into consideration, such that the phonetic contrasts at the beginning of words in each syntactic category show a better fit to a geometric than a power law distribution (Figure 19).

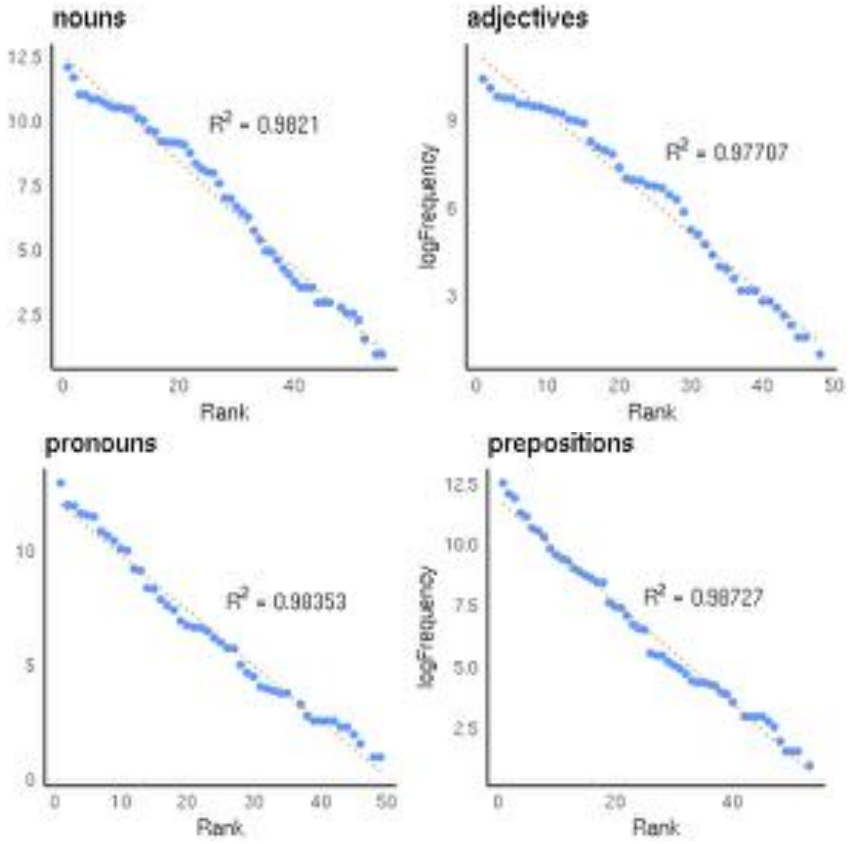

Figure 19: Log frequency x frequency rank plots of the distributions of word initial phonetic labels assigned to observed forms across parts of speech in the Buckeye Corpus of conversational speech (Pitt, Johnson, Hume et al, 2005)

Not only do these findings emphasize the idea that lexical "items" are themselves parts of systems, they also hopefully serve to illustrate how systematicity is present in human communicative codes at many (if not all) levels of abstraction.

8. Making sense of meaning

> *"The belief that words have a meaning of their own account is a relic of primitive word magic, and it is still a part of the air we breathe in nearly every discussion."*
> *Ogden & Richards (1923)*

8.1 Moving beyond compositionality

As I noted at the outset of this article, belief in compositionality – the doctrine that meaningful messages are built up out of smaller units of meaning (etc.) – has dominated and continues to dominate theories of in human communication. Indeed, at times is seems that most linguists (and other researchers in the brain and cognitive sciences) are unable to conceive of human communication in anything other than in compositional terms. As a result, it seems to fair to say that on one hand, even sensible attempts to dispense with compositionality (despite its myriad problems) are generally greeted with incredulity by the research community, whereas on the other hand, even the most outlandish suggestions tend to be given credence, just as long as they conform to the prevailing orthodoxy. For example, the problems posed by compositionality have led leading theorists to the absurdist claim that units of meaning—such as *dog, appear, carburetor, bureaucrat*, and, presumably *entropy, perplexity John, William, Mary* and *Anne*—must be (somehow) specified in the genome:

> *there is good reason to suppose that the [nativist] argument is at least in substantial measure correct even for such words as carburetor and bureaucrat, which, in fact, pose the familiar problem of poverty of stimulus if we attend carefully to the enormous gap between what we know and the evidence on the basis of which we know it. The same is often true of technical terms of science and mathematics, and it surely appears to be the case for the terms of ordinary discourse. However surprising the conclusion may be that nature has provided us with an innate stock of concepts, and that the child's task is to discover their labels, the empirical facts appear to leave open few other possibilities. (Chomsky, 2000, pp. 65–66)*

While many researchers who subscribe to compositionality would balk at endorsing this claim, it hardly matters, first because the idea of *innate concepts* is so completely vague as to contribute virtually nothing to our understanding of human communication, and second because it is clear that no better alternative account of meaning compositionality is on offer.

As I noted at the outset, after half a century of motivated effort, researchers have singularly failed to come up anything approximating a half-coherent empirical account of what a 'unit of meaning' is supposed to be (Ramscar & Port, 2015), and philosophical analyses that have long

suggested that 'meaning units' are a fundamentally misguided idea (Wittgenstein, 1953; Quine, 1960; see also Fodor, 1998). Which is to say that although most linguists (and other researchers in the brain and cognitive sciences) clearly believe in compositionality, theoretical accounts of compositional meaning themselves offer nothing beyond blind faith, vagueness, and / or mysticism.[14]

By contrast, when taken together with well-established principles from information theory, the data described here show that for personal names, the empirical facts emphatically open up possibilities for explaining human communication that seem far more helpful than hopeful appeals to innate concepts. Empirically it seems clear that name grammars are perfectly adapted to supporting a discriminative process in which speaker's signals – structured sequences of name tokens – simply serve to eliminate semantic uncertainty in a hearer. These combinations / sequences of name tokens are not compositional, nor do they 'refer to concepts' (whatever concepts are supposed to be). Rather, tokens (and token sequences) eliminatively narrow the space of possibilities within the set of identities that is correlated with them.

