Title
The Structure of Names in Memory:Deviations from Uniform Entropy Impair Memory for Linguistic Sequences

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

Authors
Dye, Melody
Johns, Brendan
Jones, Michael
et al.

Publication Date
2016

Peer reviewed

The Structure of Names in Memory:
Deviations from Uniform Entropy Impair Memory for Linguistic Sequences

Melody Dye (meldye@indiana.edu)[1], Brendan Johns (btjohns@buffalo.edu)[2], Michael N. Jones (jonesmn@indiana.edu) [1], & Michael Ramscar (michael.ramscar@uni-tuebingen.de)[3]

[1] Department of Psychological & Brain Sciences, Indiana University, Bloomington
[2] Department of Communicative Disorders and Sciences, University at Buffalo
[3] Department of Linguistics, University of Tübingen

Abstract

Human languages can be seen as socially evolved systems that have been structured to optimize information flow in communication. Communication appears to proceed both more efficiently and more fluently when information is distributed evenly across the linguistic signal. In previous work (Ramscar et al., 2013), we used tools from information theory to examine how naming systems evolved to meet this requirement historically, and how, over the past several hundred years, social legislation and rapid population growth have disrupted naming practices in the West, making names ever harder to process and remember. In support of these observations, we present findings from three experiments investigating name fluency, recognition, and recall. These results provide converging empirical evidence for an optimal solution to name design, and offer a more nuanced understanding of how social engineering has impaired the structure of names in memory.

Keywords: sequence learning; information theory; artificial grammar; associative memory; category fluency

Introduction

Names are central to identity, and the primary means by which we individuate and discriminate ourselves from our peers. However, while they are central to human life, names are unusually difficult to learn and remember. English speakers typically find a person's name more difficult to recall than any other piece of information about them, and names form the majority of naturally occurring 'tip of the tongue' (TOT) experiences (Burke et al., 1991). Moreover, the recall of names is disproportionately impaired in old age, with TOT rates for names increasing more with age than for other words. Indeed, it has been reported that older English speaking adults consider the deterioration in name recollection to be the most disturbing age-related cognitive problem they face, and the aspect of memory that they would most like to improve (Cohen & Faulkner, 1986).

What makes names so challenging? While most nouns are generic, picking out a class of objects, names are *sui generis*, picking out an individual (Cohen, 1990). Given the size of modern populations, the design of an efficient naming system must balance the need to individuate each person against the massive rise in linguistic complexity that unique identifiers would create. Using government databases and historical records, Ramscar et al. (2013) found that historically, Mandarin Chinese and English have converged on a similarly elegant solution to this problem, smoothing identifying information over several name elements each of which is distributed in a highly efficient, Zipfian manner. However, in the recent past, social forces have conspired to disrupt naming practices in the West, altering the shape of these distributions and reducing their efficiency. This presents an excellent opportunity to conduct an *in vivo* experiment on an extant linguistic sub-system, and to investigate how changes to its distributional structure have impacted its representation in memory.

While this paper takes as its focus one highly specialized linguistic subsystem – names – its broader aim is to examine how the descriptive tools offered by information theory may lend insight into how semantic categories are represented in mind.

Principles of Information Theory

In engineering efficient communication systems, the fundamental problem concerns how to code a message generated at point A (the source) into a physical signal that can be transmitted to point B (the destination) as efficiently and reliably as possible, such that the message can still be recovered intact. The answer to this challenge can be found in the set of mathematical principles and conceptual definitions elaborated by Claude Shannon (1948) in his theory of information.

Information theory specifies the theoretical bounds on this project in terms of two measures: Channel capacity defines the maximum rate at which information can be reliably transmitted; Entropy defines the shortest description length at which a random variable can be coded. These principles can be used to assess the design of communication systems of every stripe, including natural language systems.

In the simple case of a ***name grammar***—our focus here—the space of alternatives at the source consists in the population of to-be-named individuals, and the code, in a set of tokens that, in combination, efficiently discriminates among those individuals, allowing for their successful identification. An important question is how such a grammar might be designed such that it minimizes the average length of a name, yet still reliably discriminate the intended individuals. In this, two principles are key: 1) that each element of the grammar approximate a Zipfian distribution, and 2) that the uncertainty over each element remain roughly constant (Ramscar et al., 2013). We assess each precept in turn.

