Diachronic Language Change and Its Influence on Lexico-semantic Representations Across the Lifespan

Ellis Cain (ecain@ucmerced.edu)

Rachel Ryskin (rryskin@ucmerced.edu)
Cognitive and Information Science, University of California, Merced, USA

Abstract

Patterns of language use change over time and may reflect and/or impact lexico-semantic representations of individuals as they age. In the current study, we use distributional semantic word embeddings trained on corpora from different decades (HistWords) to examine language change. We first measured lexico-semantic organization in different age groups, using an open dataset of association norms, and tested how they may be related to language change. Then, using the diachronic word embeddings, we sampled English words that have changed in meaning and words that have maintained the same meaning/usage patterns between the 1950s and the 1990s. We tested how relatedness judgments for those words differ when paired with their "neighbors" from earlier vs. recent decades, for both younger and older adults. Our findings suggest that individuals continuously and rapidly update their lexico-semantic representations across the lifespan, such that earlier learned meanings have minimal impact on present-day representation.

Keywords: lexico-semantic representation; aging; language change

Introduction

Language is a complex adaptive system (Beckner et al., 2009). The behavior of speakers in a speech (or writing/signing) community is based on past interactions, social norms, and cognitive pressures. From their individual behaviors emerge the collective language usage patterns, which are recorded in corpora. Future individuals learn from past collective usage patterns, as well as from interactions with others in their community.

A consequence of this complex adaptive nature is that language changes over historical time and this can be observed in diachronic usage statistics (e.g., Bybee, 2015; Michel et al., 2011). Yet, little is known about the relationship between changes in collective usage patterns and the meaning representations of individual language users over time.

Lexico-semantic representations appear to change across the lifespan. Using word association data (De Deyne, Navarro, Perfors, Brysbaert, & Storms, 2019), Dubossarsky et al. (2017) analyzed lexical association networks and found that their organization differs by age. Namely, they observed a U-shaped developmental trajectory: networks are small and sparse (e.g., reduced average shortest path length and entropy) during language acquisition (10–18 y.o.), they are dense and well-connected in mid-life (e.g., increased in-degree and out-degree), and then they become sparse again (though larger overall), with a larger proportion of isolated, peripheral nodes, in late-life (e.g., reduced in/out degree, increased entropy).

In addition, previous studies have used electroencephalography (EEG) to explore the relationship between lexico-semantic processing and aging (e.g., Federmeier, Kutas, & Schul, 2010; Wlotko, Federmeier, & Kutas, 2012). Some signatures of predictive processing (e.g., increased anterior positivity in response to a word that is moderately vs. highly predictable from a preceding cue) are reduced in older adults, suggesting that aging may be related to changes in neural mechanisms of prediction and/or in the semantic predictability relationships among words (see also Castro, Curley, & Hertzog, 2021).

Aging is associated with myriad neural and cognitive changes which may contribute to these age-related changes in lexico-semantic representations (e.g., Payne & Silcox, 2019). Importantly, older adults also differ from younger adults in that they have experienced language over a longer stretch of time, during which usage patterns and meanings of some words may have shifted. Individual language users must update their lexico-semantic representations with experience (e.g., learning a new meaning for "tweet") but the form of this updating is unknown. Does this adaptation happen rapidly, such that an individual's representations primarily reflect the collective usage patterns of the current moment, or is it slower and more cumulative, such that an individual's representations reflect the usage patterns of the present as well as the patterns they experienced in the past?

Present research The goals of the present work are to probe differences in lexico-semantic organization across different age groups and test how well these are explained by diachronic language change at the collective level.

In the first experiment, we used representational similarity analysis (RSA; Kriegeskorte, 2008) to compare the lexico-semantic information (in the form of representational similarity matrices (RSM)) derived from diachronic word embeddings ("HistWord" embeddings trained on text corpora from different decades; Hamilton, Leskovec, & Jurafsky, 2018) with lexico-semantic information obtained from word association data from individuals across the lifespan (De Deyne et al., 2019). See Figure 1 for a schematic overview of the approach.