At which point, it seems worth re-emphasizing that accounting for the workings of personal names has traditionally proven to be a stumbling block for compositional theories, to the extent that mainstream linguistic theories typically either ignore names altogether, or else relegate them to an afterthought. By contrast, names appear to be an aspect of human communication that is readily explained in discriminative terms.[15] The reason is likely to be the same in both cases. On one hand the *sui generis* semantics that are inevitably associated with identities are difficult to reconcile with the fact that names themselves are rarely unique, such that name tokens don't appear to correspond to generic 'concepts.' On the other, the self same *sui generis* semantics make a discriminative account intuitive because the very fact that, in context, names more often than not succeed in picking out a unique correlated identity means that names are usually

[14] A previous version of this work was criticized for making "little connection with relevant current work in theoretical pragmatics" and "not much connection with the state of the art in theoretical linguistics, pragmatics, psycholinguistic processing, or children's semantic/pragmatic development," while a reviewer complained that "the specific cases … discussed are all to do with words (names, verbs, nouns, gender systems), while syntax, the key driver of linguistic compositionality, is not mentioned." What I hope is clear to the careful reader (and even future reviewers) is that I hold out no hope that a successful theory of human communication can be built on the idea of 'units of meaning' at any possible level of description, and that as a result, I have little to say about work founded on this idea (the 'state of the art in theoretical linguistics,' 'pragmatics,' 'syntax') other than to note that if the foundations of a scientific theory are wrong, it seems reasonable to assume that its ultimate contributions to human understanding are likely to be minimal.

[15] I acknowledge that many readers will find this account of the communicative function of names unsatisfactory, because will feel that this account fails to satisfy their intuitions about lexical semantics and meaning. I can only reiterate here that what one knows / feels etc., about an identity is entirely independent of how human communicative codes are used to signal them, and note that a large part of the development of a mature science of human communication will involve a re-appraisal of the kinds of phenomena that we can plausibly seek to explain (i.e., it seems likely linguists are no more likely to develop a 'comprehensive theory of meaning' than physicists or horologists are to develop a 'theory of time travel').

unambiguous in context. This means that in many contexts the successful communication of a name will result in the elimination of uncertainty on the part of a hearer (13), as opposed to merely its reduction (which appears to be sufficient for a great many communicative purposes).

(13) John: *Who wrote that book you told me about?*

Ann: *Richard Feynman*

However, as (14) illustrates, contexts where the successful communication of a name does not result in the elimination of uncertainty on the part of a listener clearly exist.

(14) John: *Do you like David Bowie?*

Ann: *It depends*

John: *Do you like David Bowie in his Ziggy period?*

Ann: *Yes*

Of course, names are only one, fairly small aspect of human communication. Providing an account of all the other myriad ways in which people communicate using words and sequences of words is beyond the scope of this paper (and its author's current abilities). There are however, many good reasons to believe that accounting for the rest of human communication will be best achieved by an extension of the approach I have outlined here. One obvious reason to support this suggestion is the fact that, as noted above, research has shown compositional accounts of communication to have little going for them other than people appear to be predisposed to intuitively believe them. For most of the past century adherence to these beliefs has led to a situation where the failure to find evidence for compositional theories has *not* been taken as evidence against them. Rather, the absence of evidence has been taken to support claims that people are miraculously able to learn and use all of the various inexplicable aspects of compositional communicative codes in spite of the 'poverty of the stimulus' from which they are supposed to learn them. (Hence the appeal to innate – though ill-defined – stocks of concepts in the quote above.)

Red Orange Yellow Green Blue Purple Brown Magenta Tan Cyan Olive Maroon Navy
Aquamarine Turquoise Silver Lime Teal Indigo Violet Pink Black White Grey/Gray

Table 11: 24 common English color words (from Wikipedia)

To further this line of thought, Table 11 lays out a set of 24 common English color words (taken from wikipedia). Perhaps unsurprisingly given the foregoing, an analysis of the frequency

distribution of these words in COCA (plotted in Figure 20) shows that they are geometrically distributed. That is, it seems that when people talk about color in English, they use *white* exponentially more frequently than *black,* which they use exponentially more frequently than *red* which they use exponentially more frequently than *green* which they use exponentially more frequently than *brown* which they use exponentially more frequently than *blue;* and this pattern continues successively for *grey/gray* then *yellow* then *silver* then *orange* then *pink* then *navy* then *olive* then *purple* then *lime* then *tan* then *violet* then *turquoise* then *maroon* then *indigo* then *teal* then *magenta* then *aquamarine* then *cyan*. This pattern is not predicted by compositional theories, but neither does it contradict them. A compositional theorist might suppose that it reflects some bias, or constraint, on people's color concepts, a factor that is independent of communication, and appeal to physiologists or physicists to explain its basis.

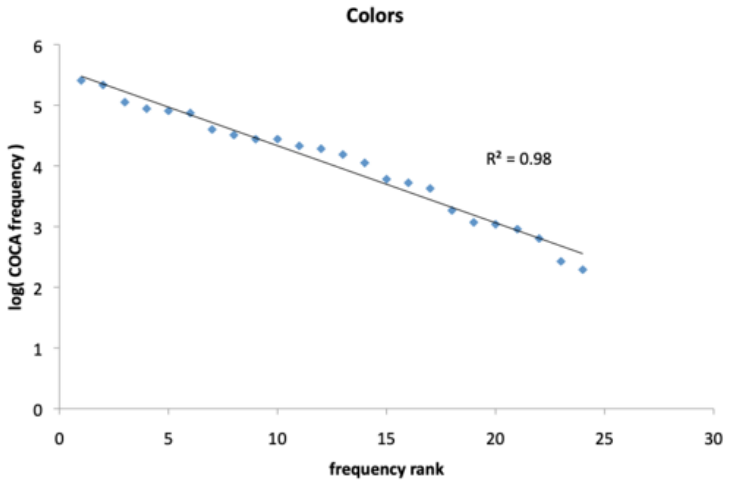

Figure 20: Log frequency x frequency rank plot of 24 color English color words (R^2=.98).

Mother, Father, Son, Daughter, Brother, Sister, Uncle, Grandmother, Aunt, Grandfather, Cousin, Grandson, Nephew, Niece, Granddaughter

Table 12: Set of English kinship terms defined by Kemp & Regier (2012).

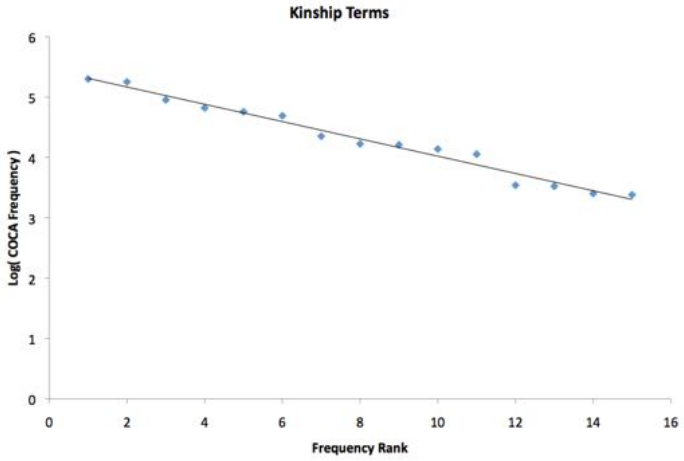

Figure 21: Log frequency x frequency rank plot of the English kinship terms defined by Kemp & Regier (2012; R^2=.98).