Zipf's Law. Zipf's (1949) Law is a universal scaling law for lexical frequencies and is one of the most striking and empirically robust regularities known to language. For a given sample of text, the frequency distribution of its lexical items approximates an inverse power law, in which a steep decline in frequency over the highest ranked words eases off as rank grows, producing a long tail with similarly low frequencies. Stated formally: If r is rank, $p(w_r)$ is the probability of a word of rank r, and α is the law's exponent, varying between 0 and 1, then:

$$p(w_r) \propto r^{-\alpha}$$

This principle appears to hold not only for words, but for linguistic units at every scale, such as letters, syllables, and even word meanings.

The existence, and the import, of this peculiar distributional feature has been the subject of some debate. In his 'principle of least effort', Zipf (1949) proposed that communicators seek to balance efficiency on the one hand, and comprehensibility on the other, and that these opposing forces minimize communicative effort over time, giving rise to language's characteristic distribution (an intuition formalized by Ferrer-i-Cancho & Solé, 2003). Other explanations have been given in terms of optimal coding theory (Mandelbrot, 1953), preferential attachment (Simon, 1955), and principles of memory search and information retrieval (Anderson & Milson, 1989). Amidst these diverse accounts, what is shared is a notion that the nature of the distribution is—in some sense—optimal, and accordingly, that deviations from it represent departures from optimality.

Regardless of one's theoretical commitments, it is straightforward to empirically establish the cost of deviating by quantifying the increase in entropy that results from altering the shape of the distribution (c.f. Ramscar et al., 2013). The broader question is how or why this might matter.

Entropy Rate Constancy Over the past decade, linguists have interested themselves in an important corollary of Shannon's work on channel capacity: For a particular linguistic sequence, comprised of a set of distinct elements, information transfer is at its theoretical maximum when the entropy over each element is *constant* (and continuously approximating the channel capacity). A growing number of studies have demonstrated that speakers ably conform to this principle by smoothing information over discourse—avoiding excess peaks or troughs in uncertainty through an array of articulatory strategies and lexical and syntactic choices (c.f. Genzel & Charniak, 2002; Keller, 2004; Aylett & Turk, 2004; Bell et al., 2009; Qian & Jaeger, 2012).

Languages themselves also appear to be structurally adapted for entropy management. Nouns provide a particular challenge in this regard, as they are—in many languages—the most diverse part of speech. Accordingly, lexical uncertainty will tend to be at its highest in any given slot where a noun occurs. The impact this has on processing is well-attested: One of the most common sites of disfluencies is at the determiner preceding a noun, and the less likely the noun, the more likely a disfluency (Clark & Wasow, 1998). Nouns are also the most common sites for incorrect lexical retrieval, and an array of other processing difficulties (Vigliocco et al., 1997). However, various linguistic subsystems—such as grammatical gender—can help manage nominal entropy by using preceding elements to subdivide the space of possible upcoming nouns, thereby reducing uncertainty in context (Dye et al., 2015).

Names may pose yet a greater challenge, as they draw from a wider array of phonological sequences than other nouns (Brennen, 1993) and may be less well discriminated by semantic context. However, they appear to be governed by a similar principle. One way to conceptualize a name grammar, which consists in multiple, sequential elements, is as a ***decision tree***. In a classic tree structure, the root node represents a space of mutually exclusive alternatives that are systematically partitioned at various branching points, giving way to specific outcomes at the terminals (Quinlan, 1986). The outcomes can be understood as the to-be-identified individuals; the branch points, as the requisite yes/no questions asked by a clever guesser attempting to discriminate the named individual. In a well-designed system, the information contributed by each branch should be identical, such that the (conditional) information required from one branch to the next remains constant, while the total (unconditional) information grows linearly as a function of path length.

Ideally then, name elements should serve to iteratively increase the degree to which an individual can be identified, while minimizing the uncertainty over each element in context. Experiment 1 tests this claim by varying entropy rate in an artificial name grammar.

Experiment 1

This study examines whether naming systems are more easily learned and remembered when there is a uniform entropy rate across name elements, such that identifying information is evenly distributed over the full name.

