Evolution of emotion semantics (Supplementary Information)

Aotao Xu[1*], Jennifer E. Stellar[2], Yang Xu[3]

[1]Department of Computer Science, University of Toronto

[2]Department of Psychology, University of Toronto

[3]Department of Computer Science, Cognitive Science Program, University of Toronto

[*]To whom correspondence should be addressed: John Xu, a26xu@cs.toronto.edu.

Contents

List of Figures

List of Tables

1 Word age and prototypicality

Here we further demonstrate the differences between prototypicality in emotions and prototypicality in birds. Previous work has suggested that prototypical emotion concepts are well-defined and may have particularly strong social or cultural scripts (Barrett, 2006; Fehr & Russell, 1984; Russell & Barrett, 1999). Thus it is conceivable that words for prototypical emotion concepts exist in a lexicon prior to those less prototypical emotion words. In contrast, prototypicality of birds is based on biological taxonomy (Boster, 1988) and grounded in sensory and visual perception (Garrard, Lambon Ralph, Hodges, & Patterson, 2001), and we do not expect prototypicality to be reflected in the age of a word; it is likely that a relatively newly documented passerine (e.g., bluebird) entered the lexicon after well-established non-passerines (e.g., chicken). To test these ideas, we analyzed the correlation between word age and prototypicality in the categories of emotion and birds.

Following existing work (Xu, Ramiro, & Xu, 2019), we obtained the age of a word from the Historical Thesaurus of English (HTE) (Kay, Roberts, Samuels, & Wotherspoon, 2017). For each word entry, the HTE provides a list of senses of the word, and for each sense, the HTE provides the word class associated with that sense of the word and the date of first appearance of the sense in historical records. We operationalized the date of (the first) emergence of a word to be the earliest date among the dates of first appearance across all of its senses. Since our analyses focused on nouns, we considered only noun senses. We did not analyze the age of French words due to the unavailability of comparable French dictionaries.

We analyzed the same lists of English emotion words and bird names described in *Section 3.1* and *Section 4.1*, which we intersected with the HTE data. For English emotion words, the Pearson correlation between emerging date and prototypicality is -0.366, $p < 0.001$, $n = 135$. This indicates that prototypical emotion words emerged earlier in the history of English. The same pattern does not hold for the bird category. For Rosch bird names, the Pearson correlation between emergence date and prototypicality is -0.0657, $p = 0.643$, $n = 52$. See Figure S1 for illustration.

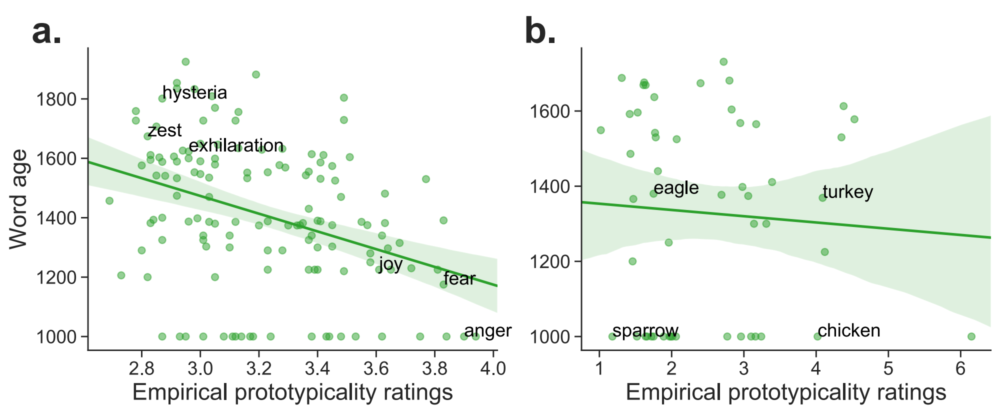

Figure S1: Scatter plots showing the relations between prototypicality and word age for a) emotion words and b) bird names. Each dot corresponds to a word (with a sample set of words annotated), and each band shows a 95% confidence interval for regressions between prototypicality and age.

2 Evaluation of the nearest-neighbour measure

We show that the nearest-neighbour measure of semantic change described in *Section 2* is 1) robust to variation in neighbourhood size (denoted by k), and 2) interpretive for a word's semantic change based on nearest neighbours retrieved at different time points.