Table 12 then lays out a set of 15 kinship terms taken from Kemp & Regier (2012). An analysis of the frequency distribution of these words in COCA (plotted in Figure 21) shows that they too are geometrically distributed. This pattern is not predicted by compositional theories either, but once again, neither does it contradict them. Our compositional theorist might suppose that it too reflects some bias, or constraint, this time on people's kinship concepts, a factor that is also independent of communication, and appeal to sociologists or anthropologists to explain its basis (again, contrary to the claim noted above, Chomsky, 2000, it seems that the empirical facts always appear to leave open many possibilities).

By contrast, a key piece of evidence I offered in support of the idea that names support a discriminative communicative function is the way they are distributed (and from the apparent universality of these distributions). Section 5 further showed that even if one were to object to my seeking to abstract my information theoretic account of communication from names to the rest of human communication by arguing that names are somehow 'special,' it turns out that many other 'more usual' communicative concepts (such as those associated with nouns and verbs) have the same distributional structure. Moreover, as Figure 22 appears to indicate, many of the 'biases and constraints' that shape people's use of words across different domains appear to result suspiciously similar patterns. These patterns make perfect sense from a discriminative, information theoretic, perspective, but once again, while they are not at odds with compositional accounts, neither are they predicted by the various intuitions that appear to drive people's implacable faith in them.

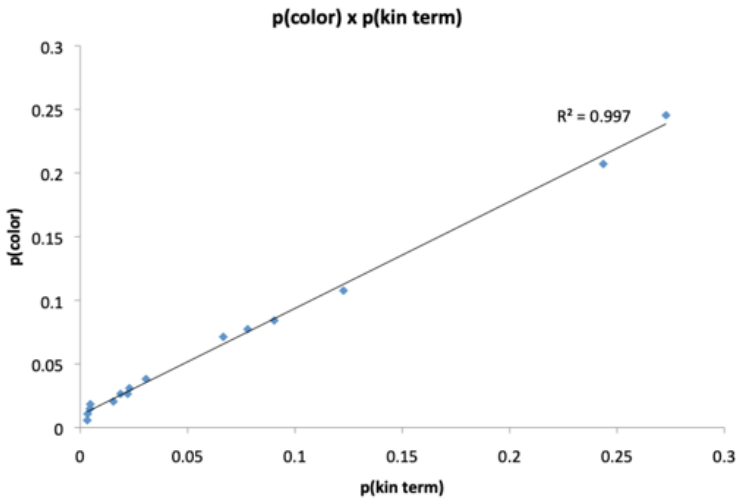

Figure 22: Point-wise comparison of the rank probabilities of English color words (calculated over a set of 24 items) and English kin terms (calculated over a set of 15 items; R^2=.997).

Thus while it is true that with sufficient time and effort, one could come up with plausible-sounding post hoc stories about biases, constraints etc., that account for all of the skewed variance in the frequencies of these supposedly compositional items (variance that compositional intuitions utterly fail to predict), at some point it seems that a duck test[16] might be in order. If natural languages are distributed the way that deductive, non-compositional communicative codes are distributed, if the lexical contrasts within domains maximize discrimination in the way that the codewords in a communicative code maximize discrimination, and if the information in human codes increases according to communicative demands in the way one would expect the information in a communicative code to increase as its coverage increases, this is probably because natural languages are in fact deductive, non-compositional communication systems.

8.2 Information and meaning

Since a great deal of the last part of this paper has been devoted to explaining to readers why compositional accounts of human communication – or, as these theories usually describe it, 'language' – are a theoretical dead end, I feel that it behooves me to end in a more upbeat manner by providing furthers example of how information theoretical accounts of communication can positively increase our understanding of meaningful questions. To this end, in closing I will briefly consider how the approach I have described here can be used to illuminate what is often called the "Easterlin paradox" in the relationship between wealth and human happiness. In doing

[16] If it looks like a duck, swims like a duck, and quacks like a duck, it probably is a duck.

so, I will try to highlight a few outstanding questions raised by this approach, as well as the benefits that information theory brings to this specific question.

The "Easterlin paradox" describes a curious finding in the relationship between wealth and human happiness. Put simply, it has been argued that although at any given point in time a relationship between wealth and happiness can be seen (such that, say, if richer and poorer countries / citizens are compared at a point in time, life satisfaction increases with the absolute amount of GDP per capita), across time there seems to be no significant relationship between the rate of improvement in happiness and the rate of economic growth (see e.g., Easterlin & Angelescu, 2009; Easterlin, 2013; Hills, Proto & Sgro, 2015). In this work, happiness is measured by overall life satisfaction, which is operationalized in terms of people's responses to the following question (Easterlin & Angelescu, 2009):

(15) *"All things considered, how satisfied are you with your life as a whole these days*"

Accordingly, it follows that the happiness / income paradox is more accurately put as follows: while at any point in time the degree to which people report being satisfied increases as income increases, it turns out that over time the degree to which people report being satisfied does not increase as incomes increase.

How can an information theoretical account of communication increase our understanding of this matter? To begin with, it is important to note this paradox arises under (relies on) a critical assumption, namely that when people are asked, *how satisfied are you with your life*, the meaning of *satisfied* is constant over time. This, of course, is a standard compositional assumption: *satisfied* is assumed to be associated with a 'concept,' and a further (implicit) assumption is that concepts are generally stable over time, except when they are not; at which point compositional theories usually don't have much to say about how concepts change, because compositional theories general analyses concepts in isolation (Ramscar & Port, 2015). By contrast, as I have emphasized throughout the foregoing (and as figure 12 in particular serves to illustrate) information is a property of systems, and these systems balance the constraints imposed by the requirement to communicate specific messages (e.g., about identities), with the need to make those specific messages informative (which also embraces predictable and learnable) in a community that needs a number of different specific messages to be communicable.

This view of communication is, as I have repeatedly stressed, concerned with uncertainty reduction. It follows therefore that over time, people might be expected to use specific words in some kind of relation to the degree to which they relate to the way uncertainty is distributed in the environment. Thus, for example, in a series of analyses (Ramscar, 2015) of *old-* word sequences

such as *old man*, *old woman,* I found that the frequency of *old-* word sequences declined as the number of elderly people in the population increased across the late 20th Century. While this might seem paradoxical from a composional perspective (where on might expect that the presence of more old men might predict more talk of *old men*), it is entirely consistent with the account of names presented earlier, since it follows that the more old men there are, the less informative (discriminative) talk of *old men* will become in context.