If lexico-semantic representations are rapidly updated

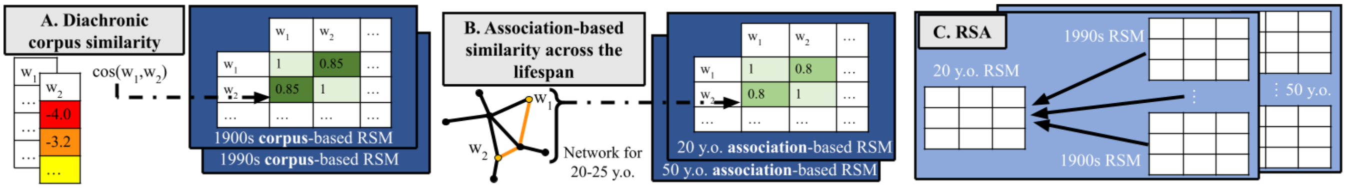

Figure 1: Schematic of data and representational similarity analyses (RSA) in Experiment 1. A) For each decade in the diachronic word embeddings, representational similarity matrices (RSM) are generated where each row/column represents a word, and the cell is the cosine similarity between those two word embeddings. This is repeated for each decade. B) Using the lexical association networks from Small World of Words, RSMs are generated where each cell is the similarity derived from a Katz random walk function over the network. This is repeated for each age cohort. C) RSA is used to compare each RSM from the association data with the RSMs from the diachronic word embeddings.

through experience, association-based RSMs should be best predicted by corpus-based RSMs from the most recent decades, regardless of age. Moreover, if lifelong language learning involves the cumulative integration of statistical information experienced over the lifespan, corpus-based RSMs from earlier decades should explain more variance in association-based RSMs for older adults relative to RSMs for young adults.

In the second experiment, we used HistWords embeddings to select words which underwent substantial meaning change between 1950 and 1990 and words which did not change in meaning in that same time period. We then gathered word relatedness judgements for each type of word paired with a near-neighbor from the 1950s, a near-neighbor from the 1990s, or a non-neighbor word (as a control).

As in the first experiment, for words that have changed in their usage, the more recent (1990s) neighbor pairs should be judged as more related than those from the earlier decade (1950s) for both ages. However, continuous cumulative updating based on the statistics of the language also predicts that assessments of relatedness for neighbor pairs may diverge between younger and older adults, in particular from earlier decades, which may be relatively more related for older adults who have experienced the usage patterns of those decades.

Experiment 1

We aimed to quantify the extent to which the age-related differences in lexico-semantic association networks (Dubossarsky et al., 2017) are explained by diachronic changes in meaning/usage patterns. In order to relate word association data and diachronic corpora, we used RSA, which uses 2nd-order isomorphic RSMs to abstract away from the original format of the data.

Methods

Data We used the English diachronic word embeddings from HistWords (Hamilton et al., 2018) to construct representational similarity matrices (RSM) for each decade from the 1900s to the 1990s, where each cell represents the cosine similarity between two word embeddings (vectors) (Fig. 1A). The embeddings in HistWords were generated using a Word2Vec model (Skip-gram negative sampling) trained on decade-level subsets of the Google books corpus. These RSMs are intended to capture the lexico-semantic information encoded in each decade's text corpora (Firth, 1957; Wittgenstein, 1953; Harris, 1954; Lenci, 2018).

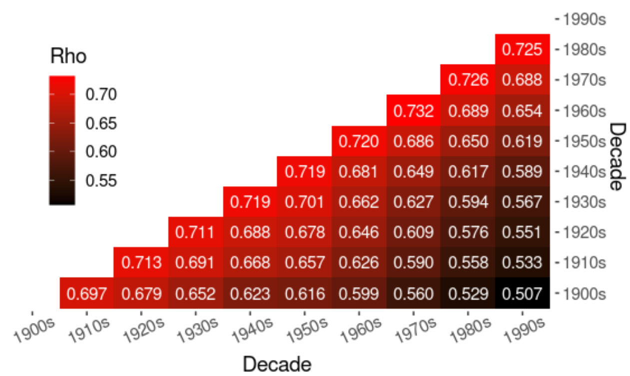

Figure 2: Heat-map of RSA between the corpus-based RSMs. Each cell represents Spearman's rho correlation coefficient. The diagonal was excluded since it would be a perfect correlation.

