Cite as: Dubossarsky, H., De Deyne, S., & Hills, T. T. (In press). Quantifying the structure of free association networks across the lifespan. *Developmental Psychology*.

Running Head: ASSOCIATION NETWORKS ACROSS THE LIFESPAN

Quantifying the Structure of Free Association Networks Across the Lifespan

Haim Dubossarsky[1], Simon De Deyne[2], and Thomas T. Hills[3]

[1]Edmond and Lily Safra Center for Brain Sciences, Hebrew University of Jerusalem
[2]School of Psychology, University of Adelaide
[3]Department of Psychology, University of Warwick

Address correspondence to:
Thomas Hills
University of Warwick
Department of Psychology
Gibbet Hill Road
CV4 7AL, Coventry UK
Tel: +44 (0) 2765 75527
Fax: +44 (0) 2765 75527
E-mail: t.t.hills@warwick.ac.uk

Acknowledgments

This work was funded by a Mid-Career Fellowship from the British Academy MD130030 to TH, and an Australian Research Foundation DECRA Grant nr 140101749 to SDD

Abstract

We investigate how the mental lexicon changes over the lifespan using free association data from over 8,000 individuals, ranging from 10 to 84 years of age, with more than 400 cue words per age group. Using network analysis, with words as nodes and edges defined by the strength of shared associations, we find that associative networks evolve in a nonlinear (U-shaped) fashion over the lifespan. During early life, the network converges and becomes increasingly structured, with reductions in average path length, entropy, clustering coefficient, and small world index. Into late life, the pattern reverses but shows clear differences from early life. The pattern is independent of the increasing number of word types produced per cue across the lifespan, consistent with a network encoding an increasing number of relations between words as individuals age. Lifetime variability is dominantly driven by associative change in the least well-connected words.

Keywords: free associations, aging lexicon, language development, small worlds, network analysis, aging

Across the lifespan humans are exposed to an ever-changing stream of language and associative information. Hart & Risley (1995) propose that by the time children are four years old they have heard between 10 and 50 million words, which further increases over the lifetime as individuals learn to read and engage in more fluent conversation. This extensive exposure to language allows us to infer meaning from word co-occurrence, which is the basis for many investigations into our use and understanding of words (Hills, 2013; Landauer & Dumais, 1997; Stefanowitsch & Gries, 2003).

One implication of increasing language exposure is the potential for a change in lexical relations (i.e., associations) across the lifespan. This potential change has been used to explain lifelong developmental changes, including language learning (Hills, Maouene, Maouene, Sheya, & Smith, 2009b) and age-related memory decline (Borge-Holthoefer & Arenas, 2010; Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014). However, thus far, no study has recorded associative change across the lifespan from early to late life.

There are many open questions relevant to associative change over the lifespan. How stable are associative representations? In what ways do associative relationships change? And which words are likely to change most? These questions have been difficult to answer in the past because data rich enough to detect changes in associations across the lifespan has not been available. Many studies, including those mentioned above, have used the well-established University of South Florida free association norms—largely collected among university students (Nelson, McEvoy, & Schreiber, 2004). These have had a huge impact on psychological and cognitive science. However, at least in practice, these require the implicit assumption of a static lifelong representation—a one-size-fits-all-ages account of lexical representation.

In the present study, we investigate lifelong changes in the mental lexicon using a large-scale cross-sectional study with word associations collected from over 8000 individuals between 10 and 84 years old. Before we describe our approach to investigating this data, we first briefly review the literature on lifelong associative change in the mental lexicon.

Associative change across the lifespan

One of the more stable findings associated with aging is that vocabulary increases across the lifespan, well into old age (Light, 1992). Recent evidence involving over 400,000 Dutch participants in a lexical decision task shows that, between the ages of 12 and 80 years, vocabulary increases from 26,000 words to almost 42,000 words (Brysbaert, Warriner, & Kuperman, 2014). There is growing evidence that the relationships between words also changes over the lifespan, beginning as early as the second year of life. Children as young as 18-months exhibit associative priming effects (Arias-Trejo & Plunkett, 2013). Other studies have found changes in the consensus among response types across individuals (i.e., the number of unique associations elicited for a cue); comparing the associations between primary and secondary school, for example, shows an increase in the frequency of the most popular responses with age (Palermo, 1964; Shapiro, 1964). This is followed by stability in midlife and—based on fairly small samples relative to the present study—mixed evidence for possible increases in stability or heterogeneity in late life (Dörken, 1956; Hirsh & Tree, 2001; Lovelace & Cooley, 1982; Riegel & Birren, 1965; Tresselt & Mayzner, 1964). Up to early adulthood, the reduction in response types is consistent with overlearning associations associated with natural mastery of language skills (Anglin, Miller, & Wakefield, 1993; Maratsos, 2005). What is happening in late life remains to be seen. The results suggest two possible alternatives: a plateauing of associative development reached in midlife or, alternatively, an inverted U-shaped pattern reflecting a reversal in language coherence into late life.

One way to investigate this change involves network analysis. Network analysis provides a flexible means for investigating the large-scale structure of the lexicon, representing words as nodes and relations between words as edges (e.g., Baronchelli, Ferrer-i-Cancho, Pastor-Satorras, Chater, & Christiansen, 2013; De Deyne & Storms, 2008a; Vitevitch, Chan, & Goldstein, 2014). One of the advantages of network analysis is that it allows researchers to study phenomena at scales that range from individual words (i.e., nodes) to the entire lexicon (i.e., the network). Network analyses have been widely used to study the large scale structures of associative networks (De Deyne, Navarro, & Storms, 2013; De Deyne & Storms, 2008a; Steyvers & Tenenbaum, 2005), the formation of categories in toddlers (Hills, Maouene, Maouene, Sheya, & Smith, 2009a), and the trajectories of network development during early life (Bilson, Yoshida, Tran, Woods, & Hills, 2015; Hills et al., 2009b).

Network analysis has also provided evidence for changes across the lifespan. A study by Zortea, Menegola, Villavicencio, & Salles (2014) compared the associative networks of teenage children (8-12), adults (17-45 years) and older adults (60-87 years). In this study, 57 individuals from each age group generated three associative responses per person to each word in a list of 87 cue words. This work connected nodes if more than one participant responded to a cue word, generating unweighted and undirected networks. Results showed an increase in network degree over the lifespan along with an increase in clustering coefficient. The present study extends this previous work by collecting data from more than 8,000 individuals with more than 400 cues, and over a much more fine-grained age progression. We also use weighted and directed network analyses where possible, allowing us to provide a detailed picture of associative network development across the lifespan. Our analyses also include a detailed investigation of the global structural changes as well as changes associated with individual words.

