simplest way of approaching the constant recalibration of the web of lexical relations is to study word to word predictivity. Of course, such a simple approach simplifies the intricacies of the language system substantially, but our current strategy is to see how far simple solutions can take us. Reassuringly, Shaoul et al. (2015) show that this approach sets up a semantic vector space, and Baayen et al. (2015a) show that measures obtained by corpus-based discrimination learning are excellent predictors of various experimental measures gauging lexical processing (see also Hendrix, 2015, for the reading of compound words).

In what follows, we will not focus on providing further empirical justification of the discriminative stance, but rather, we will start exploring what a discrimination-based statistic may reveal about language change. Before proceeding, however, a brief introduction to the basics of discrimination learning will provide a better basis for understanding the measure with which we will probe our corpus data.

Naive discrimination learning is implemented with a simple network with two layers of nodes, an input layer with *cue* nodes, and an output layer with *outcome* nodes. Each cue has a connection to every outcome. Each of these connections comes with a weight that specifies how well a given cue supports a given outcome. These weights are estimated by applying the learning equations of Rescorla and Wagner (1972). These learning equations specify how the weights should be adjusted for a given set of cues and outcomes present in a learning event. Weights on the connections from these cues to these outcomes are strengthened, while the weights to other outcomes that are not present in the learning event are weakened. The values of the weights thus change with each successive learning event. The more often an outcome is present in learning events, the stronger the weights on its incoming connections will be. However, because the Rescorla-Wagner equations also implement cue competition, the strengths of the weights are co-determined by the other cues in the learning events.

An important property of the NDL model is that it scales up to hundreds of millions of learning events extracted from large corpora. Furthermore, whenever corpora contain order information, be it the order of sentences in text, or the ordering of subcorpora in historical time, the learning events presented to the model can be sequenced such that this order information is respected. In other words, given appropriate data, the model can be used to study learning over the lifetime as well as language change in historical time.

Let us explore the connection between discrimination learning and language change in historical time a little more closely by examining lexical entrenchment, which in standard approaches is typically

