From "sooo excited!!!" to "so proud": Using Language to Study Development

Margaret L. Kern[1], Johannes C. Eichstaedt[1], H. Andrew Schwartz[1],
Gregory Park[1], Lyle H. Ungar[1], David J. Stillwell[2], Michal Kosinski[2],
Lukasz Dziurzynski[1], Martin E. P. Seligman[1],
[1] University of Pennsylvania, [2] University of Cambridge

Author Note

Margaret L. Kern, Department of Psychology, University of Pennsylvania; Johannes C. Eichstaedt, Department of Psychology, University of Pennsylvania; H. Andrew Schwartz, Computer & Information Science, University of Pennsylvania; Gregory Park, Department of Psychology, University of Pennsylvania; Lyle H. Ungar, Computer & Information Science, University of Pennsylvania; David J. Stillwell, Psychometrics Centre, University of Cambridge; Michal Kosinski, Psychometrics Centre, University of Cambridge; Lukasz Dziurzynski, Department of Psychology, University of Pennsylvania; Martin E. P. Seligman, Department of Psychology, University of Pennsylvania

Support for this publication was provided by the Robert Wood Johnson Foundation's Pioneer Portfolio, through the "Exploring Concepts of Positive Health" grant awarded to Martin Seligman and by the University of Pennsylvania Positive Psychology Center.

Correspondence concerning this article should be addressed to Margaret L. Kern. Email: mkern@sas.upenn.edu

Final accepted version, September 2013, *Developmental Psychology*. This paper is not the copy of record and may not exactly replicate the authoritative document published in the journal. The final article is available at http://dx.doi.org/10.1037/a0035048

Abstract

We introduce a new method, *differential language analysis (DLA)*, for studying human development that uses computational linguistics to analyze the big data available through online social media in light of psychological theory. Our open vocabulary DLA approach finds words, phrases, and topics that distinguish groups of people based on one or more characteristics. Using a dataset of over 70,000 Facebook users, we identify how word and topic use vary as a function of age, and compile cohort specific words and phrases into visual summaries that are face valid and intuitively meaningful. We demonstrate how this methodology can be used to test developmental hypotheses, using the aging positivity effect (Carstensen & Mikels, 2005) as an example. While this study focuses primarily on common trends across age-related cohorts, the same methodology can be used to explore heterogeneity within developmental stages or to explore other characteristics that differentiate groups of people. Our comprehensive list of words and topics are available on our website for deeper exploration by the research community.

Keywords: Emotion, Adult development, Language use, Measurement, Online social media

From "sooo excited!!!" to "so proud":

Using Language to Study Development

The recent explosion of social media has resulted in massive datasets with tens of thousands of people and millions of observations, allowing for “data intensive decision-making, including clinical decision making, at a level never before imagined” (National Science Foundation, 2012, para. 4). The social sciences have testable theories in need of rich naturalistic data, but some of the most trusted analytic tools of these fields are insufficient for datasets with millions of observations. Computer scientists are developing methods to efficiently manage and analyze the huge volumes of data generated by online human behaviors and interactions. One avenue to strategically approach such massive datasets is to combine cutting-edge methods from computer science with well-developed theories from the social sciences.

Developmental psychology in particular has been a forerunner in developing and using multiple methods (e.g., surveys, interviews, observations, quasi-experiments), modalities (e.g., self-report, observer ratings, language analysis), and statistical tools. In this paper, we add a novel instrument to the developmental methodological toolbox that combines big data available through online social media, analytic capabilities from computational linguistics, and insights and interpretations from psychology. We describe the tool and draw on a dataset of over 70,000 Facebook users to examine age-related differences in word use, highlighting special features that may be useful to developmental researchers. We test the aging positivity effect (Carstensen & Mikels, 2005) to demonstrate how the tool can be used to test developmental hypotheses.

Learning from Words

In the current investigation, we introduce a method that combines millions of thoughts, expressions, and emotions and creates language topics to make sense of individual textual statements. Our method uses *differential language analysis* (DLA) – a technique that finds distinct sets of words, phrases, and topics that distinguish groups of people based on one or more characteristics (e.g., age, gender, location, personality). Drawing on analytic methods used in computational linguistics, informative words and phrases (i.e., two or more words that occur together) are extracted from each set of text (e.g., one Facebook message). Similar to latent class cluster analysis, an algorithm iteratively finds words that cluster together, allowing the data to define categories. Visualization is an important final step in our method. Results are compiled into images (e.g., words across age; dominant words or categories distinguishing one group versus another), allowing for intuitive access to a large amount of information. We classify this method as an *open vocabulary* approach, as it does not utilize any pre-determined word-category judgments.