2.1 Robustness in neighbourhood size k

We evaluate whether the nearest-neighbour measure is robust to variation in k. Following *Section 3.1*, we again quantify semantic change by setting $t_1 = 1890$ and $t_2 = 1990$ and focus our analysis on the lists of English and French emotion words we analyzed in the main text. We show that for $k = 20, 40, 60, 80$, and 100, resulting degrees of semantic change $rate(w, t_1, t_2)$ are highly correlated. The correlation results are summarized in Figure S2. We observed that in both English and French, the degrees of change are strongly correlated between any two of the predetermined settings of k, with a small decrease in correlations for the lowest value of $k = 20$.

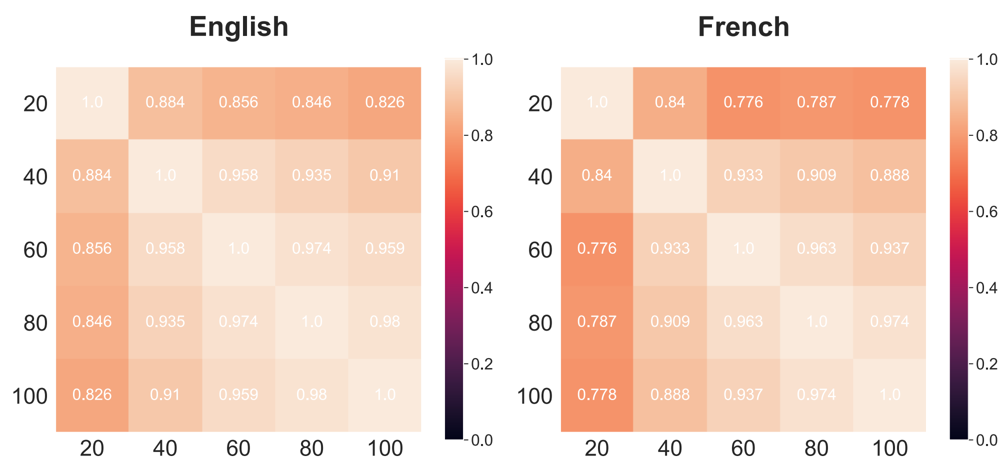

Figure S2: Robustness of *kNN* demonstrated using emotion words. The first column shows results for English and the second column shows results for French. Each cell shows the Pearson correlation between changes measured by x-nearest neighbours and by y-nearest neighbours. All p-values are significant and less than 0.001.

2.2 Examples of emotion semantic change

Qualitative changes in the nearest neighbours of a word offer interpretability for semantic change. We provide examples of emotion words that underwent the most and the least changes and their nearest semantic neighbours in Tables S1, S2, S3, and S4. For example, in Table S1, we observed *zest*, which used to primarily convey joy but later became primarily associated with food, is among the most changing emotion words; on the other hand, in Table S2, we observed words like *surprise* has barely changed in meaning. Similarly, in Table S3, we observed *stupéfaction* in French become less associated with despair and anxiety over time; in Table S4, the least changing words on the other hand tend to preserve similar emotion words as neighbours.

Most Changing	Nearest Neighbours in 1890s	Nearest Neighbours in 1990s
zest	relish, enjoyment, sprightliness	juice, teaspoons, vinegar
infatuation	priestcraft, devastations, misanthrope	inhomogeneity, palates, pleurisy
sentimentality	cant, sentimentalism, rusticity	polyphony, sterne, mandel
optimism	pessimism, aptness, sentimentalism	pessimism, insecurity, enthusiasm
exhilaration	mountebank, festivity, tulip	joy, sadness, excitement
aggravation	misery, symptoms, consequences	proxies, sleeplessness, stressor
exasperation	vehemence, peevishness, pitch	astonishment, amazement, disgust
glee	merriment, wassail, delight	claps, shouts, megaphone
cheerfulness	hopefulness, sprightliness, vivacity	serenity, blasphemies, cannonade
gaiety	sprightliness, vivacity, gayety	pontellier, faints, plaudits
fondness	liking, peevishness, passion	affection, pianists, groanings
hysteria	neurasthenia, simulation, melancholia	neurosis, hypochondriasis, psychosis
dejection	despondency, sullenness, irresolution	pantomime, theseus, disquiet
elation	sullenness, despondency, peevishness	despair, revulsion, dread
ferocity	fierceness, cruelty, prowess	vigor, fury, proverb
revulsion	feeling, disquietude, outburst	disgust, hisses, yearnings
isolation	loneliness, disorganization, seclusion	monger, characterization, coli
alienation	eviction, property, repugnancy	helplessness, blauner, resentment
hopelessness	uselessness, futility, helplessness	helplessness, despair, frustration
rapture	ecstasy, delight, joy	joy, indignation, outcasts

Table S1: Top 20 most changing English emotion words along with their 3 nearest neighbours in the flanking decades.