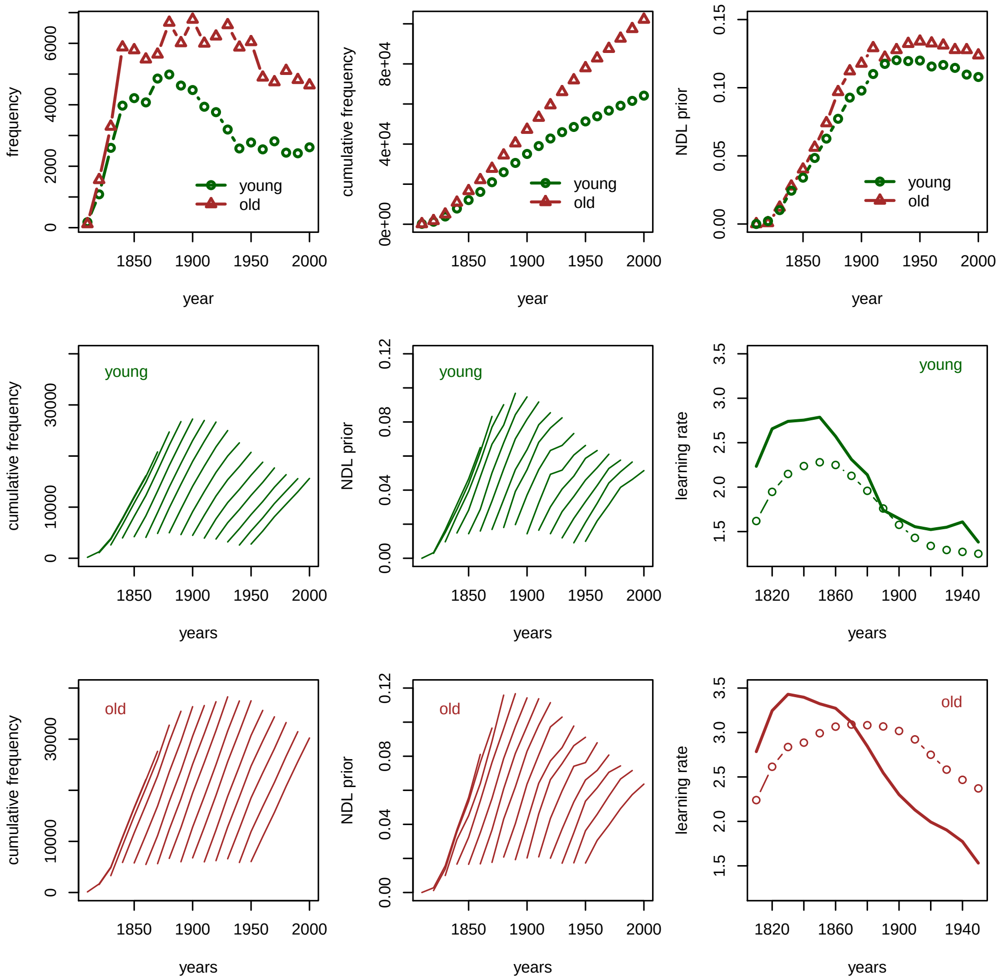

Figure 9: Changes in entrenchment for the words *young* and *old* over the lifetime, in American English in the period 1800-2000 (data from COHA).

gauged by means of frequency of occurrence. Within the framework of naive discriminative learning, a measure developed by Milin et al. (2015) is of particular interest for assessing entrenchment. This study found the median absolute deviation (MAD) of the weights on the connections from the cues to a given outcome to be especially effective for predicting lexical decision latencies (see also Baayen et al., 2015a, for replication and extension of these findings). This measure, to which we refer here as the NDL network prior, reflects how well an outcome is entrenched in the network. Words that are well entrenched have many strong connections. Thanks to these strong connections, they acquire a higher prior availability for lexical processing. The NDL network priors are correlated with frequency of occurrence, but unlike frequency, they incorporate the history of past learning and also take into account the kind of co-learning exemplified above in Figure 7. One can also regard the NDL network prior as an abstract statistic for the behavioral profile (see Hilpert, this volume), but not of construction, but of a word.

In what follows, we explore the potential for understanding language change of the NDL network prior (as a measure of entrenchment) by means of a data set that we compiled from the COHA corpus by extracting the counts, across 20 decades in the 19th and 20th century, for the top 100 three-word phrases returned when querying the COHA website for phrases of the form

$[a \mid the]$ $[young \mid old]$ $[*]$

and phrases of the form

$[a \mid the]$ $[*]$ $[boy \mid girl \mid man \mid woman]$.

This resulted in a total of 1192 unique phrases. The upper left panel of Figure 9 plots the frequency of the words *old* and *young* in these phrases as a function of time. We see that the word *old* is more frequent than the word *young*. Furthermore, the word *young* reaches its peak use earlier, and shows a greater decline in use over time compared to the word *old*.

From the perspective of accumulation of knowledge, the binning of experience is unrealistic: Learning does not start from zero with each new decade. The second top panel therefore plots the cumulated frequencies for 1800–2000. Instead of inverse U-shaped curves, we now see monotonically increasing curves with points of inflection.

The third top panel plots the NDL network prior. The learning events for the network were based on the individual phrases. For each phrase (e.g., *the young boy*), we constructed two learning events. For the first learning event, the cues were *the* and *young*, and the outcome was *boy*. For the second learning event, the cues were *the* and *boy*, and the outcome was *young*. In other words, the network was trained to predict one of the two content words from the other two words in the phrase. We made use of the NDL2 package for `R` (Shaoul et al., 2014) to train the network on the 279,192 phrase tokens of the 1192 phrase types it our data. The network was trained decade by decade. As we did not have information on the order in which phrases appeared within a given decade, the phrases within a given decade were presented for learning in a random order. In other words, the network was updated, phrase by phrase, respecting (to the extent possible) the order in which these phrases occurred over the last two centuries. For each of the successive 20 decades, we saved the current state of weights in the network. The upper right panel of Figure 9 presents the development of the NDL network priors for the words *young* and *old* across the last two centuries. After about 120 years, both curves reach their maximum, following which they show a small decrease.