Participants

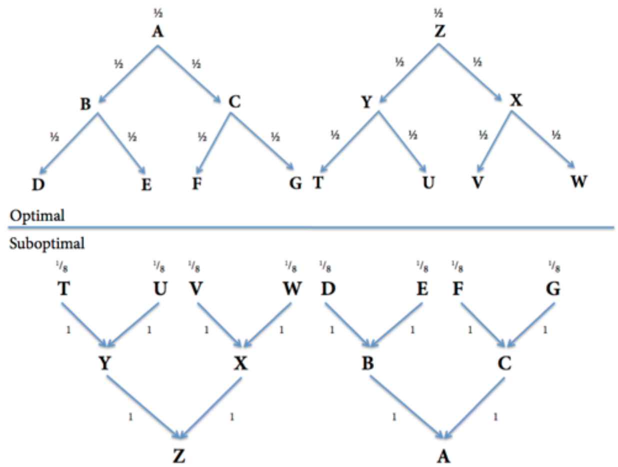
40 students (15 males and 25 females) at Indiana University participated for course credit in an undergraduate psychology course. All were native English speakers, who had completed a prosopagnosia prescreen and simple face recognition task to ensure they could recognize and discriminate faces similar to the experimental stimuli.

Materials

The present study employs a pair of artificial name grammars, comprised entirely of novel names. Each grammar consists in a set of eight full names, drawn from a finite pool of first, middle, and last name elements. The first grammar has a highly efficient information structure, with a small pool of first names followed by successively larger pools of middle and last names. The second grammar, devised to be inefficient, replicates this structure, but with the key difference that the first and last name populations are *reversed*. While names in both grammars convey the same amount of total information, the peak entropy for the suboptimal grammar is three times that of the optimal one. Where the entropy of the first name element is computed independently:

$$H(p) = H(X) = -\sum_{x \epsilon X} p(x) log_2 p(x) = \sum_{x \epsilon X} p(x) \frac{1}{log_2 p(x)}$$

and the entropy of the second and third name elements is conditioned on the preceding string:



$$H(Y|X) = -\sum_{x \epsilon X} p(x) H(Y|X = x) = -\sum_{x \epsilon X} \sum_{y \epsilon Y} p(x, y) log_2 p(y|x)$$

Figure 1: In an optimal name grammar (top), there is a small pool of first names and successively larger pools of second and last name elements. Identifying information is spread evenly over each name element, with elements sharing identical (conditional) probabilities across positions. However, if the grammar is reversed (bottom), then only the first name element is informative; the subsequent elements each have a conditional probability of 1, and thus add no information.

Names Name elements are four-letter, one-syllable nonces drawn from the English Lexicon Project (Balota et al. 2007), with the number of phonological neighbors held constant, and with accuracy and reaction time in pseudolexical decision tasks within one standard deviation of the mean.

Faces Faces are drawn from a body of similarly posed and cropped portraits taken at the Museum of Modern Art in New York City, with faces occupying near-identical spatial positions in each photograph. The set of faces were chosen to be easily discriminable, and were evenly distributed across ethnic groups. To ensure that, for the purposes of the naming study, subjects could easily re-identify the faces in each set, the stimuli were piloted in an old-new recognition task (n=41), with similar looking distractors. Selected faces averaged 91% accuracy.

Procedure

Training At study, participants were asked to learn a series of sixteen face-name pairs, eight of which were female and eight male. The female faces were assigned one set of names, the males another. Which set of name elements was assigned (A or B), and to which face-gender, was counterbalanced across participants. Name assignment within gender was randomized. The choice of grammar (optimal or sub-optimal) was varied across participants, with even assignment according to participant gender.

On each trial, a face appeared on-screen and then a name was read aloud, with the spelling of each name element appearing as it was articulated. The full name was repeated twice in view of the image, followed by a brief mask. The training sequence consisted in three repeated blocks. To control for sequential dependencies at encoding, the order of training was pseudo-randomized, such that faces from each gender were spread evenly across trials, and the same name element was never seen on adjacent trials.