The Current Study

Our aim here is to characterize the word and structure level changes in the associative lexicon in sufficient detail to help understand the directionality and timescale of associative change. In particular, we are interested in understanding to what extent there may be distinct stages in associative development and how the connectivity of words change over the course of development. We also examine the shape of this change across the lifespan. Among other things, this should directly inform our understanding of age-related differences in language related tasks (e.g., Hills, Mata, Wilke, & Samanez-Larkin, 2013; Ramscar et al., 2014).

Below we describe a basic word association task in which a cross-sectional sample of participants respond to a short list of word cues with the first words that come to mind. Next, we outline (a) a traditional word level approach to study how word meaning consensus changes over the lifespan and (b) a network level analysis which aims to explore structural changes in terms of global connectivity.

Methods

Participants

We collected data from 9 age groups, ranging from the 4th year of primary school (10 years old) to persons older than 68 years. For each group a total of 16,800 responses were collected to 420 words. Table 1 provides basic statistics for our participants and their responses after excluding unknown or missing responses (hence the number in R1, R2 and R3 are somewhat lower than 16,800). A number of criteria were used to decide on the inclusion of a participant. First we only considered individuals who indicated that they knew at least 30% of the cues and provided at least 25% of secondary and tertiary responses (each participant was asked to provide three responses to each cue word). These criteria were chosen to avoid excluding too many younger participants who might not know all the cues from a particular list or were not sufficiently motivated to complete the entire list of cues. Next, as some of the developmental data was collected as part of an ongoing study, we tried to match the number of males and females for each cue in each age group and obtain the same number of responses for each cue word (40 primary, secondary and tertiary responses) by randomly selecting a subset of participants meeting these criteria. For each age group, there were approximately an equal number of males as there were females, except for the 18-year-old participants who had on average 60% of responses from females. Education level was available for a subset of participants. When this information was available, it was used to exclude adult (age groups 30 to +68) participants who did not finish high-school (see Appendix for education based analysis).

Table 1

Age Group	Average Age	#Participants	R1	R2	R3	Total responses	Unique responses	Unique/total responses ratio
9-10	9.2	490	14453	12393	9598	36444	6441	0.177
11-12	10.5	466	15227	13728	11364	40319	6904	0.171
13-14	13.5	502	15982	14600	12043	42625	7970	0.187
17-19	18.3	1081	16709	16364	15557	48630	8663	0.178
28-32	31.0	1136	16769	16623	16221	49613	8947	0.180
38-42	41.0	1152	16759	16624	16243	49626	9501	0.191
48-52	51.0	1223	16789	16645	16254	49688	10280	0.207
58-62	61.0	1279	16777	16665	16364	49806	11144	0.224
+68	71.9	1222	16787	16595	16126	49508	12538	0.253

Note: Details for the different age-groups showing their average age, the number of participants, numbers of primary (R1), secondary (R2) and tertiary (R3) responses, number of total responses as well as the number of unique responses used (vocabulary size), and their ratio (type/token ratio).

Each response was converted to lowercase letters and checked against a list of common spelling mistakes. Different word forms were then normalized to match the word forms in the cue list (e.g., the response *apples* becomes *apple*). Non-alphanumeric characters and particles were removed ("a dog" becomes "dog") and responses like "is red" were changed to "red". For responses that contained extra information between brackets such as "play (theater)" only the part outside the brackets was retained. In a small number of cases, more than three responses were given. Only the first three responses were entered into the data. The total number of responses that were retained after these steps and the number of unique responses are shown in Table 1 for each of the age groups.

The age resolution for the younger participants reflects anticipated changes in childhood and early adulthood, while the remainder of the participants' ages was separated by an average 10 years. The participants were drawn from a large and ongoing online study currently involving over 120,000 persons described in De Deyne et al. (2013). In this study, participants volunteered for a short online word association task. For the current study, participants that matched our criteria for age were detected and presented with different stimuli. This procedure was feasible for all but the youngest participants. In order to gain responses from younger age groups as well, additional participants were recruited from the 4th, 5th, 6th and 8th grades by recruiting them at primary and high schools in Flanders. Though this represents a difference in recruitment, it was felt that this recruitment strategy would produce a better random sample than imposing a similar recruitment strategy across ages.

Stimuli

The stimuli were 420 Dutch words selected from the list of cues present in the adult word association norms (8,974 words) that were completed at the start of the study in 2009 and reported in De Deyne et al. (2013). The list of cues was designed to be a representative sample over common words, composed of 216 nouns, 102 adjectives and 102 verbs. These were randomly selected from words with known age-of-acquisition and imageability ratings (see De Deyne & Storms, 2008a) in order to compile a list of cues containing concrete and abstract words acquired at different ages.

Procedure

We used the continued word association task from De Deyne et al. (2013): Each participant was asked to provide three associations for each of a short list of cues. To reduce possible chaining, the instructions stressed the fact that responses should only be given to the cue word and should not be based on the previous responses. The number of cues varied from 15 to 42. This was adapted to the place of recruitment, with participants who performed the task at home responding to fewer cues and those performing the task at school doing more. This difference is made apparent in column three of Table 1. From 17 onwards all participants performed the task online. Since not all of the youngest participants were able to use a computer keyboard, a pen and paper procedure was used in the primary school participants and at certain high schools where computer facilities were unavailable. The instructions were similar in all age-groups except that one or two examples were provided by the experimenter for the classrooms with the youngest children. The final dataset consisted of 120 responses per cue (40 for each of the primary, secondary, and tertiary responses) from each age group. Words that elicited less than 25% of the required responses for any age group were not further analyzed. This procedure removed 16 words from the set of cues.

Word level analysis

The goal of the word level analysis was to investigate how the heterogeneity in responses changes as a function of age. This allows us to investigate possible changes in consensus across individuals by comparing response entropy of each cue for each age group. This also allows us to consider the similarity between consecutive age groups by comparing similarities in their distributions over targets using cosine distance. These measures are described below.

Entropy. Response counts were aggregated for each cue word and each age group separately, and then transformed to probabilities. This created nine response probability vectors per cue word, one for each age group. In order to evaluate the diversity in responses, taking into account both the number of different responses and their probabilities, the normalized or metric entropy[1] of each cue's response probability vector was computed as follows:

$$H = \sum_{i=1}^{n} \frac{p(x_i)\log(p(x_i))}{\log(n)} \qquad (1)$$

Where $p(x_i)$ represents the proportion of response type x_i and n represents the total number of response types that were produced as associations for that cue. This results in values bounded between 0 and 1. This is obtained by normalizing the entropy with the information length $log(n)$ in equation 1. Entropy is low for words where participants provide the same associates and high for words where participants provide more diverse response associations. An example is shown in Figure 1. A word with low entropy, like *lemon* (top panel), has most of the probability mass concentrated in just a few responses. A word with high entropy, like *bank* (bottom panel), has its probability mass more equally distributed across a variety of words.

[1]Normalized entropy and metric entropy are used interchangeably in the literature to denote the same thing.