Unfortunately, the curves shown in the upper center and right panels of Figure 9 are still highly unsatisfactory, as speakers don't live to be 200 years old. The network is accumulating too much experience, and the asymptotes we see reflect more the limit of what can be learned (see, e.g., Danks, 2003) than what individual speakers could have learned over their lifetimes.

We therefore calculated the accumulation of experience for lifespans of 60 years, with the decades 1810, 1820, ..., 1950 as starting years. The second row of panels of Figure 9 plots the results for the word *young*, and the bottom row of panels shows the results for *old*. The first panels in these rows evaluate the accumulation of experience by cumulating frequency counts, whereas the second panels clarify the development of the NDL network priors. Unsurprisingly, the entrenchment of the words *young* and *old* increases over the lifetime, irrespective of whether entrenchment is estimated by cumulative frequency or by NDL network prior. Equally unsurprising is that more experience accumulates for *old* than for *young*. What is more interesting, however, is the fact that the slope of the lifetime curves decreases with increasing starting decade. The third panels for *young* and *old* plot the slopes obtained by fitting individual linear models to each indivdual lifetime curve. The solid lines represent the slopes for the learning curves based on NDL network priors, whereas the dotted lines represent the frequency-based learning curves. After an initial increase in slopes, indicating more rapid entrenchment, all four curves show a decrease.

Let us now consider the differences between the lifetime curves based on frequency and those derived from discrimination learning. First note that the lifetime curves in the center panels reveal

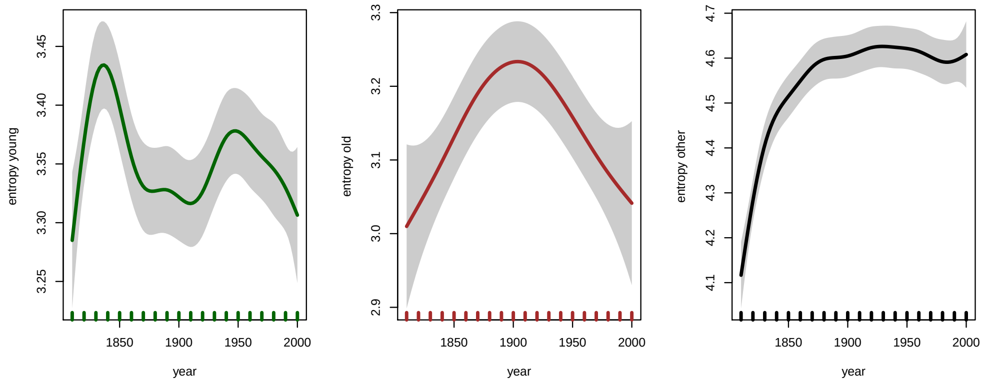

Figure 10: Development of paradigmatic entropy for the word following *young* (left), the word *old* (center) and the word preceding *boy, girl, man* or *woman* (right) in COHA.

more wiggliness. This is because, unlike frequency, network priors are sensitive to differences in burstiness and dispersion, as well as to co-learning with the other words in the phrases in our data set. Second, the learning rates (i.e., the slopes of the lifetime curves) decrease much faster when the NDL network priors are used, compared to when cumulated frequency counts are used. In other words, the discriminative learning model brings to the fore a learning problem that is less well visible in a model using frequency of occurrence as a measure of entrenchment. Third, the decline in learning rate is greater for the more frequent of the two words, i.e., for *old*.