Since human communication can in many cases be seen as the utilization of a code in order to reduce uncertainty,[17] it follows that when changes in experiential context change the informativity of an aspect of the code, people's use of that aspect of the code is likely to change. It also follows that its informativity within the code, both in context and perhaps across contexts, will change. In this case, the information that *old* contributes to a listener's expectation that *man, men, woman* or *women* will occur in spoken contexts can be shown to have changed considerably in the past 50 years. It follows also that the meaning of *old man, old men, old woman* and *old women* are likely to have changed as well. While I have no idea how one might hope to quantify what these changes in meaning are (or even if this question is well-posed, Wittgenstein, 1953), an information theoretic approach to communication does at least offer methods for evaluating whether we might expect changes to have taken place.

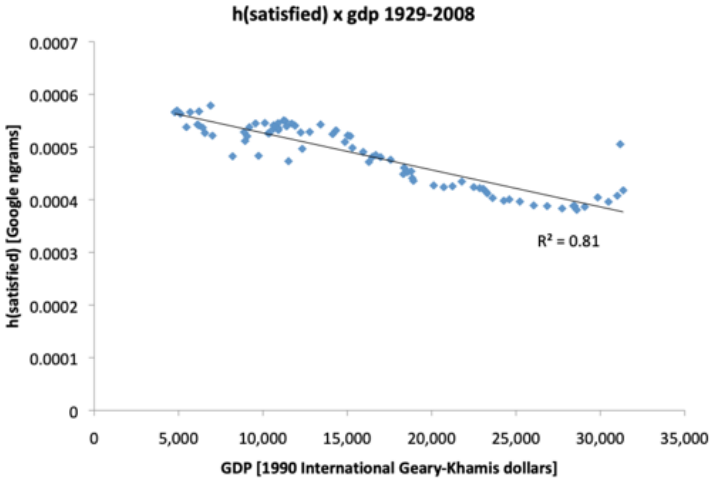

Figure 23: the entropy of the word *satisfied* in Google ngrams from 1929-2008, plotted against US per capita GDP[18] over the same period.

[17] This is not to deny that non-verbal and non-lexical codes also clearly exist, or that they contribute enormously to human communication.

[18] Measured in 1990 International Geary-Khamis dollars. Data from: www.ggdc.net/maddison/historical_statistics/

With this in mind, and returning to the happiness – wealth paradox, I analyzed the informativity of the word *satisfied* (operationalized as its entropy in Google ngrams) in relation to US per capita GDP over the period 1929-2008 (Figure 23). As can be seen, it is clear that across this period, increases in wealth are strongly associated with changes in the informativity of *satisfied* (R^2=.81).

Of course, it might be objected that since my analyses of Delaware names, *old men* etc. and *satisfied* all reveal declines in the informativity of specific words across time that this might just be a general property of the code: perhaps all words decline in informativity. To at least attempt to control for this possibility, I analyzed the informativity of the word *poverty* (operationalized as its entropy in Google ngrams) in relation to US per capita GDP over the period 1929-2008. Since I argued that the increases in general satisfaction that result from increases in GDP might lead people to talk less about being *satisfied* (because *satisfied* will become less informative when more people are satisfied), it follows that the opposite ought to be true of *poverty*. As levels of poverty decrease, then the word poverty ought to increase in its informativity (*poverty* will be uninformative when everyone is in poverty, and its informativity will increase as poverty becomes less widespread).

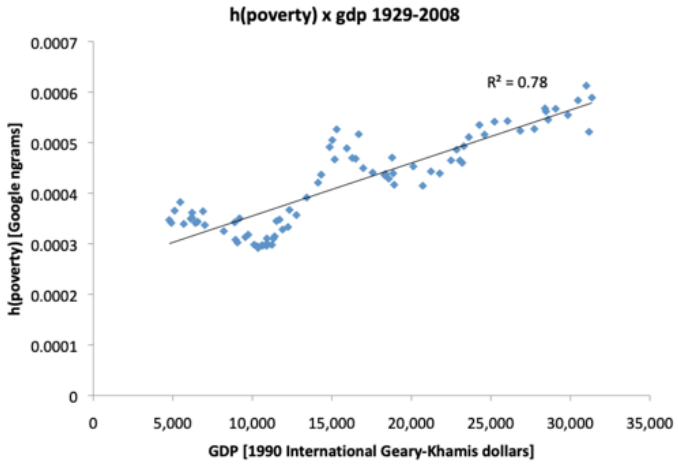

Figure 23: the entropy of the word *poverty* in Google ngrams from 1929-2008, plotted against US per capita GDP over the same period.

As figure 23 shows, whereas increases in wealth were strongly associated with a decline in the use of *satisfied* between 1929 and 2008, the opposite was true of *poverty* – as America's wealth increased, talk about *poverty* amongst American increased as well (R^2=.78). In line with my remarks above, it is worth highlighting two things that these results seem to indicate: first, as

people have become wealthier, they talk about being *satisfied* less, presumably because, if we suppose *satisfied* means 'satisfied,' the fact that more and more people are satisfied makes talk about being satisfied less meaningful; and second, because people talk about being *satisfied* less as they become wealthier, it follows that talk about being *satisfied* must increasingly occur in fewer, more specialized contexts, such that if we accept that the use of *satisfied* contributed something to the meaning of 'satisfied' in 1929, then the meaning of 'satisfied' must have changed by 2008. [19] (And of course, when it comes to *poverty*, the opposite pattern seems to hold.)

Interestingly, an analysis of the informativity of the word *happy* (operationalized as its entropy in Google ngrams) in relation to US per capita GDP over the period 1929-2008 appears to confirm the wealth-happiness paradox, since it revealed the same U-shaped relationship between wealth and poverty reported by Easterlin and colleagues (Figure 24).

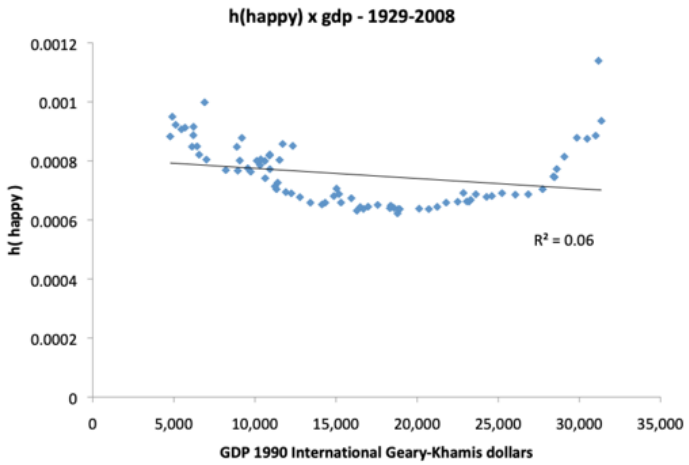

Figure 24: the entropy of the word *happy* in Google ngrams from 1929-2008, plotted against US per capita GDP over the same period.