Test At test, subjects completed a cued recall task, in which they were presented with a face and asked to produce its name. Subjects were instructed that partial responses would be counted. Subjects were given 30 seconds in which to complete typing. All pairs were tested exactly once, and response latencies were recorded. Specific face-name pairings were randomized between subjects. Test order was randomized in the same way as training.

Scoring Name accuracy was scored by two raters blind to the hypotheses tested. 1 point was awarded for every correct name element in the correct position. ½ a point was awarded if the element was correct, but in the wrong position, or if the name element was incorrect, but highly similar (phonologically) to the target. Any disagreements between raters were resolved through discussion.

Results

There was substantial variability in subject performance on the recall task in both conditions, with mean performance of 53.6% and a standard deviation of 26.7%. A by items analysis revealed that names trained with the optimal grammar were recalled significantly faster and more accurately than their rearranged versions (Correct Log RT: $t(15)=-3.36$, $p=.002$; Recall Score: $t(15)= 3.91$, $p<.001$; Figure 2). A parallel analysis for faces yielded comparable results (Correct Log RT: $t(15)=-4.42$, $p<.001$; Recall Score: $t(15)=2.57$, $p=.01$). An independent-samples by subjects analysis conformed with the general direction of results, but did not reach significance (Correct Log RT: $t(34)=-1.26$, $p=.108$; Recall Score: $t(38)= 1.04$, $p=.15$).

Discussion

This study revealed that names that complied with entropy rate constancy were more quickly and accurately recalled than those that did not.

The Price of Deviating

Over the past several hundred years, social legislation and rapid population growth have disrupted naming practices in the West. In particular, the rise of nation states has coincided with the adoption of census policies that have made family names hereditary, rendering them fixed, rather than mutable. In English-speaking countries, where the family name has traditionally been positioned last, this has created significant pressure on the first name to provide the individuating information, resulting in a greater diversity of higher entropy first names. Curiously, the collapse of efficient naming has not affected society equally: In the US, women and African Americans are disproportionately burdened with the least efficient names. Evidence from modeling simulations suggests that this flattening of the distribution should have negative consequences for name memory (Ramscar et al., 2014).

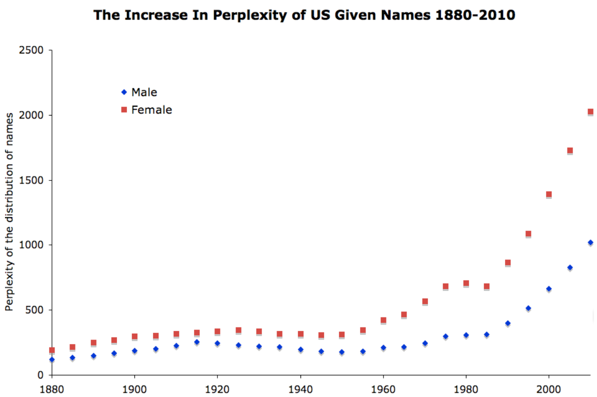


Figure 2: The perplexity of male and female names with a count ≥ 5 in US Social Security applications at 5 year intervals from 1880 to 2010 (Ramscar et al., 2013).

Are male names indeed easier to recollect than female names in American English? One of the more commonly used measures to assess semantic memory is ***category fluency***, in which participants produce as many items from a specific category as possible within a specified time frame. Fluency data are then analyzed by counting the number of items produced, and assessing the factors that influence transitions between category items.

In a fluency task, subjects tend to be superior at producing items that they have more familiarity with, all things being equal. For example, a doctor might more readily reel off a string of diseases, and a chef, a list of cooking implements. Domain-specific knowledge and experience are also reflected in gender differences: Capitani et al. (1999) report that men are advantaged in recalling man-made items, such as tools and vehicles, whereas women show more facility at food stuffs, like fruits and vegetables.

In Experiment 2, subjects of both genders were tested on their ability to produce male and female names. Given sex segregation in peer and social groups across development and early adulthood (Mehta & Strough, 2009), there is good reason to suspect that males and females will have varying experience with names in accordance with their gender. If, on average, male subjects have more experience of male names and female subjects of female names, than a double dissociation in subject performance should be observed, wherein males produce more male names, and females more female names.