These observations raise the question of what gives rise to this declining entrenchment. To address this question, consider Figure 10, which graphs lexical entropy as a function of decade. The left panel presents the changes in entropy for the word following the word *young* in the phrases in our sample. Entropy increased during the early years of the the industrial revolution in the United States. Entropy peaks a second time after the end of the second world war, strikingly, this is a time strongly associated with the onset of what has become known as the baby boom. For the word *old*, the entropy of the following word reaches its maximum around the turn of the century. Although interesting by themselves, these changes do not help explain the stronger declines in entrenchment predicted by the learning model, because the model learns to predict the adjectives from the third word and the article in the phrases. We therefore consider the entropy of the words preceding these third words (in the present exploratory case study, these third words are restricted to *boy, girl, man*, and *woman*). The third panel of Figure 10 shows that during the nineteenth century, this entropy was undergoing a marked increase. This increase was severely attenuated during the early decades of the twentieth century, and peaked in the years of the great economic crisis.

The increase in entropy reflects that a wider variety of words came into use (plots graphing counts of words look similar as those in Figure 10, but fail to take differences in token frequency into account). As a consequence, it became increasingly difficult to correctly predict the use of the word *old* (or *young*) in phrases of the form [*a* | *the*] [∗] [*boy* | *girl* | *man* | *woman*]: Every time that a word other than *old* (or *young*) was used, the weights on the connections from *boy* or *girl* or *man* or *woman* to *old* (or *young*) are downgraded, resulting in decreased entrenchment. This is again the Ecclesiastes Principle at work: The increase in prenominal lexical diversity comes at the cost of decreasing entrenchment of high-frequency words such as *young* and *old*. We think that the rise and fall in the frequency of use of these words, as documented in the upper left panel of Figure 9 is due to exactly this trade-off between the increasing demands on discriminating between — in the present example — events and other human agents in an increasingly more complex world, and the dynamics of discrimation learning.

This case study illustrates the possibilities offered by discrimination learning for research on language change. The present implementation of learning events is very simple, and based on selected examples of language use. A more in-depth analysis, which is beyond the scope of the present paper, would profit from richer learning events (i.e., learning events with more words and their grammatical functions, such as number, tense, aspect, thematic role, etc.) and training on the full COHA.

5 Final remarks

In this study, we have sketched a perspective on language change in which changes in language use over the course of healthy aging share critical dynamics with changes in language as used in societies characterized by a steady growth in the accumulation of knowledge. Our study focused on the consequences of the increase of knowledge for the domain of lexis. As onomasiological demands on a language increase, languages have to find ways to meet these demands. Creating new words is one option, as exemplified by the rise of compounding in Chinese and Vietnamese, which originally were isolating languages. Systematic changes in fine phonetic detail provide a complementary means for discrimination.

The accumulation of knowledge comes with undeniable costs. Older subjects are slower respondents in chronometric tasks (Ramscar et al., 2014), but this is the price for knowing more: Older subjects stunningly outperform young subjects on accuracy. Older subjects likewise reveal greater sensitivity to collocational patterns in the language. That they perform less well on matching word pairs that make no sense underlines the Ecclesiastes Principle, since once one learns that two words do not belong together, it follows that one must overcome this prior learning before one can learn to pair them. Accordingly, although it might seem that being less effective at learning nonsense is a cognitive deficit, further reflection indicates that learning to discard irrelevant associations is an evolutionary advantage (see Trimmer et al., 2012, for Rescorla-Wagner learning in evolutionary contexts).

There is one important dimension on which the parallel between society and the individual breaks down. The knowledge accumulated in our present-day society far surpasses what any individual can ever know. The renaissance ideal of the homo universalis is farther away than ever. The corpora that are now becoming available are far larger than the experience any single user can gather over a lifetime. As a consequence, the knowledge we harvest from corpora reveals more about us as a social species than about the individual.