Least Changing	Nearest Neighbours in 1890s	Nearest Neighbours in 1990s
grief	sorrow, anguish, joy	sorrow, sadness, anguish
pity	compassion, love, sympathy	compassion, shame, sadness
misery	wretchedness, miseries, degradation	sorrow, bitterness, anguish
disgust	horror, aversion, indignation	sadness, annoyance, amazement
anger	indignation, resentment, rage	resentment, rage, frustration
surprise	astonishment, amazement, dismay	astonishment, amazement, dismay
sorrow	grief, anguish, sadness	grief, sadness, misery
affection	affections, tenderness, esteem	admiration, sympathy, love
happiness	felicity, prosperity, welfare	prosperity, joy, enjoyment
despair	desperation, dismay, rage	anguish, frustration, sadness
fear	dread, anger, shame	dread, resentment, anger
horror	terror, astonishment, amazement	terror, astonishment, amazement
regret	disappointment, grief, sorrow	disappointment, sadness, bitterness
envy	jealousy, uncharitableness, hatred	jealousy, hatred, resentment
disappointment	mortification, grief, regret	frustration, sadness, regret
rage	fury, anger, indignation	anger, fury, indignation
shame	disgrace, infamy, blush	guilt, pity, humiliation
astonishment	amazement, surprise, dismay	amazement, dismay, surprise
joy	gladness, delight, grief	delight, sorrow, excitement
sympathy	sympathies, compassion, affection	affection, admiration, compassion

Table S2: Top 20 least changing English emotion words along with their 3 nearest neighbours in the flanking decades.

Most Changing	Nearest Neighbours in 1890s	Nearest Neighbours in 1990s
stupéfaction	indiscrétion, désespoir, anxiété	désapprobation, émotion, allégresse
suspicion	inculpation, défiance, prévention	méfiance, défiance, incertitude
culpabilité	réussite, identité, présomption	infériorité, persécution, châtiment
déplaisir	étonnement, inquiétude, appréhensions	plaisir, mâle, océans
torpeur	engourdissement, apathie, léthargie	apathie, consternation, stupeur
extase	contemplation, stupeur, somnambulisme	contemplation, joie, angoisse
soupçon	soupçons, équivoque, délit	préjugé, partialité, préjugés
hystérie	épilepsie, diabète, étiologie	névrose, épilepsie, névroses
séduction	adultère, entraînements, cruauté	immédiateté, impiété, éloquence
désolation	épouvante, dévastation, misère	pauvreté, saleté, nausées
froideur	bonhomie, trousseau, bassesse	arrogance, ingratitude, insolence
excitation	irritation, nerfs, nerf	tension, angoisse, agitation
intimidation	violence, ruse, corruption	menaces, coercition, chantage
timidité	fierté, délicatesse, naïveté	docilité, avidité, découragement
tension	volts, pression, potentiel	appareillage, excitation, tensions
intérêt	intérêts, utilité, équité	intérêts, utilité, rentabilité
espérance	espoir, espérances, désir	espoir, espérances, désir
dépit	mépris, défaillances, hésitations	grâce, précocité, conséquence
allégresse	fierté, épouvante, gaieté	joie, émotion, gaieté
effusion	sang, tendresse, larmes	sang, amertume, lucre

Table S3: Top 20 most changing French emotion words along with their 3 nearest neighbours in the flanking decades.