This raises some interesting questions, an obvious one being whether the tendency in the literature to equate people's *being satisfied* with *happiness* in comparisons of wealth and attitude is justified, given that, in terms of informativity at least, *satisfied* and *happy* interact very differently with wealth over time. Among the other obvious questions it raises are whether these correlations actually mean anything, and if they do, what, exactly are they supposed to be telling

[19] This paragraph clearly highlights the perils involved in using words to talk about word meanings.

us? I will not attempt to answer these questions here, not least because I suspect that each of them represents a serious research topic in its own right. But I will try to briefly sketch out why I feel they are useful questions, and why I think an information theoretic approach can shed useful light on them.

First, what I hope is clear is that while simply assuming that words like *satisfied, poverty* and *happy* have fixed associated concepts will serve to affirm the paradox and satisfy people's compositional intuitions, it will fail to explain the systematic changes observed in the use of the words *satisfied* and *poverty* over time, which, if one cares about relations between wealth and well-being, seem worth exploring. Second, while it is true that the comparison between the informativity of *happy* and gdp observed in figure 24 supports the wealth – happiness paradox, the fact that this analysis revealed the same U-shaped relationship between wealth and poverty reported by Easterlin and colleagues supports the overall suggestion here that informativity may be a useful way of looking at meaning over time, albeit that this matter is muddied considerably by the fact that data reporting attitudes analyzed by Easterlin and colleagues were responses to satisfaction questions. This raises still more questions, not the least of which is whether the answers that people provide in response to survey questions answer the questions that researchers want to ask (Bertrand & Mullainathan, 2001).

Which is to say, finally, that if one cares about things like relations between wealth and well-being, then if people's ideas of wealth and well-being are at all influenced by the way words like *wealth and well-being* are used in the codes we use to communicate them, improving our scientific understanding of these codes can only help. In this regard, it is interesting to consider the degree to which we might expect the words *happy, satisfied* and *poverty* to be directly related to observable economic data in the first place. To take the last of these first, if we assume that *poverty* is typically used to talk about existing in a state where ones day to day basic living needs are not met (i.e., where one daily faces some kind of existential threat), then it seems reasonable to expect that the informativity of the word *poverty* might be expected to change as general living standards improve. Similarly, if we assume that people tend to be *satisfied* when their day to day basic living *are* met (i.e., where they are not faced with any kind of existential threat), then it also seems reasonable to expect that the informativity of the word *satisfied* might be expected to change as general living standards improve.

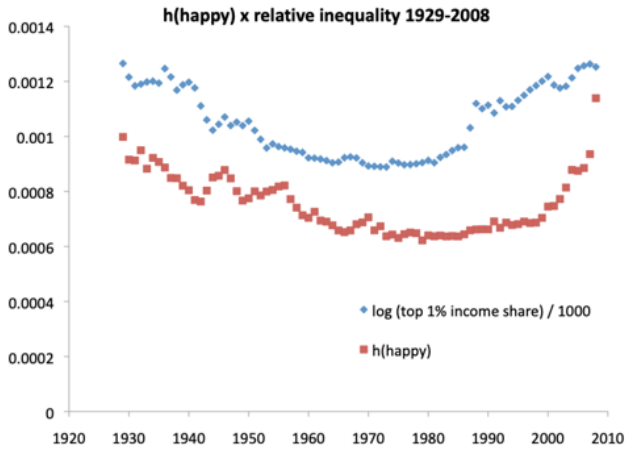

Figure 25: the entropy of the word *happy* in Google ngrams from 1929-2008, plotted against the scaled log of the income share of the top 1% of earners[20] over the same period.

This then leaves us with *happy,* which seems to have a rather more complex relationship to wealth than *poverty* or *satisfied,* not least because, for example, getting a 20% pay rise might alleviate the poverty of a low paid worker (and increase their general satisfaction), it might not necessarily make them happy, especially if they learn that all of their colleagues are getting, say, an 80% pay rise, an analysis that seems supported by relationship between the entropy of *happy* and a common measure of relative income inequality (Figure 25; R^2=.47). It is worth stressing again that I do not want to claim that these analyses are "right." That isn't the point of this exercise. Rather what I hope is obvious is that addressing these questions from a perspective that treats words as parts of information systems can cast an interesting and potentially helpful light on them. It can help illuminate some important factors that one might to control for if one does care about things like relations between wealth and well being, as many governments seem to do, and it raises – and allows one to even explore objectively – a very interesting question: if one wants to get at the answer to a complex question like the relationship between wealth and well being, is it better to ask people what they think (at a given point in time and a particular context), or to look at the way that people's communicative behavior reflects what they think across time, and across contexts.

[20] Data from https://eml.berkeley.edu/~saez/TabFig2012prel.xls

9. Conclusion

I have described an account of how human communication works based on well-established theories of learning[21] and communication, and have used the predictive power of this account to uncover and describe the way in which the cultural environment has evolved a set of remarkable structures to support human communication and the learning of human communicative skills; structures that were hitherto undiscovered. Finally, I have described how although these structures are largely incompatible with – and unpredicted by – contemporary linguistic theory, they are entirely consistent with what we know about formal theories of communication, learning and coding (Shannon, 1949; Ramscar et al, 2010). Given this last point, perhaps the most surprising aspect of these data is just how surprising most students of language and communication will find them to be.

[21] Across species learning has evolved in a way that enables individuals to respond in ways that are highly sensitive to information in the environment (Rescorla, 1988). Meanwhile, the development of cognitive control / selective attention (which allows individuals to self-direct their learning) develops very slowly in humans, such that young humans appear to be particularly pre-disposed to the learning of the conventions that communications systems appear to rely on (Ramscar & Gitcho, 2007). Taken together with the data described here, these considerations indicate that interactions between linguistic behavior, learning and its development within communities are capable over time of producing self-organizing communication systems, as well as the remarkable statistical – and ontological – structures described here.

References

Ackerman, F., & Malouf, R. (2013). Morphological organization: The low conditional entropy conjecture. Language, 89(3), 429-464.

Asr, F. T., Willits, J., & Jones, M. (2016). Comparing Predictive and Co-occurrence Based Models of Lexical Semantics Trained on Child-directed Speech. In Proceedings of the 38th Annual Conference of the Cognitive Science Society.

Aylett, M., & 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(1), 31-56.

Baayen, R. H. (2001). Word frequency distributions (Vol. 18). Springer Science & Business Media.

Baayen, R. H., McQueen, J. M., Dijkstra, T. & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In Baayen, R. H. & Schreuder, R. (eds.), Morphological Structure in Language Processing, Berlin: Mouton de Gruyter, 355–370.

Baayen, R. H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. Aphasiology, 30(11), 1174-1220.