Similarly, we used English word association data from the Small World of Words (De Deyne et al., 2019) dataset to construct separate RSMs for each age cohort. Similarity is calculated using a spreading activation Katz random walk method, which compares the distributional overlap of direct and indirect paths between two words (Fig. 1B). Between 2011–2018, De Deyne and colleagues collected up to 3 responses for 12,292 cues (such as "couple", "plug", "condense"). They had 88,722 participants in total (mean age = 36 y.o., SD = 16 y.o., female = 38%). Age bins/cohorts for RSMs were selected to balance disparities in sample size with covering a large range of participant age (20–25: n=1891, 25–30: n=1539, 30–35: n=1322, 35–40: n=1267, 40–45: n=1285, 45–50: n=913, 50–70: n=891). These RSMs are intended to capture the lexico-semantic representations of each age cohort.

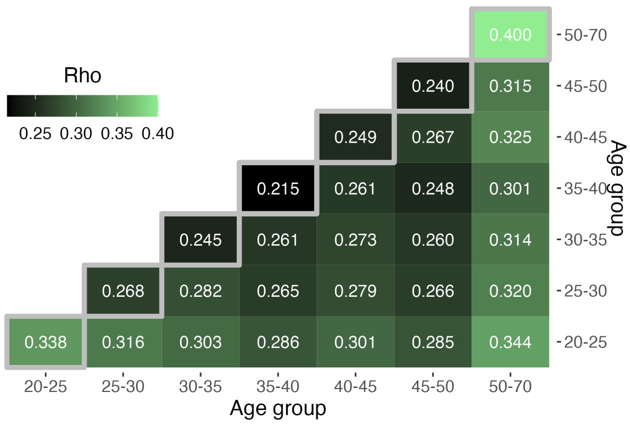

Figure 3: Heat-map of RSA between the association-based RSMs. Each cell represents Spearman's rho correlation coefficient between RSMs of two decades. The diagonal values represent internal reliability, which were adjusted using Spearman-brown prophecy formula.

Results

We first conducted RSA using Spearman-rank correlation to compare the semantic spaces from HistWords across decades (Fig. 2), from Small World of Words across age cohorts (Fig. 3), and between the two sources for all decade-age pairs (Fig. 1C; Fig. 4).

For the corpus-based semantic organization (Fig. 2), the RSMs were most similar for adjacent decades ($0.697 \leq \rho \leq 0.732$), and as the temporal distance between decades increase, the similarity gradually decreased, with the 1900s and 1990s being the least similar (ρ = 0.507). Importantly, lexico-semantic organization appears to change non-trivially within the lifespan of an older adult living in the 2000s: the correlation between similarities from the 1990s and 1980s is $\rho = 0.725$, whereas between the 1990s and 1940s it is $\rho = 0.589$. Still, the minimum correlation was $\rho \approx 0.5$, showing that while there may be broad changes in linguistic meaning over time, there is substantial stability as well.

The results of all pairwise correlations between association-based RSMs from different age groups can be seen in Fig. 3. To measure the internal reliability of the data, we calculated the average correlation between two RSMs generated from five random splits (in half) of an age cohort's association data, and then adjusted the ρ correlation coefficient using the Spearman-brown formula (Eisinga, Grotenhuis, & Pelzer, 2013; Kelley, 1925), in order to correct for the fact that splitting the sample leads to an underestimate of the reliability. The association-based RSMs were less correlated than the HistWords-based RSMs overall ($\rho \approx 0.3$) and had relatively low internal reliability ($\rho \approx 0.2 - 0.4$ after Spearman-brown prophecy correction). Notably, the 50-70 age group had the highest internal reliability, potentially due to the fact that it included data from a wider range of ages (in order to make a bin with similar sample size to the others). The second highest internal reliability was found in the 20-25 age group data, which had the largest sample size. No gradient change in similarity related to temporal distance in age cohorts was apparent.

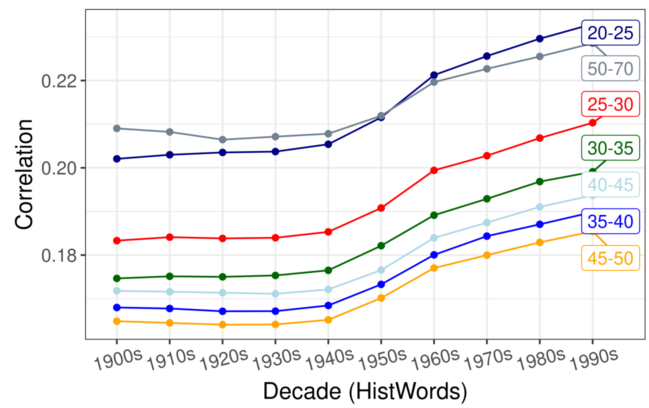

Figure 4: Spearman's ρ correlations between the association-based and corpus-based RSMs over time. Each line connects the correlation values for a specific age cohort.