Least Changing	Nearest Neighbours in 1890s	Nearest Neighbours in 1990s
tristesse	angoisse, amertume, effroi	amertume, douleur, angoisse
tendresse	bienveillance, sympathie, sollicitude	douceur, compassion, amour
patience	courage, persévérance, prudence	courage, persévérance, audace
orgueil	vanité, amour, ambition	arrogance, insolence, vanité
horreur	honte, effroi, angoisse	opprobre, effroi, tristesse
effroi	tristesse, consternation, terreur	tristesse, consternation, horreur
indignation	admiration, effroi, cri	enthousiasme, cri, admiration
joie	tristesse, douleur, bonheur	tristesse, enthousiasme, douleur
honte	horreur, chagrin, humiliation	opprobre, humiliation, peur
jalousie	haine, ambition, convoitise	arrogance, haine, rancune
colère	désespoir, mécontentement, anxiété	désespoir, fureur, émotion
douleur	chagrin, souffrance, douleurs	souffrance, tristesse, douleurs
stupeur	tristesse, angoisse, effroi	consternation, surprise, effroi
vengeance	haine, ressentiment, fureur	haine, jalousie, orgueil
bonheur	malheur, gloire, joie	malheur, joie, plaisir
chagrin	douleur, tristesse, honte	tristesse, douleur, amertume
terreur	effroi, horreur, haine	horreur, effroi, anarchie
enthousiasme	ardeur, joie, admiration	joie, admiration, indignation
impatience	anxiété, indignation, angoisse	joie, humeur, espoir
souffrance	douleur, tristesse, angoisse	douleur, angoisse, souffrances

Table S4: Top 20 least changing French emotion words along with their 3 nearest neighbours in the flanking decades.

3 Additional analyses of emotion semantic change

We describe three additional analyses that corroborate our findings on emotion semantic change in the main text. The first analysis rules out the possibility that our findings are an artifact of the non-emotion senses of polysemous emotion concepts (e.g., *zest*). The second analysis shows that additional predictors based on hypernymy-hyponymy and degrees of polysemy do not subsume the effects of prototypicality on emotion semantic change. The third analysis provides evidence that our results on emotion concepts were not caused by artifacts in our estimation of prototypicality.

3.1 Category-bounded analysis

We investigate the robustness of our semantic change analyses by considering a variant of the nearest-neighbour measure discussed in *Section 2*. Originally in *Section 2*, the degree of semantic change of a word w is defined as the Jaccard distance between its nearest neighbours at time t_1, $kNN(t_1)$, and its nearest neighbours at a later time, $kNN(t_2)$, where kNN is restricted to nouns in the entire lexicon and determined by cosine similarity over word vectors. Since the meaning of every emotion word is represented by one word vector only, the set kNN might also capture meaning change with respect to the word's polysemes and homonyms, i.e., meaning change outside the category of emotion. To assess how such meaning change might affect our results, we restricted the set of nearest neighbours so that only the list of emotion words are included, i.e., a category-bounded analysis of emotion semantic change. Since the set of emotion words is much smaller than the full lexicon, we set the size of the neighbourhood to be $k = 25$.

We first provide evidence that this variant of the nearest-neighbour measure is also capable of capturing semantic change by showing 1) this measure is positively correlated with the original nearest-neighbour measure, and 2) this measure captures the negative relationship between frequency and semantic change (Hamilton, Leskovec, & Jurafsky, 2016). We obtain degrees of change under this variant measure by following the same procedure described in *Section 2*. In the case of English emotion words, the Pearson correlation between degrees of semantic change measured by this variant and degrees obtained by the original measure is 0.751, $p < 0.001$, $n = 123$; the Pearson correlation between degrees of change measured by the variant and frequency is -0.489, $p < 0.001$, $n = 123$. In the case of French emotion words, the Pearson correlation between degrees of change measured by this variant and degrees obtained by the original measure is 0.604, $p < 0.001$, $n = 112$; the Pearson correlation between degrees of change measured by the variant and frequency is -0.203, $p = 0.0318$, $n = 112$. These results suggest this variant is capable of replicating patterns of change identified previously.

After validating this variant measure, we also repeated the analyses on emotion semantic change described in the main text. Figure S3 shows a significant negative correlation between prototypicality and degree of semantic change: for English, $\rho = -0.535$, $p < 0.001$, $n = 123$; for French, $\rho = -0.558$, $p = 0.002$, $n = 112$. Figure S4 shows multiple regression results similar to the results presented in the main text. The adjusted R^2 for English is 0.432, with $p < 0.001$, $n = 123$. Mean regression coefficients for prototypicality ($\beta = -0.479$, $p < 0.001$) and frequency ($\beta = -0.0356$, $p < 0.001$) remained negative and significant, whereas valence ($\beta = 0.0112$, $p = 0.091$) is insignificant. Again, results hold similarly for French with the adjusted $R^2 = 0.381$, $p < 0.001$, $n = 112$ (prototypicality $\beta = -0.600$, $p < 0.001$; frequency $\beta = -0.0208$, $p < 0.001$; valence $\beta = -0.0015$, $p = 0.599$). Compared to the main results, we observed that prototypicality remains a competitive predictor of semantic stability relative to frequency.