Baek, S.K., Kiet, H.A.T, Kim, B.J. Family name distributions: Master equation approach, Physical Review E 76, 046113, (2007)

Bahl, L., Baker, J., Jelinek, E., & Mercer, R. (1977). Perplexity—a measure of the difficulty of speech recognition tasks. In Program, 94th Meeting of the Acoustical Society of America 62:S63, Suppl. no. 1

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. Journal of Memory and Language, 60(1), 92-111.

Benford, F. (1938). The law of anomalous numbers. Proceedings of the American philosophical society, 551-572.

Bertrand, M., & Mullainathan, S. (2001). Do people mean what they say? Implications for subjective survey data. American Economic Review, 91(2), 67-72

Blevins, J. P. (2006). Word-based morphology. Journal of Linguistics, 42(3), 531-573.

Blevins, J. P. (2016). Word and paradigm morphology. Oxford University Press.

Blevins, J. P., Ackerman, F., Malouf, R., & Ramscar, M. (2016). Morphology as an adaptive discriminative system. Morphological metatheory, 271-302.

Blevins, J. P., Milin, P., & Ramscar, M. (2017). The Zipfian paradigm cell filling problem. Perspectives on Morphological Organization: Data and Analyses, 10, 141.

Bloomfield, L. 1933. Language. New York: Holt.

Boersema, D. B. (2000). Wittgenstein on Names. Essays in Philosophy, 1(2), 7.

Burge, T. (1974). Reference and proper names. The Journal of Philosophy, 70(14), 425-439.

Cameron‐Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. Cognitive Science, 27(6), 843-873.

Carey, S., Shusterman, A., Haward, P., & Distefano, R. (2017). Do analog number representations underlie the meanings of young children's verbal numerals?. Cognition, 168, 243-255.

Chomsky, N. (2000). New horizons in the study of language and mind. Cambridge, England: Cambridge University Press

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. Behavioral and brain sciences, 31(5), 489-509.

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. SIAM review, 51(4), 661-703.

Corbett, G.G. 1991. Gender. Melbourne: Cambridge University Press.

Crook, A L (2012) Personal naming patterns in Scotland, 1700 - 1800: a comparative study of the parishes of Beith, Dingwall, Earlston, and Govan. MPhil(R) thesis. University of Glasgow

Culicover, P. W. (1999). Syntactic nuts: Hard cases, syntactic theory, and language acquisition (Vol. 1). Oxford University Press on Demand.

Culicover, P. W.(1999) Syntactic Nuts: Hard Cases in Syntax. Oxford University Press, Oxford.

Cutler, A., McQueen, J., & Robinson, K. (1990). Elizabeth and John: Sound patterns of men's and women's names. Journal of linguistics, 26(2), 471-482.

Davies, M. (2009). The Corpus of Contemporary American English (COCA): 400+ million words, 1990-present (2008).

Donnellan, K. S. (1972). Proper names and identifying descriptions. In Semantics of natural language (pp. 356-379). Springer, Dordrecht.

Du, R., Yuan, Y., Hwang, J., Mountain, J., & Cavalli-Sforza, L. L. (1992). Chinese surnames and the genetic differences between north and south China. Journal of Chinese Linguistics Monograph Series, (5), 1-93

Dye, M., Milin, P., Futrell, R., & Ramscar, M. (2017). A functional theory of gender paradigms. In F. Kiefer, J. P. Blevins, & H. Bartos (Eds.), Perspectives on morphological organization: Data and analyses (pp. 212–239). Leiden: Brill

Dye, M., Milin, P., Futrell, R., & Ramscar, M. (2018). Alternative solutions to a language design problem: The role of adjectives and gender marking in efficient communication. Topics in cognitive science, 10(1), 209-224.

Easterlin, R. A. (2013). Happiness, growth, and public policy. Economic Inquiry, 51(1), 1-15.

Easterlin, R. A., & Angelescu, L. (2009). Happiness and growth the world over: Time series evidence on the happiness-income paradox. Forschungsinstitut zur Zukunft der Arbeit IZA discussion paper 4060

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. Behavioral and Brain Sciences, 32(5), 429-448

Faarlund, J. T., & Emonds, J. E. (2016). English as North Germanic: A Summary. Language Dynamics and Change, 6(1), 1-17.

Fara, D. G. (2015). Names are predicates. Philosophical Review, 124(1), 59-117.

Farmer, J. D., & Geanakoplos, J. (2006). Power laws in economics and elsewhere. Santa Fe Institute Technical Report.

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. Proceedings of the National Academy of Sciences, 109(44), 17897-17902

Ferrer i Cancho, R., & Solé, R. V. (2001). Two Regimes in the Frequency of Words and the Origins of Complex Lexicons: Zipf's Law Revisited*. Journal of Quantitative Linguistics, 8(3), 165-173.

Firth, J.R. (1957). A synopsis of linguistic theory 1930-1955. In Studies in Linguistic Analysis, pp. 1-32. Oxford: Philological Society. Reprinted in F.R. Palmer (ed.), Selected Papers of J.R. Firth 1952-1959, London: Longman (1968)

Fodor, J. A. (1998). Concepts: Where cognitive science went wrong. Oxford University Press.

Fox, W. R., & Lasker, G. W. (1983). The distribution of surname frequencies. International Statistical Review/Revue Internationale de Statistique, 81-87.

Frege, G. (1892). On sense and meaning. Translations from the philosophical writings of Gottlob Frege , 3 , 56-78.

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. Proceedings of the National Academy of Sciences, 112(33), 10336-10341.

Galbi. D.A. (2002) Long-Term Trends in Personal Given Name Frequencies in England and Wales, Working paper version 1.1, Federal Communications Commission, Washington DC, USA

Gallager, R., & Van Voorhis, D. (1975). Optimal source codes for geometrically distributed integer alphabets (corresp.). IEEE Transactions on Information theory, 21(2), 228-230.

Gerlach, M., & Altmann, E. G. (2013). Stochastic model for the vocabulary growth in natural languages. Physical Review X, 3(2), 021006.

Golomb, S. (1966). Run-length encodings (Corresp.). IEEE transactions on information theory, 12(3), 399-401.

Gray, A. (2014). Name-bearing, reference, and circularity. Philosophical Studies, 171(2), 207-231.

Grünwald, P. D., & Vitányi, P. M. (2003). Kolmogorov complexity and information theory. With an interpretation in terms of questions and answers. Journal of Logic, Language and Information, 12(4), 497-529.

Guo , Jin-Zhong, Chen, Qing-Hua, and Wang, You-Gui. "Statistical distribution of Chinese names." Chinese Physics B 20.11 (2011): 118901

Gureckis, T. M., & Goldstone, R. L. (2009). How you named your child: Understanding the relationship between individual decision making and collective outcomes. Topics in Cognitive Science, 1(4), 651-674.