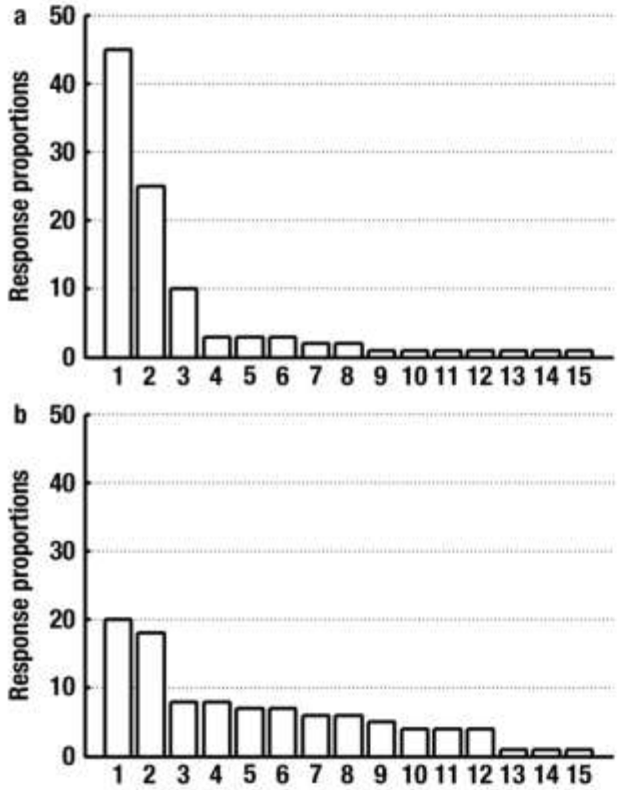

Figure 1. Illustration of two different response profiles for two cue words (top: *lemon*, bottom: *bank*). Responses (x axis) are sorted according to their response proportions (y axis).

Cross year associative change. In order to evaluate the cross-year change in a word's associations, we computed the cosine distance between consecutive age groups' response probability vectors as follows:

$$W = log\left(1 - \frac{\sum_{i=1}^{n} x_{i,t} * x_{i,t+1}}{\sqrt{\sum_{i=1}^{n}\left(x_{i,t}\right)^2} \times \sqrt{\sum_{i=1}^{n}\left(x_{i,t+1}\right)^2}}\right) \qquad (2)$$

with $x_{i,t}$ and $x_{i,t+1}$ representing the proportion of response type x_i for consecutive age groups, t and $t+1$, and n representing the number of different response types elicited by that word in the two age groups. In order to meet the assumptions of normalcy for later statistical analysis, these data was log transformed to reduce skewness.

Network Analysis

Each of the network measures we use captures increasingly more global information. This allows us to investigate how words change over time, taking into account different amounts of information about its structure as we go from local structure (degree centrality) to global structure (the small world index). Before explaining what each of these measures mean, we first explain how the network was derived.

One-mode network. Constructing a graph based on cues and their associates results in a two-mode graph, also known as bipartite graph, with one node type denoting cue words, and a second node type denoting associations (or target words). To facilitate graph analysis, the two-mode graph was compressed into a one-mode (square matrix) graph using the projection method developed by Newman (2001) and Opsahl (2013). This allows the use of the entire response repertoire in the network construction; it retains cue words as nodes and relations between target words as directed edges. Two cue words in the projected graph share directed edges if they both have edges to the same target node (i.e., if they led to the production of the same associate). The directed edges are weighted in proportion to their relative production of shared targets, meaning the edges between words represent the strength with which they produce shared associates. The new graph therefore represents the general structure of associations, with cue words with shared patterns of association linked together. The Appendix provides a detailed account of this projection method along with further rationale.

For each of the nine graphs, degree centrality, clustering coefficient, average shortest path, and small world index were computed using generalized methods for weighted directed graphs (Opsahl, Agneessens, & Skvoretz, 2010; Opsahl & Panzarasa, 2009). Below we provide a short summary of each of these measures (for further details about each of these measures see the Appendix).

In-degree and *out-degree*. In and out degree (k^{in}, k^{out}) represent the centrality of the nodes in the network on the local level and distinguishes between ingoing and outgoing edges. Our measures of in- and out-degree use Opsahl's method (Opsahl et al., 2010). This method combines both edge count and edge weight into an integrated measure, allowing us to evaluate the impact of the projection across edge count and weights. Our results show the same qualitative pattern for both degree and node strength (i.e., weighted degree) measures. Henceforth, we refer to Opsahl's integrated measure as degree.

Clustering coefficient. The clustering coefficient (C) indicates the interconnectivity among the neighbors of a node. Words whose immediate neighbors are connected among themselves have higher clustering coefficients than words whose neighbors are not connected.

Average shortest path. The average shortest path (L) between a pair of nodes indicates how well a node is connected to any other node in the network, and therefore measures its role in the entire network structure. A central word that is well connected would have a smaller average shortest path length than a more peripheral word.

Small world index. A small world network has higher relative clustering coefficient and average shortest path than a random network of the same size and density (the probability of an edge in the network). Humphries & Gurney (2008) proposed a small-world index (*SWI*) that measures how much a network deviates from randomness in relation to its small-world properties and this is the measure we use here.

Statistical tests

We compare changes in the above measures to null hypothesis distributions based on simulated Erdös-Renyi random networks. This was done for each measure and each age group separately. In addition, in order to assess associative changes, the data for each measure (entropy, clustering coefficient, average shortest path-length, and in/out degrees) were submitted to a multivariate analysis of variance (MANOVA) with age group as the independent variable. Bayesian Information Criterion (BIC; Schwarz, 1978), which penalizes models according to their complexity, was used to evaluate the shape of the curve across the lifespan. Regression analysis was used to predict associative changes between consecutive age groups. A hierarchical regression was used in order to assess the contribution of different factors to associative change. Additional details on each of these tests along with controls for education and word knowledge, mentioned in the Discussion, can be found in the Appendix.

Results

In order to describe the lexicon structure and its change, we used two types of measures: word level measures (entropy and associative change) that are computed directly on the raw association data, and network level measures (in/out degree, clustering coefficient, average shortest path and small-world index) that are computed on the network projection.

Entropy has a U-shaped structure across the lifespan

The results for changes in the entropy of associations are shown in Figure 2. The U-shaped pattern for normalized entropy reveals that associations tend to become more predictable as individuals age from childhood up until age 30. After age 30, the pattern reverses, with older individuals producing increasingly dissimilar responses with age. These results indicate that words differ in the entropy of their response associations between age groups (results of a one-way ANOVA: $F_{(8,3605)}=17.42$, $MSE=.002$, $\eta^2=.037$, $p<.001$). Permutation tests confirm this is not simply a result of changes in density ($p<.001$ for each age group).