Figure 4 shows the results of correlations between RSMs for each decade and RSMs for each age cohort. Across all of the age cohorts, the overarching trend was the same: increased correlation for more recent decades, with a particular inflection after the 1940s. Comparing the cohorts, the youngest (20–25 y.o.) and oldest (50–70 y.o.) RSMs had the highest correlation to the 1990s RSM (and all other decade RSMs), while the 35–40 y.o. and 45–50 y.o. RSMs had the lowest correlation to the 1990s RSM (and all other decade RSMs). However, this may be primarily related to measurement issues. Unsurprisingly, given that the internal reliability of a measure places an upper bound on any potential correlations with other measures, the maximum correlation for a given age group fairly closely tracked the age cohort's internal reliability (see Fig. 3).

Interestingly, though overall correlations for the 50–70 y.o. cohort and the 20-25 y.o. cohort are similar, the slope of the increase across decades was shallower for the 50-70 y.o. cohort than the 20-25 y.o. cohort; the 50-70 y.o. cohort had the higher correlations for earlier decades ($\rho \approx 0.21$ at 1900s), but this switched after the 1950s to the 20-25 y.o. group having the higher correlations ($\rho \approx 0.23$ at 1990s).

To further quantify the relative predictive ability of each decade's RSM for each age cohort's RSM, we used a linear model to predict each age cohorts' RSM as a combination of the decade-level RSMs and ablated the individual decade to quantify its impact via model comparison.

The equation for the full (non-ablated) model was as follows:

$$Sim.Assoc = \beta_0 + (\beta_1 * Sim.1900) + \cdots + (\beta_{10} * Sim.1990) + \varepsilon$$

where Sim.Assoc refers to a similarity value from a given age cohorts' RSM, and Sim.*Year* refers to the similarity from

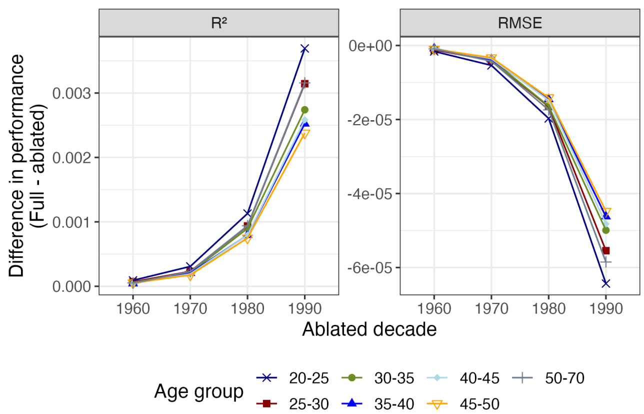

Figure 5: Impact of ablation on model performance for each age cohort, from 1960-1990s. Ablating decades from 1950s and earlier had no impact on performance. Each dot represents the difference between full and ablated model performance.

a given decade's RSM.

Figure 5 shows the average difference in root-mean-square error (RMSE), and R^2 between the full and ablated models across 5 cross-validation folds. The most recent decades have the largest impact on the models' ability to predict the association-based similarity for all age groups. Further, the impact of the most recent decades appears largest for the youngest age cohorts and less for the older age cohorts. One exception to this pattern is the 50–70 cohort, which includes a larger age range than the other cohorts due to smaller sample size among the oldest age participants and has the highest internal reliability. F-tests comparing the model fit of the full and ablated models showed that the full model always predicts the association-based similarity values significantly better than the ablated models ($ps < 0.001$).

In sum, we see that the corpus-based semantic similarity spaces changed gradually over time, while the association-based similarity was largely consistent across age cohorts. All age cohorts' semantic organizations were most similar to the 1990s corpus-based semantic organization. With the exception of the 50–70 y.o. cohort, similarity to the corpus-based organization generally decreases with age, with the 20–25 y.o. having the highest correlation, and the 40–50 y.o. having the lowest.