Especially in the domain of lexis, we are faced with the problem that although the highest frequency words are common knowledge, as we move out into the low-frequency tail of Zipfian word frequency distributions, knowledge fractionates across individuals. Both classical factorial (Carroll and White, 1973) and recent crowdsourcing studies (Keuleers et al., 2015) highlight the specialized, and hence restricted, knowledge of individual language users. But perhaps knowledge specialization is the evolutionary answer to the limits on what an individual member of a commmunity can achieve. From this perspective, the registers, genres, and specialist vocabularies appear as just another variation of nature on intra-species variagation and eusociality.

For historical corpora such as COHA, we have illustrated one way in which the fractionation problem can be addressed, at least in part, by zooming in on fictive individual speakers, representing (equally fictive) generations. By restricting generational lifespans to 60 years, with different decades at which generational learning is initiated, we were able to show how increasing paradigmatic lexical diversity comes, by the Ecclesiastes Principle, at the cost of a reduction in the entrenchment of high-frequency words.

More in general, we think it is worth reflecting on parallels between language change within the lifespan of an individual, and language change in the course of the histories of a given society. It is not the case that by the age of twenty-one, a language has been learned, to remain stationary and unaltered over the remaining lifetime. Over the lifetime, new words and expressions are constantly encountered as speakers read, watch TV, travel to new places with unfamiliar names for streets and buildings, meet new people, and buy novel products. This accumulation of experience is unlikely to be uniform across the lexicon and the constructicon, and we anticipate that trade-offs at individual and aggregate levels, such as the adaptation towards pronouns under onomasiological overload, or the increase in the use of compounds (Scherer, 2005), are more widespread than we can currently imagine.

References

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, 15:147–149.

Arcodia, G. F. (2007). Chinese: a language of compound words? In *Selected proceedings of the 5th Décembrettes: Morphology in Toulouse*, pages 79–90.

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

Aylett, M. and Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, 119:3048–3058.

Baayen, R. H. (2011a). Corpus linguistics and naive discriminative learning. *Brazilian Journal of Applied Linguistics*, 11:295–328.

Baayen, R. H. (2011b). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, 5:436–461.

Baayen, R. H. (2014). Multivariate Statistics. In Podesva, R. J. and Sharma, D., editors, *Research Methods in Linguistics*. Cambridge University Press, Cambridge.

Baayen, R. H., Milin, P., Filipović Durdević, D., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118:438–482.

Baayen, R. H., Milin, P., and Ramscar, M. (2015a). Frequency in lexical processing. *Aphasiology*, to appear.

Baayen, R. H. and Ramscar, M. (2015). Abstraction, storage and naive discriminative learning. In Dabrowska, E. and Divjak, D., editors, *Handbook of Cognitive Linguistics*, pages 99–120. De Gruyter Mouton, Berlin.

Baayen, R. H., Shaoul, C., Willits, J., and Ramscar, M. (2015b). Comprehension without segmentation: A proof of concept with naive discriminative learning. *Language, Cognition, and Neuroscience*.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3):445–459.

Bannard, C. and Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, 19:241–248.

Bien, H., Baayen, R. H., and Levelt, W. (2011). Frequency effects in the production of dutch deverbal adjectives and inflected verbs. *Language and Cognitive Processes*, 27:683–715.

Carroll, J. B. and White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quaterly Journal of Experimental Psychology*, 25:85–95.

Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology*, 47(2):109–121.

Davies, M. (2010a). The corpus of contemporary american english as the first reliable monitor corpus of english. *Literary and linguistic computing*, 25(4):447–464.

Davies, M. (2010b). The corpus of historical american english (coha): 400+ million words, 1810-2009. *Online at http://www. americancorpus. org.*

des Rosiers G. and Ivison, D. (1986). Paired-associate learning: normative data for differences between high and low associate word pairs. *Journal of Clinical Experimental Neuropsychology*, 8:637–642.