An alternative possibility, assessed here, is that there may be an interaction between both the frequency with which a category is sampled and its distributional structure—a structure which may be more or less optimal for retrieval purposes (Anderson & Milson, 1989). If the organization of female names makes them harder to access in memory, this may diminish, or even eliminate, the predicted female advantage for female names.

Experiment 2

This study examines the effect of the distributional structure of male and female names on name fluency. Subjects were asked either to recall 1) the names of as many famous living people as they could, or 2) as many first names as they could.

Participants

146 students (73 males and 73 females) at Indiana University participated for course credit in an undergraduate psychology course. All were native English speakers. Subjects were randomly assigned to one of the task versions. 15 subjects were dropped from analysis for failing to follow instructions (e.g., by mixing genders), leaving n=33 for first names collected separately, 36 for first names collected together, 28 for famous names collected separately, and 34 for famous names collected together.

Design

The fluency tasks were administered in the online survey program Qualtrics. Two versions of each task were administered. In the first version of the task, no gender was specified, and subjects simply recalled as many names as they could within two minutes. In the second, subjects were asked to produce the names for one gender, and then the other, with ninety seconds for each task. Which gender was requested first was counterbalanced according to the gender of the participant.

Results

A 2 (subject gender) x 2 (name gender) repeated measures ANOVA was run over each version of the task design, revealing the same basic pattern of results across tasks: a significant main effect of name gender – more male names were produced overall – and a significant interaction of subject gender and name gender. In each case, post hoc analyses (Tukey HSD) indicated that while female subjects produced similar numbers of male and female names, male subjects produced significantly more male names than female names (p<0.001).

Modeling Fluency

To model participants' search through memory in the first name fluency task, the Luce Choice Axiom was used to quantify the contribution of different linguistic factors to first name retrieval. Given a set of possible responses, the axiom defines the probability of responding to stimulus S_i with response R_j as:

$$P(R_j|S_i) = \frac{\beta_j^{\lambda_0} S(i,j)^{\lambda_1}}{\sum_{k=1}^{n} \beta_k^{\lambda_0} S(i,k)^{\lambda_1}}$$

where β_j is the response bias for item j, and $S(i,j)$ is the similarity between items i and j. In our application of the decision rule, the response set is the full set of first names produced in the task, and $P(R_j|S_i)$ is the probability of producing a particular name given the last name that was produced. The response bias β_j scales with name frequency, and the item similarity S(i,j) with the semantic similarity between one name and the next. Frequency estimates were calculated from a 20 million sentence Wikipedia corpus, and semantic similarity from the vector cosine of the names' semantic representations, as computed by the distributional semantic model BEAGLE (Jones & Mewhort, 2007) run over that same corpus. The parameters λ_0 and λ_1 control the relative contributions of frequency and semantic similarity in producing the response. For each participant's name sequence, we determined the set of parameters that were most likely to have generated the observed data, using a model comparison technique (see Johns et al., 2013 for detailed methodology).

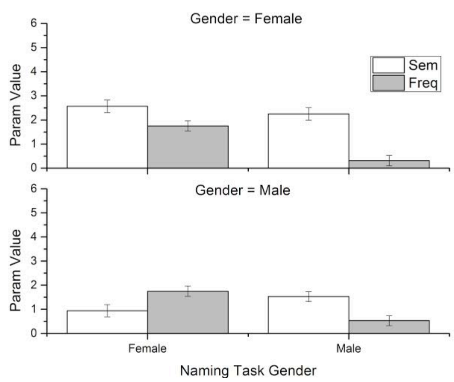


Figure 3: Fluency model of first name production.

Across subject genders, there was a significant effect of the semantic parameter [$F(1,65)=4.65$, $p<0.05$], indicating that overall, female subjects relied more on semantic information in producing first names than their male counterparts. In comparing the two tasks – producing male names vs. producing female names – there were main effects of both semantics [$F(1,65)=8.716$, $p<0.01$] and frequency [$F(1,65)=33.805$, $p<0.001$], suggesting that the influence of these parameters varied by name gender. In particular, male naming relied more heavily on semantics, while female naming relied more on frequency, overall. However, there was also a significant interaction of semantics with subject gender [$F(1,65)=17.417$, $p<0.001$]: female subjects made significantly more use of semantics in generating female names than did male subjects.