Experiment 2

The dataset of word associations used in Experiment 1 yielded valuable insights but was limited in multiple ways: the behavioral responses were based on associations as opposed to relatedness (which is what the similarities of HistWords embeddings are thought to represent), the sample sizes across age bins were uneven, and the internal reliability metrics of the RSMs were modest. Moreover, the cue words for the Small World of Words dataset were selected to maximize the number of associates produced by participants of all ages. These then may not be optimal for studying the interaction between language change and age-related differences in lexico-semantic representations.

Therefore, in Experiment 2, we collected explicit word relatedness judgments (following Hill, Reichart, & Korhonen, 2015; Gerz, Vulić, Hill, Reichart, & Korhonen, 2016), from both young (22–33 y.o.) and older adults (63–81 y.o.), comparing words that changed in their usage/meaning over the approximate lifetime of the older adults (1950 to 2000) to those that didn't.

Methods

Participants 1000 English-speaking participants were recruited through the online crowd-sourcing platforms Amazon Mechanical Turk and Prolific. All participants correctly answered 5 catch trials and completed the task in a reasonable time-frame ($>$ 4 min). Participants were split into two groups based on their age: younger adults (YA $\leq$ 33 y.o., born in $\sim$1990), older adults (OA $\geq$ 63 y.o.; born in $\sim$1960). The YA group included 500 participants ($M = 27.7$ y.o., $SD = 4.04$, Female = 46%), and the OA group included 500 participants ($M = 68.04$ y.o., $SD = 4.33$, Female = 53.6%). Participants took on average 13.4 minutes to complete the task.

Stimuli The stimuli for this experiment were pairs of words selected using the HistWords diachronic word embeddings. The target words were drawn from the HistWords vocabulary and filtered by frequency based on SUBTLEX (Brysbaert & New, 2009), such that they included a wide range of usage frequencies (range 10 – 27616). The most recent decade available in the HistWords embeddings is 1990–2000s. We used the decade-level word embeddings to identify 150 target words that had undergone the most change in meaning (i.e., cosine similarity between 1950s and 1990s was < 0.35) and 150 target words that had not undergone meaning change between the 1950s and 1990s (300 words total). (Since the HistWords embeddings were aligned using an orthogonal Procrustes, we can directly compare the word embeddings across decades to quantify the extent of change.) We chose the 1950s since this decade seemed to be a key transition point in the results of Exp. 1 (e.g., the pre-1950s RSMs were most correlated with the 50–70 year old group's RSMs, but the post-1950s RSMs were most correlated with the 20–25 year old group's RSMs). Examples of changed words would be "icon" or "broadcast", while an unchanged word would be "daughter".

For each of the 300 target words, we selected 10 nearest neighbors from the 1950s and 10 neighbors from the 1990s (20 neighbors per word, 6000 word-neighbor pairs). To serve as an unrelated baseline, 10 unrelated words were randomly selected for each target word from the inverse set of its neighbors (3000 non-neighbor pairs), bringing the total stimulus set to 9000 word pairs.

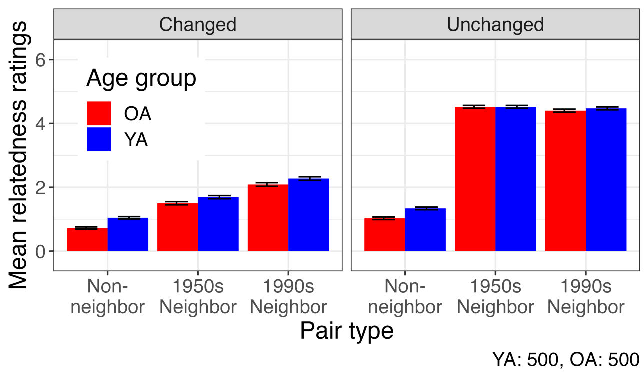

Figure 6: Mean relatedness ratings, from Experiment 2, by target word type (changed vs. unchanged in meaning from 1950 to 1990) and pair type. Error bars represent 95% confidence intervals over participant mean.