Dufouil, C., Pereira, E., Chêne, G., Glymour, M. M., Alpérovitch, A., Saubusse, E., Risse-Fleury, M., Heuls, B., Salord, J.-C., Brieu, M.-A., et al. (2014). Older age at retirement is associated with decreased risk of dementia. *European journal of epidemiology*, 29(5):353–361.

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, 27(1):1–24.

Ferrer-i Cancho, R., Debowski, L., and Moscoso del Prado Martín, F. (2013). Constant conditional entropy and related hypotheses. *Journal of Statistical Mechanics: Theory and Experiment*, 07:L07001.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84(3):474–496.

Gahl, S., Cibelli, E., Hall, K., and Sprouse, R. (2014). The "up" corpus: A corpus of speech samples across adulthood. *Corpus Linguistics and Lingustic Theory*, 10(2):315–328.

Gahl, S., Yao, Y., and Johnson, K. (2012). Why reduce? phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66(4):789–806.

Galbi, D. A. (2002). Long-term trends in personal given name frequencies in england and wales. *A Journal of Onomastic*, 51(4):105–32.

Gick, B. (2002). The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association*, 32:113–121.

Gries, S. and Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on alternations. *International Journal of Corpus Linguistics*, 9(1):97–129.

Gries, S. T. and Divjak., D. S. (2009). Behavioral profiles: A corpus-based approach towards cognitive semantic analysis. In Evans, V. and Pourcel, S. S., editors, *New directions in cognitive linguistics*, pages 57–75. John Benjamins, Amsterdam & Philadelphia.

Hendriks, P., Englert, C., Wubs, E., and Hoeks, J. (2008). Age differences in adults' use of referring expressions. *Journal of Logic, Language and Information*, 17(4):443–466.

Hendrix, P. (2015). *Experimental explorations of a discrimination learning approach to language processing*. PhD thesis, University of Tübingen.

Hooper, C. R. and Cralidis, A. (2009). Normal changes in the speech of older adults: You've still got what it takes; it just takes a little longer! *SIG 15 Perspectives on Gerontology*, 14(2):47–56.

Jones, B. F. (2005). The burden of knowledge and the 'death of renaissance man': Is innovation getting harder? *Working Paper 11360*.

Kemps, R., Ernestus, M., Schreuder, R., and Baayen, R. H. (2005a). Prosodic cues for morphological complexity: The case of Dutch noun plurals. *Memory and Cognition*, 33:430–446.

Kemps, R., Wurm, L. H., Ernestus, M., Schreuder, R., and Baayen, R. H. (2005b). Prosodic cues for morphological complexity in Dutch and English. *Language and Cognitive Processes*, 20:43–73.

Keuleers, E., Stevens, M., Mandera, P., and Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology*, pages 1–62.

Kuperman, V., Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2006). Morphological predictability and acoustic salience of interfixes in Dutch compounds. *JASA*, 122:2018–2024.

Levelt, W., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22:1–38.

Lüdeling, A. and Evert, S. (2005). The emergence of non-medical -itis. Corpus evidence and qualitative analysis. In Kepser, S. and Reis, M., editors, *Linguistic evidence. Empirical, theoretical, and computational perspectives*, pages 315–333. Mouton de Gruyter, Berlin.

Marsolek, C. J. (2008). What antipriming reveals about priming. *Trends in Cognitive Science*, 12(5):176–181.

Meibauer, J., Guttropf, A., and Scherer, C. (2004). Dynamic aspects of German -er-nominals: a probe into the interrelation of language change and language acquisition. *Linguistics*, 42:155–193.

Meylan, S. C. and Gahl, S. (2014). The divergent lexicon: Lexical overlap decreases with age in a large corpus of conversational speech. In Bello, P., Guarini, M., McShane, M., and Scassellati, B., editors, *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, pages 1006–1011. Austin, TX: Cognitive Science Society.