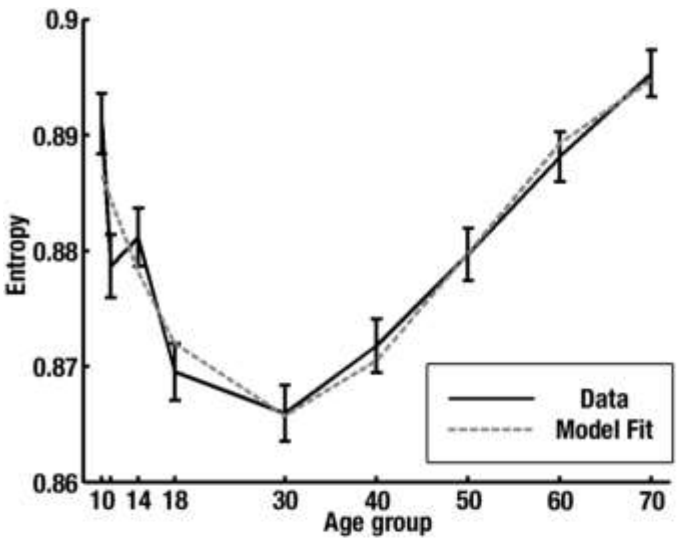

Figure 2. Average entropy of words associations across age groups. Bars represent standard error of the mean. The dashed gray line represents the cubic polynomial fit which was found to be the best fitting polynomial model (table 2).

To investigate the non-linear pattern in Figure 2, we calculated the BIC for linear and various polynomial models listed in Table 2. As shown, the optimal model for entropy (H) was a cubic model, confirming the non-linear U-shaped pattern visible in Figure 2.

Table 2.

	Linear	Quadratic	Cubic	4th degree
C	-87.36	-92.67	-95.83	**-116.90**
L	-34.23	-42.89	-49.83	**-56.09**
k^{in}	18.73	11.42	5.77	**4.30**
k^{out}	18.49	11.21	5.73	**4.21**
H	-82.07	-92.10	**-98.40**	-96.93

Note: Bayesian Information Criterion (BIC) scores for five measures. The best fitting model is marked in bold.

Network measures have a U-shaped structure across the lifespan

Figure 3 presents the network structure for a representative sample of ages across the lifespan. The networks visually present a gradual nonlinear change in structure across the lifespan. This is evident in the number of isolates shown in the hemisphere around the larger central component. It is also apparent in the number of interconnections in the central component, which is sparsest in early and late life. In what follows we present the network statistics that support this visual progression.

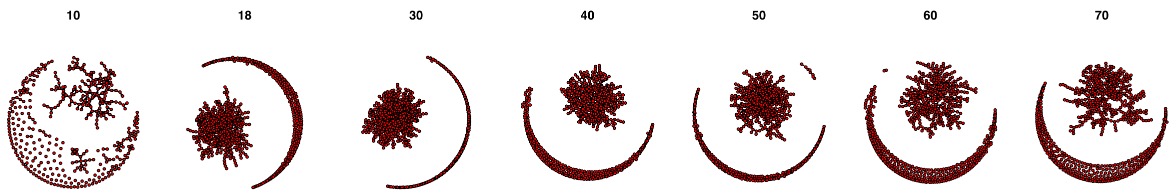

Figure 3. Free association networks across the lifespan. These networks were produced by setting a threshold of 5 for each directed edge. Isolates and small components are combined in the crescents around the outer perimeter. The giant component is centered in each image.

A multivariate analysis of variance for the four network-level measures (MANOVA) revealed that the network structure indeed changes across life (Wilk's λ=.747,

$F_{(32,13285)}$=*34.114, p<.001*). Figure 4 presents the results for each of the network's measures over the lifespan. For the in- and out- degree we find that the cue words start with relatively few *in* and *out* links to other words, followed by a dramatic increase into midlife, and then a drop in late life (Figure 4a and 4b, respectively). This is supported by one-way ANOVA's for both the *in*- and *out*- *degree* (k^{in}: $F_{(8,3605)}=34.93$, $MSE=75.87$, $\eta^2=.072$, $p<.001$; k^{out}: $F_{(8,3605)}=28.27$, $MSE=89.02$, $\eta^2=.059$, $p<.001$), confirming these measures differ between the age groups. For the average shortest path (Figure 4c), cue words start with relatively long paths, followed by shorter paths in midlife, and then a lengthening toward old age, $F_{(8,3605)}=116.77$, $MSE=.077$, $\eta^2=.206$, $p<.001$. This decrease and later increase in the average shortest path indicates that the words move towards more densely associative patterns of connectivity in midlife, but become more distant in their associations later in life.

Clustering coefficient (Figure 4d) shows a decrease throughout life, with a possible increase in later life, $F_{(8,3605)}=21.91$, $MSE=.004$, $\eta^2=.046$, $p<.001$. The large decrease indicates that the words' immediate environment become less clustered over development, i.e. words neighbors become less connected among themselves. All of the network results are further supported by permutation tests, indicating these patterns are not driven by the underlying constraints inherent in the network density (*p<.001* for each measure at each age group).