Discussion

Across four name fluency tasks, a robust interaction between subject gender and name gender was observed: While female subjects recalled similar numbers of male and female names, male subjects recalled significantly more male names than female names. The absence of a mirror effect across genders is surprising: As a consequence of gender homophily in their social networks, female subjects should have more experience, on average, with female names, which should support fluency. Yet in all but one condition, female subjects actually produced more male names overall.

There were also clear differences in the strategies deployed in generating male and female names from memory. An analysis of the factors contributing to name production indicated that subjects relied more on semantics in generating male names and more on frequency in generating female names. This difference in strategy may be telling. In related work on verbal fluency, Johns et al. (2013) found that subjects who would go on to be diagnosed with mild cognitive impairment showed increased reliance on frequency cues in sampling the semantic space. Similarly, Taler et al. (2013) report that when bilinguals are forced to switch between languages, they reduce their use of semantic cues, favoring frequency instead. This suggests that when the task is made more difficult, or when the semantic space is degraded, speakers switch more towards a baseline sampling procedure that prioritizes frequency as the guiding cue.

Taken together, these findings lend support to the hypothesis that the distributional properties of female names make them harder to recall than male names. However, there is another possibility: Female subjects may simply have more *equal* experience with male and female names, given the overrepresentation of males in the broader cultural canon (e.g., educational materials and popular media). Experiment 3 was designed to assess and control for the effects of prior familiarity.

Experiment 3

If female names are less memorable than male names in American English, they should take longer to recognize in association with an individual, even when familiarity and name length are taken into account. In the first half of the experiment, our subjects were tested on their ability to recognize famous faces, which can be interpreted as a latent name generation task. Subjects viewed images of famous individuals intermixed with similarly posed and photographed non-famous distractors, and were asked to indicate whether they thought they recognized the individual or not. In the second half of the experiment, subjects viewed the same famous individuals, but this time, paired with either their actual name (hit) or a same-gender lure (false alarm). Subjects were asked to determine whether the names and faces were accurately matched or not. Across both tests, accuracy and response latency were recorded.

Participants

48 students (24 males and 24 females) at Indiana University participated for course credit in an undergraduate psychology course. All were native English speakers.

Materials

The famous names collected in Experiment 2, along with search volume data from Google Trends and overall Google hits, was used to create an inventory of a hundred famous individuals that college-age subjects were likely to recognize. **Names** The famous men and women whose names were accurately paired with their pictures (hits) were matched for name length and fame. This was also true for the men and women whose faces were mismatched with novel names (lures). Novel names were generated using naming data from the last fifty years of the Social Security Index, and were matched on length, and first and last name popularity. Notably, it was not possible to precisely match either set of names for syllable length, although this was balanced to the extent possible. Popular American female names are longer, on average, and contain more syllables than male names.

Faces Photographs of each famous individual were representative forward-facing poses taken from the first page of Google Image search results, and were resized and cropped, such that the face of the individual took up a similar portion of the frame. Photographs of foils were taken from similarly posed and cropped professional portraits taken at the Museum of Modern Art in New York City.

Results

In the face recognition task, a 2 (subject gender) x 2 (face gender) repeated measures ANOVA revealed a significant interaction of subject gender and face gender for both accuracy [$F(1,46) = 19.58$, $p < .001$] and correct log RT [$F(1,46) = 6.59$, $p=.0135$]. Post hoc analyses (Tukey HSD) indicated that female subjects correctly identified more women than men ($p<.05$), and that male subjects not only identified more men than women ($p<.02$), but were also faster in those correct identifications ($p<.001$).

In the name recognition portion, a similar analysis revealed a significant interaction of subject gender and name gender on accuracy [$F(1,46) = 5.937$, $p=.019$], and both a main effect of name gender [$F(1,46) = 18.09$, $p<.001$] and a significant interaction of subject gender and name gender [$F(1,46) = 16.08$, $p<.001$] on correct log RT. Post hoc analyses revealed that female subjects correctly identified the names of more women than men ($p=.045$), and that male subjects not only correctly identified more men than women ($p=.012$), but again, were significantly faster in their identifications of them ($p<.0001$).