Milin, P., Ramscar, M., Baayen, R. H., and Feldman, L. B. (2015). Cornering segmentation: the perspective from discrimination learning. *Manuscript submitted for publication, University of Tübingen.*

Palmore, E. B. and Manton, K. (1973). Ageism compared to racism and sexism. *Journal of Gerontology*, 28:363–369.

Plag, I., Homann, J., and Kunter, G. (2014). Homophony and morphology: The acoustics of word-final s in english. *Manuscript, University of Düsseldorf.*

Ramscar, M., Dye, M., and McCauley, S. M. (2013a). Error and expectation in language learning: The curious absence of mouses in adult speech. *Language*, 89(4):760–793.

Ramscar, M., Hendrix, P., Love, B., and Baayen, R. H. (2013b). Learning is not decline: The mental lexicon as a window into cognition across the lifespan. *The Mental Lexicon*, 8:450–481.

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., and Baayen, R. H. (2014). Nonlinear dynamics of lifelong learning: the myth of cognitive decline. *Topics in Cognitive Science*, 6:5–42.

Ramscar, M., Smith, A., Dye, M., Futrell, R., Hendrix, P., Baayen, R. H., and Starr, R. (2013c). The 'universal' structure of name grammars and the impact of social engineering on the evolution of natural information systems. In *Proceedings of the 35th Meeting of the Cognitive Science Society.* Berlin, Germany.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., and Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, 34(6):909–957.

Rescorla, R. A. (1988). Pavlovian conditioning. It's not what you think it is. *American Psychologist*, 43(3):151–160.

Rescorla, R. A. and Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Black, A. H. and Prokasy, W. F., editors, *Classical conditioning II: Current research and theory*, pages 64–99. Appleton Century Crofts, New York.

Scherer, C. (2005). *Wortbildungswandel und Produktivität: Eine empirische Studie zur nominalen'-er'-Derivation im Deutschen*, volume 497. Walter de Gruyter.

Schmid, H.-J. and Küchenhoff, H. (2013). Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics*, 24(3):531–577.

Schönle, P. W., Gräbe, K., Wenig, P., Höhne, J., Schrader, J., and Conrad, B. (1987). Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain and Language*, 31:26–35.

Shaoul, C., Schilling, N., Bitschnau, S., Arppe, A., Hendrix, P., and Baayen, R. H. (2014). *NDL2: Naive Discriminative Learning.* R package version 1.901, development version available upon request.

Shaoul, C., Willits, J., Ramscar, M., Milin, P., and Baayen, R. H. (2015). A discrimination-driven model for the acquisition of lexical knowledge in auditory comprehension. *under revision.*

Singh-Manoux, A., Kivimaki, M., Glymour, M. M., Elbaz, A., Berr, C., Ebmeier, K. P., Ferrie, J. E., Dugravot, A., et al. (2012). Timing of onset of cognitive decline: Results from whitehall II prospective cohort study. *British Medical Journal*, page 344.

Stefanowitsch, A. and Gries, S. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2):209–243.

Stemberger, J. P. and MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory and Cognition*, 14:17–26.

Tabak, W., Schreuder, R., and Baayen, R. H. (2010). Producing inflected verbs: A picture naming study. *The Mental Lexicon*, 5(1):22–46.

Tomaschek, F., Tucker, B. V., Wieling, M., and Baayen, R. H. (2014). Vowel articulation affected by word frequency. In *Proceedings of 10th ISSP, Cologne*, pages 429–432.

Trimmer, P. C., McNamara, J. M., Houston, A. I., and Marshall, J. A. R. (2012). Does natural selection favour the Rescorla-Wagner rule? *Journal of Theoretical Biology*, 302:39–52.

Walling, A. D. and Dickson, G. M. (2012). Hearing loss in older adults. *American family physician*, 85(12):1150–1156.

Wood, S. N. (2006). *Generalized Additive Models*. Chapman & Hall/CRC, New York.

Zipf, G. K. (1949). *Human Behavior and the Principle of the Least Effort. An Introduction to Human Ecology*. Hafner, New York.