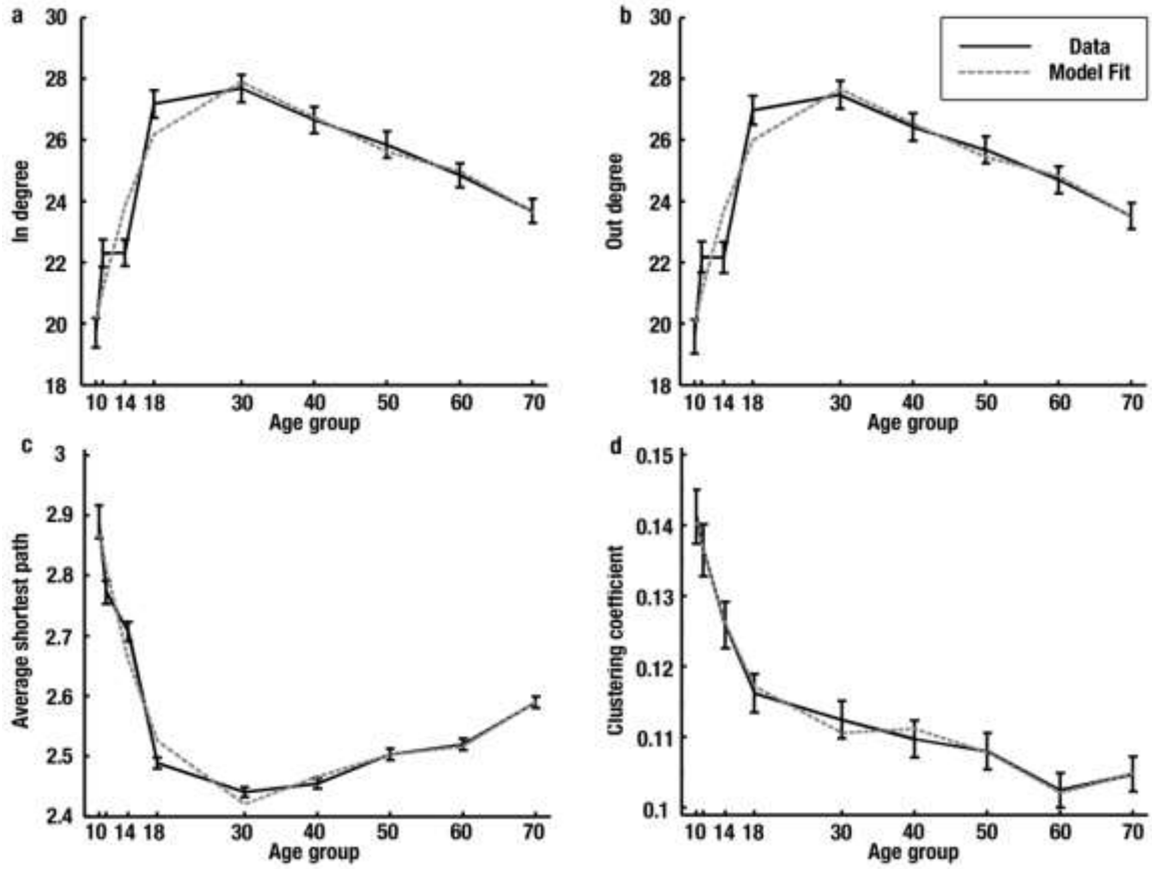

Figure 4. Network analysis measures across the lifespan for in-degree (a), out-degree (b), average shortest path (c), and clustering coefficient (d). Bars represent standard error of the mean. Dashed gray lines represent the best fitting polynomial models from Table 2.

The average shortest path and in and out degree clearly conform to a U-shape pattern throughout life, as confirmed by the non-linear polynomial fits (Table 2). This U-shape pattern is in contrast to the near monotonic decrease in clustering coefficient. Finally, the small-world index in Figure 5 shows a similar U-shaped pattern, with small world indices greatest in early and late life.

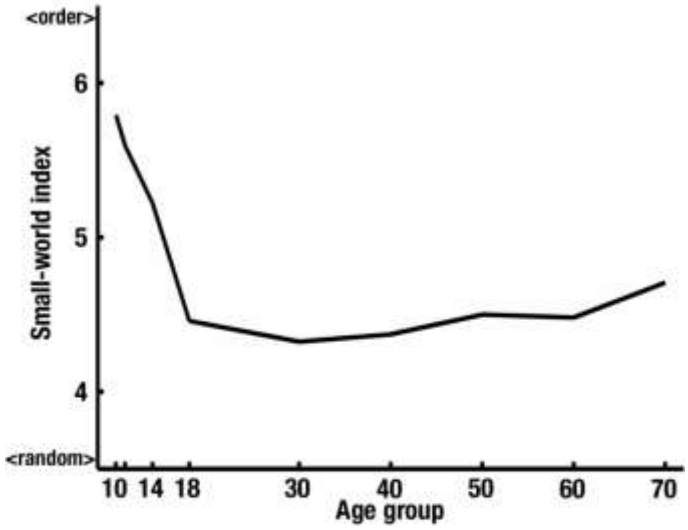

Figure 5. The small-world index across the lifespan.

Cross-year associative change is driven by words at the periphery

To understand which words are changing the most across the lifespan, we ran a regression predicting cross-year associative change using the network measures. Table 3a shows for each pair of consecutive age bins that the network-level measures are all significant predictors of cross-year change (*p*<.001). Notably, words that have lower clustering coefficients, higher average path length to other nodes, higher in-degree and lower out-degree are the words most likely from one year to the next. Because in-degree and out-degree are on the same scale (and well correlated), we can see that lower out-degree has the strongest effect. Collectively, these measures suggest that words that are least well connected to other nodes are the words that change the most from one year to the next.

Table 3b shows that entropy becomes the most important predictor when entered into the regression alongside the network statistics. The positive coefficient is consistent with the results we observe above for the network statistics. Words with numerous weak associates, that are least predictable in their relations with other words, the greater the change in associations over the lifespan.

Table 3a

	10-11	11-14	14-18	18-30	30-40	40-50	50-60	60-70
C	-.116*	-.148*	-.135*	-.119*	-.103*	-.148*	-.093*	-.118*
L	.175*	.251*	.151*	.201*	.153*	.219*	.087*	.100*
k^{in}	.140*	.145*	.103*	.142*	.164*	.142*	.130*	.111*
k^{out}	-.368*	-.213*	-.292*	-.157*	-.193*	-.132*	-.243*	-.202*
R^2	.511	.385	.408	.324	.322	.381	.334	.300

Table 3b

	10-11	11-14	14-18	18-30	30-40	40-50	50-60	60-70
C	-.036	-.018	-.016	-.010	-.016	-.002	.015	-.014
L	.020	.078	-.001	.035	-.021	.045	.002	-.043
k^{in}	-.041	-.055	-.029	-.021	-.044	-.007	-.001	-.036
k^{out}	-.045	.063	-.016	.057	.007	.034	-.035	.006
H	.539*	.530*	.543*	.571*	.547*	.491*	.469*	.495*
R^2	.796	.658	.735	.750	.755	.740	.713	.743

Note: Beta coefficients for the four network-level measures (3a), and including entropy (3b), in a two-step hierarchical regression. * p<.001.

Discussion

The present work makes a number of contributions to understanding how associative patterns change across the lifespan and what factors influence this change at the level of the individual word. Our results demonstrate a complex pattern of associative change, one that is reflected by a rapid increase in associative consistency during the formative years of language acquisition (up to early adult life)—as indicated by a reduction in entropy and average shortest path length. This is also indicated by an increase in *in*- and *out*-degree; our projection increases the number of edges between nodes as the associative responses become more shared across cues. In late life all measures reverse direction. Older ages showed increasing average path length, smaller in and out degree, and increasing entropy. The collective pattern indicates that the associative networks begin rather sparse, with increasing numbers of density towards midlife, followed by an increasing sparseness as individuals move into old age. The data suggests a U-shaped pattern of associative development, suggested in previous work (Zortea et al., 2014).

More specifically, however, our results show that late adult life is not merely the inverse of early life. In late life, words do not revert to the same structure found in early life, as indicated by comparing early and late life clustering coefficients. This is also reflected in the small-world index, showing that late life networks are not the same small worlds observed in early life. This reflects a strength of the network approach, as this difference is not clear from the entropy level analysis alone.

Broadly, we can characterize this fairly continuous developmental trajectory into three stages: 1) a pre-adult formative stage, 2) a midlife plateau, and 3) a late-life expansion. The network breathes, contracting into an ordered mid-life stage, that then loses coherence into late life. This is a simplification, of course, as the demarcations of these stages follow different patterns for each of our measures. The *in* and *out degree* change the most at age 18, whereas the *small word index* is lowest at age 50, with each of the other measures suggesting ages somewhere in between.