Design Participants were presented with 75 word pairs and asked to rate their relatedness on a 7-point Likert scale, where 0 was "unrelated" and 6 was "very closely related". Each participant saw 12 (or 13 depending on the counterbalancing list) target words that had changed in meaning/usage, according to the HistWords embeddings, and 13 (or 12) that hadn't changed. Each target word was rated 3 times, paired with: 1) one of its nearest neighbors based on the 1950s embeddings, 2) one of its nearest neighbors based on the 1990s embeddings, and 3) an unrelated, non-neighbor. Participants were randomly assigned to one of 120 counterbalancing lists and the order of pairs was randomly generated for each participant.

To ensure that participants were paying attention, there were five catch trials throughout the experiment, where participants were asked to identify the pair with a non-word from a set of five pairs. Participants who did not answer all five of the catch trials correctly were excluded from the analyses.

Results

Due to the random assignment to counterbalancing lists, each target word and its pairs were rated by different numbers of participants. On average, each pair was rated by 9.51 participants (M_{YA} = 4.76; M_{OA} = 4.75 participants).

Figure 6 shows the average relatedness ratings for the three types of pairs based on whether or not the target word changed in meaning, for both age groups. We used a Bayesian multilevel ordinal model[1] to test the relationship between age group, pair types (1990s, 1950s, or non-neighbor), meaning change, their interactions, and the average relatedness ratings for each target-neighbor pair. The YA group, unchanged target words, and non-neighbor pairs were coded as the reference levels for the corresponding variables.

For the unchanged target words, the 1950s and 1990s neighbor pairs were both rated as highly related, relative to the non-neighbor pairs ($Est._{1950}$ = 4.76, 95% CI = [4.55, 4.97]; $Est._{1990}$ = 4.69, 95% CI = [4.48, 4.90]) Both of these pair types were consistently rated as more related than the non-neighbor pairs ($Est._{Non-neighbor}$ = -0.40, 95% CI = [-0.50, -0.30]). For the changed target words, the non-neighbor pairs also had lower rating across age groups ($Est._{Non-neighbor}$ = -1.06, 95% CI = [-1.26, -0.87]) relative to the neighbor pairs ($Est._{1950}$ = 0.35 SD, 95% CI = [0.14, 0.53]; $Est._{1990}$ = 1.21 SD, 95% CI = [1.01, 1.41]). Both neighbor pair types were rated as less related for the changed target words, relative to unchanged ($Est._{Changed} = -0.57$ SD, 95% CI = [-0.74,-0.40]).

Overall, OA rated pairs as less related than YA ($\beta_{OA} = -0.79$ SD, 95% CI = [-1.00,-0.60]). There was an interaction between age group and meaning change, such that the difference between the ratings of YA and OA is smaller for the unchanged than the changed pairs ($\beta_{OA:Changed} = -0.19$ SD, 95% CI = [-0.33,-0.05]). Age group interacted with pair type, such that both 1950s and 1990s neighbor pairs were rated as more related (relative to non-neighbors) by OA than YA ($\beta_{OA:1950}$ = 0.85, 95% CI = [0.63, 1.06]; $\beta_{OA:1990}$ = 0.75, 95% CI = [0.54, 0.96]), and the difference between age groups in relatedness ratings between non-neighbors and neighbors was smaller for words that changed meaning than for the unchanged words($\beta_{Changed:1950} = -3.58$, 95% CI = [-3.83, -3.34]; $\beta_{Changed:1990} = -2.73$, 95% CI = [-2.97, -2.48]).

There was a three-way interaction between age, meaning change, and pair type, such that the rating difference between non-neighbors and 1950s neighbors was larger for OA than YA with unchanged words, but this difference switches for changed target words, with the difference being larger for YA than OA ($\beta_{OA:Changed:1950} = -0.33$, 95% CI = [-0.55, -0.12] The CI for the 1990s neighbor includes zero, suggesting a lack of three-way interaction for the 1990s neighbors ($\beta_{OA:Changed:1990} = -0.19$, 95% CI = [-0.39, 0.01]).

However, crucially, the difference in relatedness ratings between 1950s and 1990s neighbor pairs was similar across the two age groups (β_{OA} = -0.89, 95% CI = [-1.06, -0.70]; β_{YA} = -0.84, 95% CI = [-1.00, -0.69]).