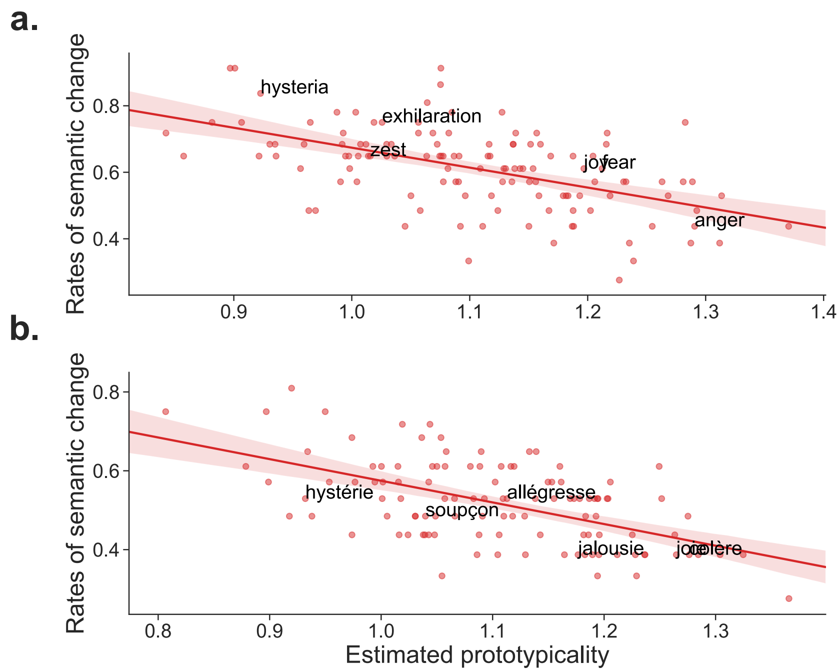

Figure S3: Scatter plots showing the negative correlations between emotion prototypicality and rates of semantic change between the 1890s and 1990s, in a) English and b) French. Each dot corresponds to an emotion term (with a sample set of words annotated), and each band shows a 95% confidence interval for regressions between prototypicality and rates of semantic change.

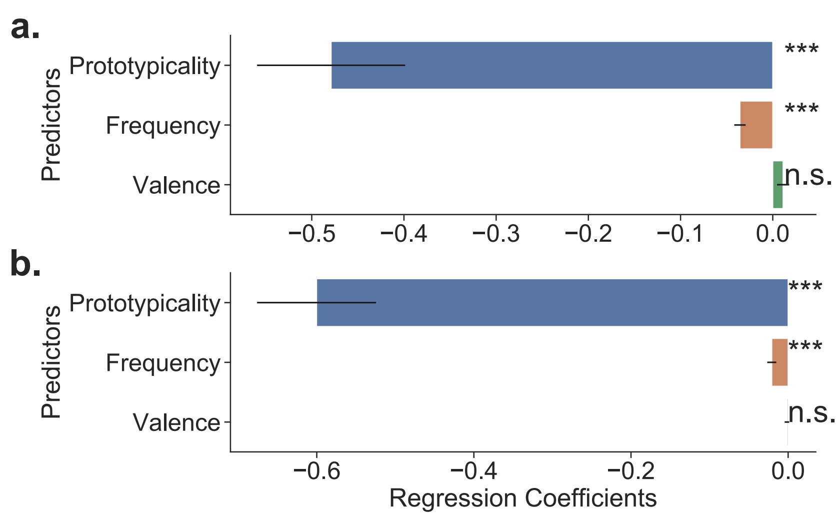

Figure S4: Predictor coefficients from multiple regressions on rates of semantic change. Error bars show standard error, and “n.s.”, “*”, “**”, “***” denote no significance at $p < 0.05$, and $p < 0.05$, $p < 0.01$, $p < 0.001$ respectively. a) shows results for English, and b) shows results for French.