The difference in inflection points for the U-shaped patterns may reflect complex developmental interactions between structural change, such as vocabulary learning, and cognitive control. These have different characteristic dynamics across the lifespan (Salthouse, 2009). Past research has often confounded the independent roles of cognitive structures and the cognitive control processes that access those structures (Jones, Hills, & Todd, 2015)—they are not the same and both are likely to contribute to age-related changes in cognition. Several recent studies have demonstrated how changes in cognitive control processes can lead to different patterns of retrieval in both memory search and problem solving (Hills et al., 2013; Hills, Todd, & Goldstone, 2010; Hills & Pachur, 2012). Alternatively, numerous studies have shown that word knowledge is acquired throughout adulthood and is a language learning capacity persevered into old age (e.g., Brysbaert et al., 2014). In part, this follows naturally from the Zipfian nature of language: many words are rare—encountering and learning them requires extended exposure to language (cf. Landauer & Dumais, 1997). The result is that associations are likely to reflect changing control processes searching over a gradual accumulation, and possible decay, of lexical knowledge across the lifetime.

Lifespan research is often subject to criticisms regarding potential cohort differences. Older individuals may have different associates for words like *computer* and *tablet* not because of age-related effects in cognition, but because these words had different meanings 70 years ago than they do now. To investigate this, we used a chronological dictionary of Dutch (Van der Sijs, 2001) to remove 36 words that were either introduced in Dutch before 1930 or missing from this dictionary. Excluding these words did not affect our results in any way.

Different age groups may also differ in their levels of education. Unfortunately, we do not have education data on all our participants. However, for those for which we do have this data, controlling for education levels provided results nearly identical to those we report above. However, since our data on education does not cover our entire sample, further research is needed.

Finally, it might be that there are systematic differences between the different age groups because the young participants were recruited differently (through parents' consent in schools) compared to the older participants. This interpretation is unlikely. The differences within the young participants for all measures were extreme, indicating that our recruitment method did not yield a homogenous group. As a more conservative test, we also considered evidence based only on the age groups recruited on a voluntary basis (18- to 68-year-olds). For all analyses, the same qualitative results were obtained. A variety of additional analyses showed that the nature and magnitude of the effects are quite robust against cohort differences (see Appendix for further details on the above tests). Though more costly and time consuming, longitudinal data would of course be ideal.

Notwithstanding the obvious caveats, our results are quite promising with respect to the predictive utility of associative norms during different stages of development. In particular, they offer an inroad for understanding the relative stability of word meanings over time. The prominent use of the University of South Florida free association norms (Nelson, McEvoy, & Schreiber, 2004) in cognitive modeling work has been extremely productive (e.g., Griffiths, Steyvers, & Firl, 2007; Hills et al., 2009b). Our results suggest that age-appropriate norms may further enhance this productivity.

Our results also offer insight into word-level factors that influence associative change across the lifespan. Words with more heterogeneous connections (as measured by entropy) were the most likely to change their associative structure over the lifespan; the more diverse the response profile of a word, the more likely it was to show an age-related change in its associations. In our results, the pattern is one where words that are less well connected become more well-connected into midlife, and then reverse this pattern in old age. More poetically, the lexicon appears to breathe—with an inhalation and ordering peaking in mid-life followed by an exhalation and relaxing of order into late life.

What determines a word's capacity to change are the associations it has already acquired. This is referred to as entrenchment, following Stefanowitsch & Gries (2003), meaning the degree to which the formation and activation of word associations is routinized and automated in the mental lexicon. This correlates with the frequency of occurrence with associations (Langacker, 1987; Schmid, 2010). Importantly, recent results by Baayen, Tomaschek, Gahl, & Ramscar (2016) showed that it is more difficult to learn new associations for well entrenched words relative to less entrenched words, as evaluated by their lexical entropy. Our results are consistent with this interpretation.

Finally, we note that like previously existing association norms, our norms are aggregated across individuals and may therefore not reflect the lexical representations of any single individual. Inferences from aggregated association norms are generally the rule in cognitive psychology and they have been highly successful at predicting behavior (e.g., Griffiths et al., 2007; Hills et al., 2009b). Nonetheless, corroborating inferences about individual change are naturally limited by the difficulty of acquiring longitudinal data from individuals. Although it might be impossible to track the lifetime development of the lexicon of an individual comparable to what we have reported here (covering more than 60-years), a longitudinal study of a more modest scale would be a natural next step.

References

Anglin, J. M., Miller, G. A., & Wakefield, P. C. (1993). Vocabulary Development: A Morphological Analysis. *Monographs of the Society for Research in Child Development*, *58*(10), i-186. doi:10.2307/1166112

Arias-Trejo, N., & Plunkett, K. (2013). What's in a link: Associative and taxonomic priming effects in the infant lexicon. *Cognition*, *128*(2), 214–227. doi:10.1016/j.cognition.2013.03.008

Baayen, R. H., Tomaschek, F., Gahl, S., & Ramscar, M. (2016). The Ecclesiastes Principle in Language Change Harald Baayen. In M. Hundt, S. Mollin, & S. Pfenninger (Eds.), *The changing English language: Psycholinguistic perspectives*. Cambridge: Cambridge University Press.

Baronchelli, A., Ferrer-i-Cancho, R., Pastor-Satorras, R., Chater, N., & Christiansen, M. H. (2013). Networks in cognitive science. *Trends in Cognitive Sciences*, *17*(7), 348–360. doi:10.1016/j.tics.2013.04.010

Barrat, A., Barthélemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(11), 3747–3752. doi:10.1073/pnas.0400087101

Bilson, S., Yoshida, H., Tran, C. D., Woods, E. A., & Hills, T. T. (2015). Semantic facilitation in bilingual first language acquisition. *Cognition*, *140*, 122–134. doi:10.1016/j.cognition.2015.03.013

Borge-Holthoefer, J., & Arenas, A. (2010). Semantic Networks: Structure and Dynamics. *Entropy*, *12*(5), 1264–1302. doi:10.3390/e12051264

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–11. doi:10.3758/s13428-013-0403-5

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, *45*(2), 480–498. doi:10.3758/s13428-012-0260-7

De Deyne, S., & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior Research Methods*, *40*(1), 213–231. doi:10.3758/BRM.40.1.213

Dörken, H. (1956). Frequency of common association. *Psychological Reports*, *2*(3), 407–408.

Griffiths, T. L., Steyvers, M., & Firl, A. (2007). Google and the Mind. *Psychological Science*, *18*(12), 1069–1076. doi:10.1111/j.1467-9280.2007.02027.x

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore, MD: Paul H Brookes Publishing.