A more detailed analysis, in terms of hits and false alarms, revealed a parallel pattern of results, with the additional finding that in the name recognition task, both male *and* female subjects were significantly faster at correctly rejecting male lures.

Discussion

In both recognition tasks, classification accuracy patterned by gender, lending credence to the hypothesis that our male subjects have, on average, more exposure to men and male names, and our female subjects to women and female names. However, while male subjects were consistently faster in their identifications of males across both tasks; the inverse trend did not hold for our female subjects. More tellingly, female subjects actually proved faster at correctly rejecting *male* lures in the name recognition task, suggesting that they found these names easier to process.

General Discussion

Previous empirical work supports the notion that names are uniquely difficult to learn and remember compared to other semantic categories. For example, subjects find it harder to learn to pair faces with names, than to pair faces with occupations, even when the frequencies of names and occupations are held constant, and even when the same lexical items are used (e.g., ‘baker’ as both a surname and occupation; Cohen & Faulkner, 1986; McWeeny, Young, Hay, & Ellis, 1987). The experiments and analyses presented here offer insight into how the distributional structure of names—and other semantic categories, more broadly—can affect memory for those items across an array of tasks. In particular, selectively increasing entropy over a given element appears to disrupt memory not only for that element, but for the linguistic sequence it belongs to.

Acknowledgments

This research was funded by an NSF graduate fellowship to MD. We thank Richard Shiffrin and lab for comments and discussion.

References

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*(4), 703–719.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis. *Language and Speech*, *47*(1), 31–56.

Balota, D.A. et al. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445-459.

Bell, A. et al. (2008). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111.

Brennen, T. (1993). The difficulty with recalling people’s names: the plausible phonology hypothesis. *Memory*, *1*(4), 409–431.

Burke, D. M. et al. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, *30*(5), 542–579.

Capitani, E., Laiacona, M., & Barbarotto, R. (1999). Gender affects word retrieval of certain categories in semantic fluency tasks. *Cortex*, *35*(2), 273–278.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*(3), 201–242.

Cohen, G. (1990). Why is it difficult to put names to faces? *British Journal of Psychology*, *81*(3), 287–297.

Cohen, G. & Burke, D.M. (1993). Memory for proper names: A review. *Memory*, 1, 249-263.

Dye, M., Milin, P., Futrell, R., & Ramscar, M. (2015). A functional theory of gender paradigms. In F. Kiefer, J.P. Blevins, & H. Bartos (Eds.) *Morphological Paradigms and Functions*.

Ferrer i Cancho, R., & Sole, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, 100(3), 788–791.

Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text (pp. 199–206). In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*.

Griffin, Z.M. (2010). Retrieving personal names, referring expressions, and terms of address. *The Psychology of Learning and Motivation*, 53, 345-387.

Johns, B.T. et al. (2013). Using cognitive models to investigate the temporal dynamics of semantic memory impairments in the development of Alzheimer’s disease. In the *Proceedings of the 12th ICCM*.

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1–37.

Mandelbrot, B. (1953). An informational theory of the statistical structure of language. *Communication Theory*.

McWeeny, K. H. et al. (1987). Putting names to faces. *British Journal of Psychology*, *78*(2), 143–149.

Mehta, C.M. & Strough, J. (2009). Sex segregation in friendships and normative contexts across the life span. *Developmental Review*, 29, 201-220.

Qian, T., & Jaeger, T. F. (2012). Cue Effectiveness in Communicatively Efficient Discourse Production. *Cognitive Science*, *36*(7), 1312–1336.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106.

Ramscar et al. (2013) The ‘universal’ structure of name grammars and the impact of social engineering on the evolution of natural information systems. *Proceedings of the 35th Meeting of the Cognitive Science Society*, Berlin, Germany.

Ramscar, M. et al. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science*, *6*(1), 5–42.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423, 623–656.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42, 425–440.

Taler, V. et al. (2013). A computational analysis of semantic structure in bilingual fluency. *Journal of Memory and Language*, 69, 607-618.

Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical Gender Is on the Tip of Italian Tongues. *Psychological Science*, 8(4), 314–317.

Zipf, G.K. (1949). Human behavior and the principle of least effort: An introduction to human ecology. Addison-Wesley: Cambridge, MA.