3.2 Other factors of semantic change

In the main text, we tested prototypicality as a predictor of semantic stability alongside frequency. Here we examine the role of prototypicality in predicting semantic stability by controlling for three additional predictors: 1) the degree of polysemy of a word, 2) superordinate-subordinate relations between emotion words, and 3) the age of acquisition (AoA) of a word. Firstly, similar to frequency, we control for degree of polysemy because it is a general predictor of semantic change which has been found to negatively correlate with stability in meaning (Hamilton et al., 2016; Luo & Xu, 2018). Secondly, since one function of prototypical emotion words is that they can help define more complex emotion words (Johnson-Laird & Oatley, 1989) and this anchoring function may provide relative semantic stability,[1] we examine superordinate-subordinate relations (i.e., the hierarchy in the semantic category of emotion) as a potential confounding variable influencing both prototypicality and semantic change. Thirdly, since prototypical emotion words are relatively well-defined and would be easy to learn, and that AoA is a known predictor of stability in lexical change (Monaghan, 2014; Monaghan & Roberts, 2019), we control for AoA as a potential mediator between prototypicality and semantic change.

Following Luo and Xu (2018), we operationalized the degree of polysemy of a word as the number of senses the word had at the starting time $t_1 = 1890$ according to the HTE (Kay et al., 2017). To operationalize superordinate-subordinate relations, we used WordNet (Fellbaum, 1998) provided by NLTK (Loper & Bird, 2002). Specifically, we constructed a directed graph based on hypernym-hyponym relations, where the root is the sense for *feeling*, and the other nodes correspond to the most frequent sense of an emotion word (see Figure S5 for illustration). Then, we quantified a word's degree of subordination as its depth in the graph. For example, the word *thrill* has a depth of 4, while *joy* has a depth of 2. Furthermore, to match the historical period of our analyses, we used objective, test-based measurements of AoA originally published by Dale and O'rourke (1981). A digitized version of this data was obtained from Brysbaert and Biemiller (2017), where each entry contains a word form, its meaning, and the age at which it was acquired. We computed the AoA of a word by taking the average over all entries in which it appears. Due to the lack of analogous French historical data, we only focused on English emotion words. We assumed these hypernym-hyponym relations and AoA are stable over the past century. After intersecting WordNet and AoA with our historical resources described in the main text, we had a total of 109 English emotion words.

We analyzed these factors using our materials and methods described in *Section 3.1* and *Section 3.2* of the main text. Specifically, we computed the rates of change for every emotion concept x, $rate(x, 1890, 1990)$. Then, we performed a multiple regression using the following formula:

$$rate(x, 1890, 1990) \sim p(x|c = emotion) + freq(x) + poly(x) + val(x) + depth(x) + AoA(x) \quad (1)$$

where $poly(x)$ is the degree of polysemy of the word operationalized by number of senses, $depth(x)$ is the depth of the concept in the hypernym-hyponym graph we constructed, and $AoA(x)$ is the age at which the word was acquired.

Figure S6 shows the multiple regression results suggesting the dominant roles of prototypicality and frequency. The adjusted R^2 of the model is 0.697, with $p < 0.001$, $n = 109$; mean regression

[1] For example, suppose *joy* is part of the definition of *exhilaration*; if the meaning of *joy* changed, the meaning of *exhilaration* will necessarily change as well, but not vice versa.