Hills, T. T. (2013). The company that words keep: comparing the statistical structure of child-versus adult-directed language. *Journal of Child Language*, *40*(3), 586–604. doi:10.1017/S0305000912000165

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009a). Categorical structure among shared features in networks of early-learned nouns. *Cognition*, *112*(3), 381–96. doi:10.1016/j.cognition.2009.06.002

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009b). Longitudinal analysis of early semantic networks: preferential attachment or preferential acquisition? *Psychological Science*, *20*(6), 729–39. doi:10.1111/j.1467-9280.2009.02365.x

Hills, T. T., Mata, R., Wilke, A., & Samanez-Larkin, G. R. (2013). Mechanisms of age-related decline in memory search across the adult life span. *Developmental Psychology*, *49*(12), 2396–2404. doi:10.1037/a0032272

Hills, T. T., & Pachur, T. (2012). Dynamic search and working memory in social recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 218–228. doi:10.1037/a0025161

Hills, T. T., Todd, P. M., & Goldstone, R. L. (2010). The central executive as a search process: Priming exploration and exploitation across domains. *Journal of Experimental Psychology: General*, *139*(4), 590–609. doi:10.1037/a0020666

Hirsh, K. W., & Tree, J. J. (2001). Word association norms for two cohorts of British adults. *Journal of Neurolinguistics*, *14*(1), 1–44. doi:10.1016/S0911-6044(00)00002-6

Humphries, M. D., & Gurney, K. (2008). Network “small-world-ness”: a quantitative method for determining canonical network equivalence. *PloS One*, *3*(4), e0002051. doi:10.1371/journal.pone.0002051

Jones, M. N., Hills, T. T., & Todd, P. M. (2015). Hidden processes in structural representations: A reply to Abbott, Austerweil, and Griffiths (2015). *Psychological Review*, *122*(3), 570–574. doi:10.1037/a0039248

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. doi:10.1037/0033-295X.104.2.211

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. 1). Stanford university press.

Light, L. L. (1992). The organization of memory in old age. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 111–165). Hillsdale, NJ:

Erlbaum.

Lovelace, E. A., & Cooley, S. (1982). Free Associations of Older Adults to Single Words and Conceptually Related Word Triads. *Journal of Gerontology*, *37*(4), 432–437. doi:10.1093/geronj/37.4.432

Maratsos, M. P. (2005). How Children Learn the Meanings of Words (review). *Language*, *81*(2), 495–498. doi:10.1353/lan.2005.0073

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407. doi:10.3758/BF03195588

Newman, M. E. J. (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, *64*(1), 16132. doi:10.1103/PhysRevE.64.016132

Opsahl, T. (2013). Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks*, *35*(2), 159–167. doi:10.1016/j.socnet.2011.07.001

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251. doi:10.1016/j.socnet.2010.03.006

Opsahl, T., & Panzarasa, P. (2009). Clustering in weighted networks. *Social Networks*, *31*(2), 155–163. doi:10.1016/j.socnet.2009.02.002

Palermo, D. S. (1964). Word Associations and Children's Verbal Behavior. In *Advances in Child Development and Behavior* (Vol. 1, pp. 31–68). doi:10.1016/S0065-

2407(08)60326-2

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: non-linear dynamics of lifelong learning. *Topics in Cognitive Science*, *6*(1), 5–42. doi:10.1111/tops.12078

Riegel, K. F., & Birren, J. E. (1965). Age differences in associative behavior. *Journal of Gerontology*, *20*(2), 125–130.

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*(4), 507–514.

Schmid, H. (2010). Does frequency in text instantiate entrenchment in the cognitive system ? In *Quantitative Methods in Cognitive Semantics: Corpus-driven Approaches* (pp. 101–133).

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. doi:10.1214/aos/1176344136

Shapiro, S. S. (1964). Meaningfulness values for 52 CVCs for grade-school-aged children. *Psychonomic Science*, *1*, 127–128.

Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, *8*(2), 209–243. doi:10.1075/ijcl.8.2.03ste

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science*, *29*(1), 41–78. doi:10.1207/s15516709cog2901_3

Tresselt, M. E., & Mayzner, M. S. (1964). The Kent-Rosanoff word association: Word association norms as a function of age. *Psychonomic Science*, *1*, 65–66.

Van der Sijs, N. (2001). *Chronologisch woordenboek*. LJ Veen.

Vitevitch, M. S., Chan, K. Y., & Goldstein, R. (2014). Insights into failed lexical retrieval from network science. *Cognitive Psychology*, *68*, 1–32. doi:10.1016/j.cogpsych.2013.10.002

Zortea, M., Menegola, B., Villavicencio, A., & Salles, J. F. De. (2014). Graph analysis of semantic word association among children, adults, and the elderly. *Psicologia: Reflexão E Crítica*, *27*(1), 90–99. doi:10.1590/S0102-79722014000100011

Appendix

Projecting the two-mode association network onto a one-mode network

We transformed our two-mode graph of cues and associations into a weighted directed graph based on common associations shared between cue words. In the one-mode projection, nodes represent cue words, but the edges no longer represent the associations between the words (two nodes are not tied together because one is a direct associate of the other). Instead, the edge weights are defined asymmetrically as the count of associates from word *i* to the associates shared with word *j*. This is normalized by the relative distinctiveness of the shared associates by dividing the number of times these associations are shared over all the cue words. More formally:

$$w_{ij} = \sum_{i=1}^{P} \frac{w_{i,p}}{N_p - 1}$$

where w_{ij} is an edge's weight in the one-mode graph directed from node *i* to node *j*. *P* is the number of shared associations between cue words *i* and *j*. $w_{i,p}$ is the number of times cue *i* produced *p* as an associate. N_p is the number of cue words that produced *p*. Cue words' idiosyncratic responses were removed prior to projection.

To provide some intuition for this measure, consider the cue words *fish* and *bird,* which led to the production of similar associative targets, *pet* and *food*. If no other cue words produced *pet* but many other cue words produced *food*, then *pet* will make a larger contribution to the projected edge weight than *food*, as it is a more distinct association. A threshold of $w_{ij} < 1$ was used to ensure that very weak edges were removed from the graph[2]. The projection method results in a directed graph, in which nodes are cue words, and edges convey information about their similarity as measured through their shared associates.

[2]The value of the cutoff can be varied without influencing the results.

Note that the intention of the projection is not to represent a cognitive lexical representation, but rather to capture the structural properties of the more complex bipartite free association network in a way that allows us to quantify structural properties of all of the data as it changes over the lifespan. We feel this is preferable to representations based only on cue-cue associations and undirected networks, which we can confirm show the same qualitative patterns across the lifespan as presented here.

In and out degree

In order to preserve as much information on the nodes' connections as possible, we used the Opsahl's method (Opsahl et al., 2010). This allowed us to vary the influence of the counts of the number of connections and their weights, as follows:

$$k_{in/out}(i) = k_i^{(1-\alpha)} \times w_i^{\alpha} = \sqrt{k_i} \times \sqrt{w_i}$$

Our results proved to be insensitive to particular values of α. The results we report use $\alpha = .5$, equally weighting the contributions of degree and strength for each node. We keep the traditional degree *k* notation, but use this weighted variant throughout the text.