Similar to Experiment 1, we used a model ablation approach to quantify the relative predictive ability of each decade's similarity values for the two age cohort's relatedness ratings. We focused specifically on the 1950s and 1990s neighbor pairs for changed words.

We trained four (2 age groups, 2 neighbor types) mixed-effect linear models to predict participant relatedness ratings as a combination of decade-level similarity from HistWords[2], and ablated individual decades to quantify their impact via model comparison.

Figure 7 shows the difference in root-mean-square error (RMSE) and R^2 between the full and ablated models. χ^2 tests indicate that the full model predicts the participant ratings significantly better than any of the ablated models ($p < 0.001$). For the 1950s neighbor pairs, the most distant (1940s and

[1] brm(ratings $\sim$ 1 + age group * pair type * changed + (1 + changed * pair type | subject) + (1 + age group | pair), family = cumulative("logit"))

[2] lmer(ratings $\sim sim_{1940} + sim_{1950}$ + ... + sim_{1990} + (1 | Subject)

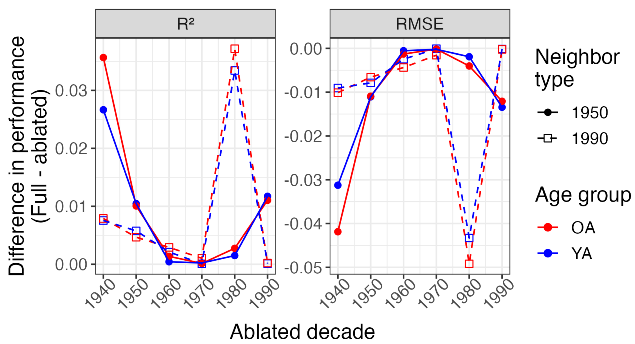

Figure 7: Ablation impact on model performance for each age cohort, by neighbor pair type (1950s or 1990s from changed only). Each dot represents the difference between full and ablated model performance.

1950s) and most recent decades (1980s and 1990s) had the greatest impact. For the 1990s neighbor pairs, the most distant (1940s and 1950s) and the 1980s had the greatest impact.

Indeed, the 1940s and the 1980s had the largest impact for the 1950s and 1990s neighbors respectively. This may be related to our stimulus selection process which searched for words which maximized the cosine similarity between the 1990s and 1950s decade embeddings, though the exact reason for this pattern is unclear. The crucial finding is that ablation impact patterns across decades were quasi-identical across the two age groups.

In sum, patterns of relatedness ratings across word and pair types were similar for young and older adults. Namely, for unchanged target words, the 1950s and 1990s neighbor pairs had very similar ratings, while for the target words that changed meaning, the 1990s neighbor pairs were rated as the most related, followed by the 1950s neighbor pairs. Additionally, with the exception of the unchanged neighbor pairs, the YA rated pairs as more related than the OA did, on average. Crucially, their ratings of 1990s neighbors relative to 1950s neighbors did not differ by age.

Discussion

In the first study, we found that the lexico-semantic representations derived from diachronic corpora changed from one decade to the next, such that the collective language use patterns experienced by an individual at different points in their life (e.g., 50 years apart) are measurably different. Yet, for individuals of all ages, their lexico-semantic representations (based on word association behavior) were most closely related to the usage patterns of the most recent decade.

Some patterns were suggestive of age-related differences in these diachronic effects: the correlations to corpus-based semantic organization generally decreased with age (with the exception of the 50–70 y.o. group). One interpretation is that, as a result of experiencing a larger variety of meanings/usage patterns, the lexico-semantic representations of older adults may become less closely tied to the usage patterns of any particular decade. However, caution is warranted in interpreting these results given the low internal reliability of the association-based similarity measures.

In the second study, we demonstrated that words which changed in meaning between the 1950s and 1990s, according to diachronic embeddings, were judged to be more related in meaning to neighbors from the 1990s than the 1950s. (This was not the case for words which didn't change in meaning: neighbors from the 1950s and 1990s were judged to be equally semantically related.) Crucially, both younger and older adults judged the words that changed in meaning to be closer to their 1990s neighbors than their 1950s neighbors to the same extent. As in Experiment 1, the lexico-semantic representations of both age groups appeared to be more closely aligned with the usage patterns of more recent decades. The older adults overall provided lower ratings than the young adults. This pattern is potentially analogous to what was observed in Experiment 1: as a result of experiencing a larger variety of meanings/usage patterns, the lexico-semantic representations of older adults may become less closely tied to the usage patterns of any particular decade. However, the influence of earlier language use patterns — which older adults, but not younger adults, would have experienced in their lifetime — appeared to be similar for both age groups.