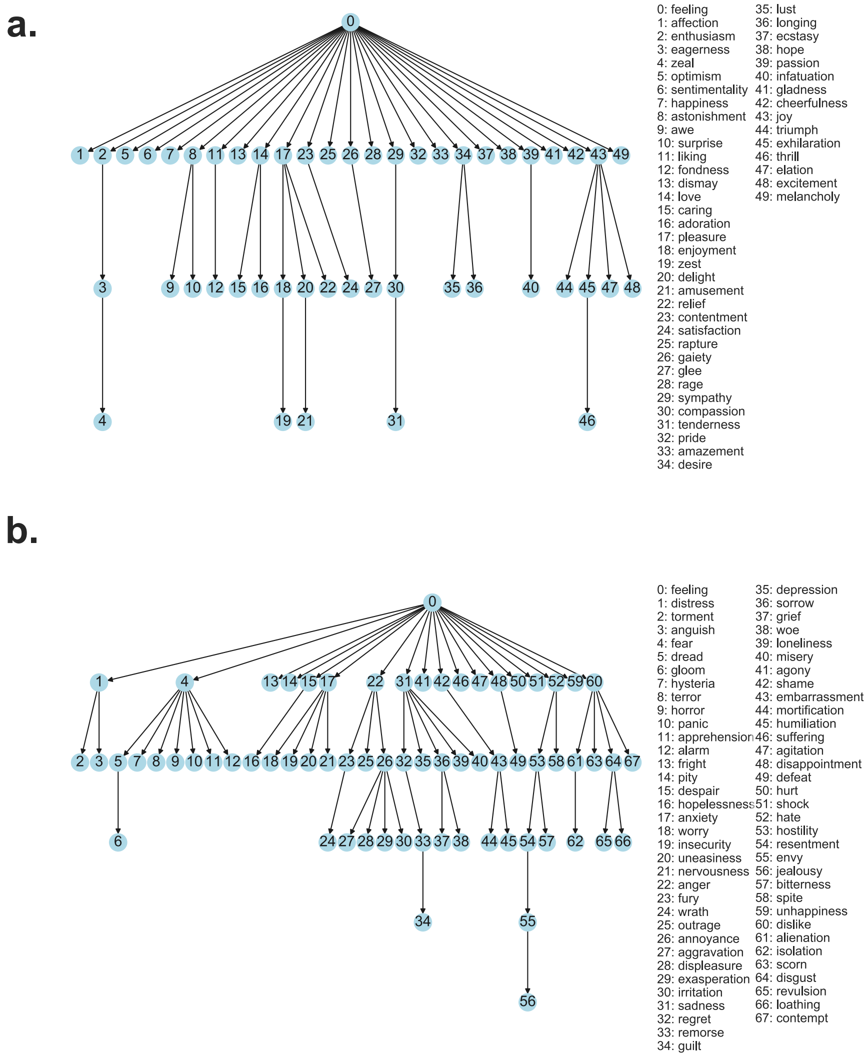

Figure S5: WordNet hierarchy of hypernyms and hyponyms for a) positive emotion words and b) negative emotion words. Valence is determined by our data described in *Section 3.1*.

coefficients for prototypicality ($\beta = -0.4704$, $p < 0.001$) and frequency ($\beta = -0.0460$, $p < 0.001$) are significant, but for valence ($\beta = 0.0049$, $p = 0.267$), number of senses ($\beta = 0.0029$, $p = 0.075$), depth ($\beta = -0.0057$, $p = 0.408$), and AoA ($\beta = 0.0036$, $p = 0.093$) it is insignificant. We observe that prototypicality still has a significant, negative effect as predicted by our hypothesis. We also observe that we can reproduce the finding by Hamilton et al. (2016) for frequency.

3.3 Human judgements of prototypicality

We repeated the analysis in the previous section by replacing estimated prototypicality with empirical prototypicality ratings. We computed the rates of change for every emotion concept x,

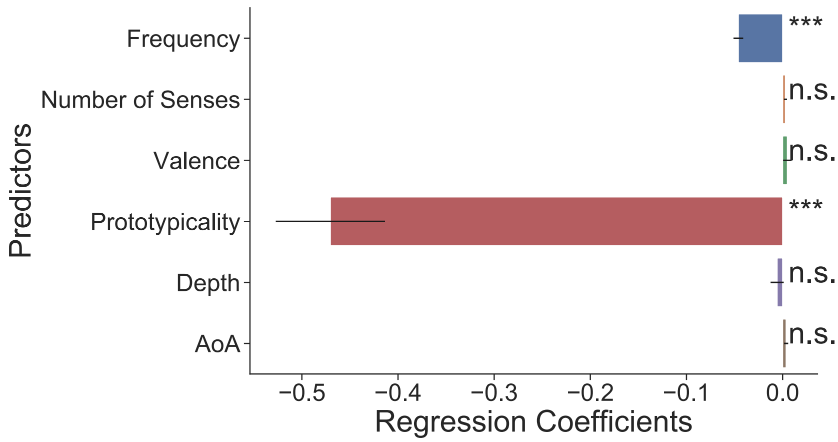

Figure S6: Predictor coefficients from multiple regressions on rates of semantic change. Error bars follow the same layout as Figure S4. Prototypicality is estimated using Equation 2 in the main text.