Clustering coefficient

We use a version of the clustering coefficient described for weighted networks (Barrat, Barthélemy, Pastor-Satorras, & Vespignani, 2004), where w_{ij} is the weight between nodes *i* and *j*, s_i is the sum of the weighted edges going out from node *i*, k_i is the number of neighbors of node *i*, and *j* and *h* represent all neighbors of node *i*. This measure is then defined as follows:

$$c_i = \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{w_{ij} + w_{ih}}{2}$$

This measure computes coherence based on the inter-connectedness of neighboring nodes, and does so by accounting for the weights of local edges to the target node. The normalization using *s* and *k* confines c_i to range between 0 and 1. It measures how much of the weights node *i* projects to its neighbors remain in the local neighborhood due to connections between its neighbors, and how much is lost due to neighbors that lack such connectivity.

Average shortest path

The shortest path, *L*, between two nodes is defined as the path that travels the shortest distance over the edges between the nodes. In Figure S1, an edge with a weight of 2, as a result of our projection, implies a path twice as short compared to an edge with a weight of 1. Therefore, the shortest path between A and C is the indirect path through B.

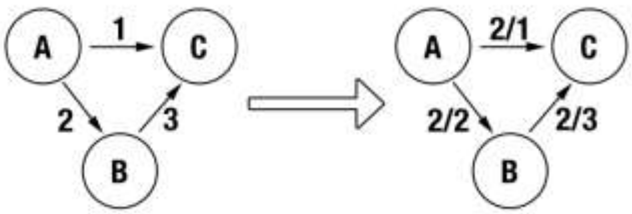

Figure S1. Transforming the network edges from weights (left) to distances (right).

In order to transform the weights to distances, the edges' weights were normalized by dividing by the average weight of the network, and then inverted using the method from Opsahl et al. (2010). In the above mentioned example (right), the direct path from A to C has a length of 2, while the indirect path has a length of 1.67.

Small-world Index

Humphries & Gurney (2008) proposed a small-world index (*SWI*) that measures how much a network deviates from randomness by taking the ratio of the normalized clustering coefficient and the normalized average path length. The normalization is computed by dividing by the average values computed for random graphs of the same size and density as follows:

$$SWI = \frac{\acute{C}}{C_{rand}^{'}} \Big/ \frac{\acute{L}}{L_{rand}^{'}}$$

Where $\acute{C}$ and $\acute{L}$ are the average clustering coefficient and average shortest path, respectively, and $C_{rand}^{'}$ and $L_{rand}^{'}$ are those measures computed for an Erdös-Renyi random graph with the same size and density (computed by randomly shuffling edges between nodes in observed networks).

Permutation tests

We compare each network statistic against a null hypothesis derived from randomized versions of the observed word level or network level properties. These allow us to conclude that the statistical patterns we observe in the data are unlikely to be an artifact of a random data generation process.

Entropy. For each cue word, the total number of responses (i.e., tokens), *n*, and the number of response types, *k*, that were produced as its associations were computed from the observed data. For each cue word, random associations were produced by sampling uniformly with replacement, *n* out of *k*, creating a response distribution for that word under a uniform probability condition. The *entropy* was then computed for the random distribution of each word according to equation 1. This was averaged across words creating the test statistic, σ. This process was repeated 10,000 times, creating a null distribution for the random entropy. The statistical significance, *p*, of the true average entropy score was defined as 1 minus the proportion of times it was smaller than σ out of the 10,000 repetitions.

Degrees, shortest-path and clustering coefficient. Random graphs were created by randomly shuffling the weighted edges of the observed graphs, creating standard Erdős–Rényi graphs with the same densities. Statistics were then computed for the random networks in the same way these were computed for the observed networks. This process was repeated 10,000 times. The statistical significances, *p*, for each statistics was defined as 1 minus the proportion of times each statistic was smaller than its random counterpart out of the 10,000 repetitions.

Education level analysis

For logistical reasons, education levels were collected for only the more recent participants that took part in the experiment: a total of 407, 396, 452, 390 and 829 participants for the age groups of 30, 40, 50, 60 and 70. Within this partial data, a large portion of participants had a Masters degree or higher (more than 4 years of University in the Belgian system). The percentages were 62% (30-years-old), 51% (40-years-old), 34% (50-years-old), 28% (60-years-old), 42% (+68-year-olds). All the remaining participants had finished some form of tertiary education.

We controlled for possible contributions of between-group differences in education levels by computing our BIC analysis for each of the 5 measures after controlling for education by using a vector rejection method. We found similar non-linear effects previously reported (see Table S1):

Table S1:

Polynomial degree	Linear	Quadratic	Cubic	4th degree
C	-93.70	-95.91	-95.83	-107.70
L	-38.69	-47.19	-52.56	-61.09
k^{in}	15.01	7.81	2.84	0.53
k^{out}	14.76	7.62	2.84	0.48
H	-83.11	-93.14	-98.93	-97.42

We repeated our analysis of variance for the six age groups between 18-68 after controlling for the effects of educational differences, with similar results as reported in the main text: entropy ($F_{(5,2418)}$ = 25.62, p<.001), k^{in} ($F_{(5,2418)}$ = 12.54, p<.001), k^{out} ($F_{(5,2418)}$ = 11.01, p<.001), *L* ($F_{(5,2418)}$ = 35.1, p<.001), and *C* ($F_{(5,2418)}$ = 3.74, p<.1).

Conservative Removal of Diachronically Suspicious Words

This list of words was removed from analysis because their first appearance in Dutch was after 1930 (Van der Sijs, 2001). This allowed us to control for words that may have entered our older participant's lexicons later in life. The removal of these words had no influence on the statistical pattern of our results.

aanraden - *to recommend*, bemind - *loved*, bikini - *bikini*, bloemen - *flowers*, CD - *CD*, concentratie - *concentration*, eeuwig - *eternal*, grappig - *funny*, gunstig - *beneficial*, horen - *to hear*, inzet - *effort*, keukengerief - *kitchen utensils*, kleding - *clothing*, metro - *metro*, muis - *mouse*, muziekinstrument - *musical instrument*, nadenken - *to think*, nakomen - *to honor*, nuttig - *useful*, ongewoon - *unusual*, opletten - *to pay attention*, sappig - *juicy*, schattig - *cute*, shoppen - *to shop*, snoep - *candy*, speels - *playful*, stank - *stench*, verdriet - *sadness*, vergissen - *to mistake*, verkiezen - *to prefer*, vriendelijk - *friendly*, vriendschap - *friendship*, waarheid - *truth*, wonde - *wound*, ziekenhuis - *hospital*, zielig - *pathetic*.