One possibility is that differences by age group may have been masked by noise. The stimuli for Experiment 2 were selected on the basis of meaning change and more frequent words are less likely to change in meaning (Hamilton et al., 2018), such that many of the target words were lower frequency ($M_{Freq} = 310$ in SUBTLEX). Participants may not have been familiar with all these words, leading to increased rates of guessing.

Another possibility is that the similarity in lexico-semantic representations across ages may reflect the fact that, within a typical individual's lifetime, there is substantial stability in meanings (see Fig. 2; $\rho_{1950-1990} = 0.62$ vs $\rho_{1980-1990} = 0.73$). Many words do not change their meanings within that time-frame. The change which does occur may be too subtle to detect with a meta-linguistic relatedness judgment task.

Taken at face value, the current results are consistent with findings that adults continue to update their language representations well into adulthood (Brysbaert, Warriner, & Kuperman, 2014; Ryskin, Qi, Duff, & Brown-Schmidt, 2017). More specifically, the minimal age-related differences suggest that individuals rapidly adapt to changes in linguistic meaning and continually track the collective usage patterns, rather than being biased by the prior usage patterns they experienced. Future work will explore whether traces of past language experience may in fact be revealed through the use of implicit neural measures (e.g., EEG).

References

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., ... others (2009). Language is a complex adaptive system: Position paper. *Language learning*, *59*, 1–26.

Brysbaert, M., & New, B. (2009). Moving beyond kučera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavior research methods*, *41*(4), 977–990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior research methods*, *46*(3), 904–911.

Bybee, J. (2015). *Language change*. Cambridge University Press.

Castro, N., Curley, T., & Hertzog, C. (2021). Category norms with a cross-sectional sample of adults in the united states: Consideration of cohort, age, and historical effects on semantic categories. *Behavior research methods*, *53*, 898–917.

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019, June). The "Small World of Words" English word association norms for over 12,000 cue words. *Behavior Research Methods*, *51*(3), 987–1006. doi: 10.3758/s13428-018-1115-7

Dubossarsky, H., De Deyne, S., & Hills, T. T. (2017, August). Quantifying the structure of free association networks across the life span. *Developmental Psychology*, *53*(8), 1560–1570. doi: 10.1037/dev0000347

Eisinga, R., Grotenhuis, M. t., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, cronbach, or spearman-brown? *International journal of public health*, *58*(4), 637–642.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and language*, *115*(3), 149–161.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*.

Gerz, D., Vulić, I., Hill, F., Reichart, R., & Korhonen, A. (2016). Simverb-3500: A large-scale evaluation set of verb similarity. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 2173–2182).

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2018, October). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. *arXiv:1605.09096 [cs]*.

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2-3), 146–162.

Hill, F., Reichart, R., & Korhonen, A. (2015). Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics*, *41*(4), 665–695.

Kelley, T. L. (1925). The applicability of the spearman-brown formula for the measurement of reliability. *Journal of Educational Psychology*, *16*(5), 300.

Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*. doi: 10.3389/neuro.06.004.2008

Lenci, A. (2018, January). Distributional Models of Word Meaning. *Annual Review of Linguistics*, *4*(1), 151–171. doi: 10.1146/annurev-linguistics-030514-125254

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, … Aiden, E. L. (2011, January). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, *331*(6014), 176–182. doi: 10.1126/science.1199644

Payne, B. R., & Silcox, J. W. (2019). Aging, context processing, and comprehension. In *Psychology of learning and motivation* (Vol. 71, pp. 215–264). Elsevier.

Ryskin, R. A., Qi, Z., Duff, M. C., & Brown-Schmidt, S. (2017). Verb biases are shaped through lifelong learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(5), 781.

Wittgenstein, L. (1953). *Philosophical investigations. philosophische untersuchungen.* Macmillan.

Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: Age-related differences in the use of sentential context. *Psychology and Aging*, *27*(4), 975–988. (Place: US Publisher: American Psychological Association) doi: 10.1037/a0029206