Hall, K. C., Hume, E., Jaeger, T. F., & Wedel, A. (2018). The role of predictability in shaping phonological patterns. Linguistics Vanguard.

Handel, Z. (2008). What is Sino‐Tibetan? Snapshot of a Field and a Language Family in Flux. Language and Linguistics Compass, 2(3), 422-441.

Hanks, P., Coates, R., & McClure, P. (2011). Methods for studying the origins and history of family names in Britain: Philology meets statistics in a multicultural context. In Facts and Findings on Personal Names: Some European Examples. Proceedings of an International Symposium in Uppsala (pp. 37-58). Uppsala: Acta Acadamiæ Regiæ Scientiarum Upsaliensis.

Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., ... & Ng, A. Y. (2014). Deep speech: Scaling up end-to-end speech recognition. arXiv preprint arXiv:1412.5567.

Hartley, Ralph. “Transmission of Information,” Bell System Technical Journal 7, no. 3 (July 1928): 535–63.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. Psychonomic bulletin & review, 7(2), 185-207.

Hills, TT, Proto, E., and Sgroi, D. (2015) Historical Analysis of National Subjective Wellbeing Using Millions of Digitized Books,” Forschungsinstitut zur Zukunft der Arbeit IZA discussion paper 9195

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. The Journal of the Acoustical Society of America, 29(2), 296-305.

Huffman, D. A. (1952). A method for the construction of minimum-redundancy codes. Proceedings of the IRE, 40(9), 1098-1101.

Hvenekilde, A., Marak, C. R., & Burling, R. (2000). Personal Names in a Mande (Garo) Village. Names, 48(2), 83-104.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. Cognitive Psychology, 61(1), 23–62.

Jeffries, J., & Fulton, R. (2005). Making a population estimate in England and Wales. National Statistics Methodological Series, 34.

Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. Psychonomic Bulletin & Review, 17(5), 662-672.

Jones, M. N., & Mewhort, D. J. (2007). Representing word meaning and order information in a composite holographic lexicon. Psychological review, 114(1), 1.

Jones, M. N., Willits, J. & Dennis, S. (2015). Models of semantic memory. Oxford handbook of mathematical and computational psychology, 232-254

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. Typological studies in language, 45, 229-254

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. Journal of neuroscience, 17(11), 4302-4311.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. Science, 336(6084), 1049-1054.

Kemps, R. J., Ernestus, M., Schreuder, R., & Baayen, R. H. (2005). Prosodic cues for morphological complexity: The case of Dutch plural nouns. Memory & cognition, 33(3), 430-446.

Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. Brain and Language, 90(1-3), 117-127.

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

Kiet, H. A. T., Baek, S. K., Kim, B. J., & Jeong, H. (2007). Korean Family Name Distribution in the Past. Journal of the Korean Physical Society, 51(5), 1812-1816

Kilarski. M. 2007. On grammatical gender as an arbitrary and redundant category. In Douglas Kilbee, editor, History of Linguistics 2005: Selected papers from the 10th International Conference on the History of Language Sciences (ICHOLS X), pages 24–36. John Benjamins, Amsterdam.

Kim, B. J., & Park, S. M. (2005). Distribution of Korean family names. Physica A: Statistical Mechanics and its Applications, 347, 683-694

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review, 104(2), 211.

Lauderdale, D. S., & Kestenbaum, B. (2000). Asian American ethnic identification by surname. Population Research and Policy Review, 19(3), 283-300.

Levin, B. (1993). English verb classes and alternations: A preliminary investigation. University of Chicago press.

Levy, R. (2008). Expectation-based syntactic comprehension. Cognition, 106(3), 1126–1177.

Lieberson, S., & Lynn, F. B. (2003). Popularity as Taste. Onoma, 38, 235-276.

Lindblom, B. (1990) ‘‘Explaining phonetic variation: A sketch of the H&H theory,’’ in Speech Production and Speech Modeling, edited by W. Hardcastle and A. Marchal Kluwer, Dordrecht", pp. 403–439

Linke M & Ramscar, M (in prep) Context-sensitive articulatory variation at word boundaries increases communicative efficiency of spoken english

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: a network model of category learning. Psychological Review, 111(2), 309.

MacKay, D. J. (2003). Information theory, inference and learning algorithms. Cambridge: Cambridge University Press.

MacWhinney, B. (2014). The CHILDES project: Tools for analyzing talk, Volume II: The database. Psychology Press.

Mandelbrot, B. (1966). Information theory and psycholinguistics: A theory of word frequencies. In P. Lazarsfield & N. Henry (Eds.), Readings in mathematical social sciences. Cambridge: MIT Press

Manin, D. (2009). Mandelbrot’s Model for Zipf’s Law: Can Mandelbrot’s Model Explain Zipf’s Law for Language? Journal of Quantitative Linguistics, 16(3), 274–285

Matisoff, J. A. (1991). Sino-Tibetan linguistics: present state and future prospects. Annual review of anthropology, 20(1), 469-504.

McDonald, S., & Ramscar, M. (2001). Testing the Distributioanl Hypothesis: The influence of Context on Judgements of Semantic Similarity. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 23, No. 23).

McDowell, J. (1977). On the sense and reference of a proper name. Mind, 86(342), 159-185.

McElduff, F., Mateos, P., Wade, A., & Borja, M. C. (2008). What's in a name? The frequency and geographic distributions of UK surnames. Significance, 5(4), 189-192.

McQueen, J. M., & Cutler, A. (2012) Cognitive Processes in Speech Perception. The handbook of phonetic sciences, 116, 489.

Meylan, S. C., & Griffiths, T. L. (2017). Word forms-not just their lengths-are optimized for efficient communication. arXiv preprint arXiv:1703.01694.

Milin, P., Đurđević, D. F., & del Prado Martín, F. M. (2009). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. Journal of Memory and Language, 60(1), 50-64.

Milin, P., Kuperman, V., Kostic, A., & Baayen, R. H. (2009). Paradigms bit by bit: An information theoretic approach to the processing of paradigmatic structure in inflection and derivation. Analogy in grammar: Form and acquisition, 214-252.

Müller, H. M., & Kutas, M. (1996). What's in a name? Electrophysiological differences between spoken nouns, proper names and one's own name. NeuroReport, 8(1), 221-225

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. Journal of experimental psychology: human perception and performance, 17(1), 3.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. Annual review of Psychology, 43(1), 25-53.

Ogden, C. K., & Richards, I. A. (1923). The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism (Vol. 29). K. Paul, Trench, Trubner & Company, Limited.

Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization, UCLA Law Review, 2010, p. 1701

Pereira, F. (2000). Formal grammar and information theory: together again?. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 358(1769), 1239-1253.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. Psychonomic bulletin & review, 21(5), 1112-1130.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. Proceedings of the National Academy of Sciences, 108(9), 3526.

Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability. Speech Communication, (45), 89–95. https://doi.org/10.1016/j.specom.2004.09.001

Port, R. F., & Leary, A. P. (2005). Against formal phonology. Language, 81(4), 927-964.

Port, R., & Leary, A. P. (2005). Against formal phonology. Language, 81, 927-964.\

Proverbio, A. M., Lilli, S., Semenza, C., & Zani, A. (2001). ERP indexes of functional differences in brain activation during proper and common names retrieval. Neuropsychologia, 39(8), 815-827

Quine, W. V. O. (1960). Word and object (Studies in Communication). New York and London: MIT Press.

Quine, W. V. O. (1960). Word and object. MIT press.

Ramscar, M. & Pain, H. (1996) Can a real distinction be made between cognitive theories of analogy and categorisation In Garrison W. Cottrell (ed.), Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society. Lawrence Erlbaum. pp. 346--351

Ramscar, M., & Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. Cognitive science, 31(6), 927-960.

Ramscar, M., Matlock, T., & Dye, M. (2010). Running down the clock: The role of expectation in our understanding of time and motion. Language and Cognitive Processes, 25(5), 589-615.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The Effects of Feature-Label- Order and their implications for symbolic learning. Cognitive Science, 34(6), 909-957.

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

Ramscar, M., Dye, M., & Klein, J. (2013b). Children value informativity over logic in word learning. Psychological Science, 24(6), 1017-1023.

Ramscar, M., Dye, M., & Hübner, M. (2013c). When the fly flied and when the fly flew: How semantics affect the processing of inflected verbs. Language and Cognitive Processes, 28(4), 468-497.

Ramscar, M., Hendrix, P., Love, B., & Baayen, R. H. (2013d). Learning is not decline. The Mental Lexicon, 8(3), 450-481.

Ramscar, M., Smith, A. H., Dye, M., Futrell, R., Hendrix, P., Baayen, H., & Starr, R. (2013e). The 'universal'structure of name grammars and the impact of social engineering on the evolution of natural information systems. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 35, No. 35).

Ramscar, M., & Baayen, H. (2013f). Production, comprehension, and synthesis: a communicative perspective on language. Frontiers in psychology, 4, 233.

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

Ramscar, M. (2015, August 5). The Unspeakable In Pursuit Of The Unrepeatable. The Importance of Being Wrong. Retrieved April 9, 2016, from https://ramscar.wordpress.com/2015/08/05/the-unspeakable-in-pursuit-of-the-unrepeatable/

Ramscar, M., & Port, RF. (2015). Categorization (without categories). In Dabrowska, E., & Divjak, D., Handbook of Cognitive Linguistics. De Gruyter Mouton, pages 75-99.

Ramscar, M., & Port, R. F. (2016). How spoken languages work in the absence of an inventory of discrete units. Language Sciences, 53, 58-74.

Ramscar, M, Sun, C. C., Hendrix, P., & Baayen, H. (2017). The Mismeasurement of Mind: Life-Span Changes in Paired-Associate-Learning Scores Reflect the "Cost" of Learning, Not Cognitive Decline. Psychological Science, 28(8), 1171-1179

Ramskir C.A (1973). Pre-Register Genealogy and a Lost Place Name. The Local historian, 217-252, National Council of Social Service, London

Reddy, M. (1979). The conduit metaphor. Metaphor and Thought, 2, 285-324.

Reed, W. J., & Hughes, B. D. (2003). On the distribution of family names. Physica A: Statistical Mechanics and its Applications, 319, 579-590.

Rescorla, R.A. and Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement, Classical Conditioning II, A.H. Black & W.F. Prokasy, Eds., pp. 64–99. Appleton-Century-Crofts.

Rissanen, J. (1984). Universal coding, information, prediction, and estimation. IEEE Transactions on Information theory, 30(4), 629-636.

Russell, B. (1919). Descriptions. Introduction to mathematical philosophy, 167-80.

Schwichtenberg, B., & Schiller, N. O. (2004). Semantic gender assignment regularities in German. Brain and language, 90(1-3), 326-337.

Scott, J. C. (1998). Seeing like a state: How certain schemes to improve the human condition have failed. Yale University Press

Searle, J. R. (1971). The problem of proper names. Semantics. Cambridge (Mass), 37-53.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. Cognition, 133(1), 140-155.

Shannon, C. E. (1956). The bandwagon. IRE Transactions on Information Theory, 2(1), 3.

Shannon, C.E. (1948), "A Mathematical Theory of Communication", Bell System Technical Journal, 27, pp. 379–423 & 623–656, July & October, 1948.

Sidwell, P., & Blench, R. (2011). 14 the austroasiatic urheimat: the southeastern riverine hypothesis. Dynamics of human diversity, 315.

Smith-Bannister, S. (1997). Names and naming patterns in England, 1538-1700. Oxford University Press.

Song, J. J. (2006). The Korean language: Structure, use and context. Routledge.

Taylor, V. M., Nguyen, T. T., Do, H. H., Li, L., & Yasui, Y. (2011). Lessons learned from the application of a Vietnamese surname list for survey research. Journal of immigrant and minority health, 13(2), 345-351.

Trimmer, P. C., Houston, A. I., Marshall, J. A., Bogacz, R., Paul, E. S., Mendl, M. T., & McNamara, J. M. (2008). Mammalian choices: combining fast-but-inaccurate and slow-but-accurate decision-making systems. Proceedings of the Royal Society of London B: Biological Sciences, 275(1649), 2353-2361.

van Rooij, J. C., & Plomp, R. (1991). The effect of linguistic entropy on speech perception in noise in young and elderly listeners. The Journal of the Acoustical Society of America, 90(6), 2985-2991.

Van Son, R. J. J. H., & Pols, L. C. (2003). How efficient is speech. In Proceedings of the institute of phonetic sciences(Vol. 25, pp. 171-184).

Wedel, A., Nelson, N., & Sharp, R. (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. Journal of Memory and Language, 100, 61-88.

Westbury, C., Shaoul, C., Moroschan, G., & Ramscar, M. (2016). Telling the world's least funny jokes: On the quantification of humor as entropy. Journal of Memory and Language, 86, 141-156.

Wittgenstein, L. (1953). Philosophical Investigations. Oxford: Blackwell.

Wrigley, E. A., Schofield, R. S., & Schofield, R. (1989). The population history of England 1541-1871 (Vol. 46). Cambridge University Press.

Yarlett, D. (2008). Language learning through similarity-based generalization. Unpublished PhD Thesis, Stanford University.

Yuan, Y. (2002). Chinese Surnames. East China Normal University Press, Sanghai, 21-57

Zipf, G. K. (1935). The Psychobiology off Language. Boston: Houghton-Mifflin.

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