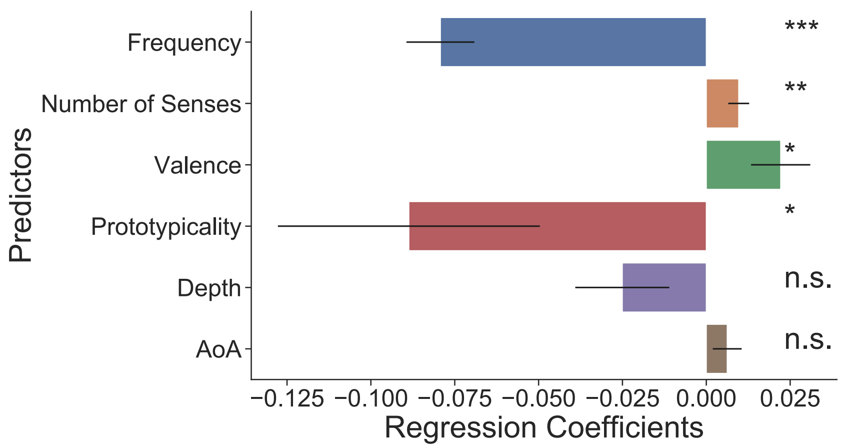

Figure S7: Predictor coefficients from multiple regressions on rates of semantic change. Error bars follow the same layout as Figure S4. Prototypicality is based on human ratings.

$rate(x, 1980, 1990)$. Then, we performed multiple regression using the following formula:

$$rate(x, 1980, 1990) \sim proto(x) + freq(x) + poly(x) + val(x) + depth(x) \tag{2}$$

where $proto(x)$ is the prototypicality rating of x obtained from Shaver, Schwartz, Kirson, and O'connor (1987).

Figure S7 shows the multiple regression results. The adjusted R^2 of the model is 0.524, with $p < 0.001$, $n = 109$; mean regression coefficients for prototypicality ($\beta = -0.0887$, $p = 0.025$), frequency ($\beta = -0.0793$, $p < 0.001$), number of senses ($\beta = 0.0097$, $p = 0.002$), and valence ($\beta = 0.0222$, $p = 0.014$) are significant, but for depth ($\beta = -0.0250$, $p = 0.077$) and AoA ($\beta = 0.0062$, $p = 0.150$) it is insignificant. We observe that empirical prototypicality also has a significant, negative effect, albeit the effect size is smaller than previously.

References

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*(1), 28–58.

Boster, J. S. (1988). Natural sources of internal category structure: Typicality, familiarity, and similarity of birds. *Memory & Cognition*, *16*(3), 258–270.

Brysbaert, M., & Biemiller, A. (2017). Test-based age-of-acquisition norms for 44 thousand english word meanings. *Behavior research methods*, *49*(4), 1520–1523.

Dale, E., & O'rourke, J. (1981). *The living word vocabulary, the words we know: A national vocabulary inventory.* Chicago: World Book.

Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, *113*(3), 464.

Fellbaum, C. (1998). *Wordnet: An electronic lexical database*. Cambridge, MA: MIT Press.

Garrard, P., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology*, *18*(2), 125–174.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.*

Johnson-Laird, P. N., & Oatley, K. (1989). The language of emotions: An analysis of a semantic field. *Cognition and Emotion*, *3*(2), 81–123.

Kay, C., Roberts, J., Samuels, M., & Wotherspoon, I. (2017). *The Historical Thesaurus of English, version 4.21.* Glasgow, UK: University of Glasgow. Retrieved from `http://historicalthesaurus.arts.gla.ac.uk/`

Loper, E., & Bird, S. (2002). Nltk: The natural language toolkit. In *In Proceedings of the ACL Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics. philadelphia: Association for computational linguistics.*

Luo, Y., & Xu, Y. (2018). Stability in the temporal dynamics of word meanings. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society.*

Monaghan, P. (2014). Age of acquisition predicts rate of lexical evolution. *Cognition*, *133*(3), 530–534.

Monaghan, P., & Roberts, S. G. (2019). Cognitive influences in language evolution: Psycholinguistic predictors of loan word borrowing. *Cognition*, *186*, 147–158.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805.

Shaver, P., Schwartz, J., Kirson, D., & O'connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52*(6), 1061.

Xu, A., Ramiro, C., & Xu, Y. (2019). A predictability-distinctiveness trade-off in the historical emergence of word forms. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society.*