Our method is not the first to automatically count word occurrence. Most familiar to the psychological literature, Pennebaker and Francis (1999) created the Linguistic Inquiry and Word Count (LIWC) software program, enabling exploration of individual differences in the frequency of single words that people write or speak. Using the program, Pennebaker and Stone (2003) compiled writing samples from 45 different studies, including over 3,000 individuals aged eight to 85 years old, and tallied word occurrence in 14 categories. Older individuals used more positive words and the future tense, whereas younger individuals used more negative words, first person pronouns, and the past tense. Although the findings suggest age-related differences in word use, the LIWC program is based on manually created categories that reflect the backgrounds and biases of the creators. The authors note: “in the years to come, a significant rethinking is needed of the ways words are used and how their usage ties to psychologically interesting variables” (Pennebaker & Francis, 1999, p. 300). Our method addresses this challenge through an open-ended analysis of the words that people voluntarily write in the course of their daily lives.

Our method is also not the first to automatically organize qualitative information. A growing number of tools and algorithms are available for analyzing interviews, books, online searches, and more (e.g., Dedoose, NVivo, MAXQDA, SAS Sentiment Analysis, WordSmith). Our method is particularly relevant for identifying characteristics that distinguish groups of people (based upon age, gender, personality, etc.) in large social media datasets, and complements other methods designed for different purposes or for different data sources.

The Age and Emotion Paradox

To demonstrate how our method can be used to test developmental theory, we explore the aging positivity effect (Carstensen & Mikels, 2005), which states that older people are happier than younger people, despite cognitive and physiological declines (e.g., Carstensen & Mikels, 2005; Isaacowitz & Blanchard-Fields, 2012; Lawton, 2001; Scheibe & Carstensen, 2010). Old age is often thought of negatively by both younger and older individuals (e.g., Garry & Lohan, 2011; Nosek, Banaji, & Greenwald, 2002), yet “the observation that emotional well-being is maintained and in some ways improves across adulthood is among the most surprising findings about human aging to emerge in recent years” (Carstensen et al., 2011, p 21). For instance, in a study in which 184 adults age 18 to 94 were paged five times per day for a week to rate 19 different emotions, the frequency of negative emotion decreased linearly through age 60 and then leveled off, whereas positive emotions remained fairly stable, such that the overall positivity ratio increased across age (Carstensen, Pasupathi, Mayr, & Nesselroade, 2000). A 10-year follow-up study further supported these trends (Carstensen et al., 2011).

Most consistently, negative emotion declines across adulthood (e.g., Carstensen et al., 2000; Charles, Reynolds, & Gatz, 2001; Mroczek, 2001; Gross, Carstensen, Tsai, Skorpen, & Hsu, 1997; Stone, Schwartz, Broderick, & Deaton, 2010). Positive emotion trends have been mixed, with some studies finding stable levels of intensity and frequency across ages (e.g., Carstensen et al., 2000), some finding increases (e.g., Biss & Hasher, 2012; Diehl, Hay, & Berg, 2011; Gross, Carstensen, Tsai, Skorpen, & Hsu, 1997), and others finding decreases (e.g., Griffin, Mroczek, & Spiro, 2006; Kunzmann, Little, & Smith, 2000). This discrepancy may be due, in part, to the emotions that are measured (Fernandez-Ballesteros, Fernandez, Cobo, Caprara, & Botella, 2010; Grühn, Kotter-Grühn, & Röcke, 2010; Pinquart, 2001). For instance, with 277 participants age 20 to 80 years, high arousal positive affect decreased from young to middle age and then remained stable, whereas low arousal positive affect increased with age (Kessler & Staudinger, 2009). In one of the largest studies of age and well-being, with 340,847 people age 18 to 85 in the U.S., hedonic well-being decreased across age, sadness was relatively stable, and worry, stress, and anger decreased (Stone, Schwartz, Broderick, & Deaton, 2010). Together, studies suggest the importance of distinguishing different emotions and intensities.

In online social media, age is currently skewed toward younger adults, although older adults are adopting social media at increasing rates (Brenner, 2012). We believe there is value in exploring age trends within the young group, particularly in the social media environment. We predicted that (1) younger people would mention negative emotions at a greater frequency than older individuals; (2) high arousal positive emotions would remain steady across age; and (3) older adults would mention low arousal positive emotions at a higher frequency than young people.

In sum, the main purpose of this paper is to introduce and apply a new tool that uses the big data available through online social media to study trends in human development. We present a series of analyses to demonstrate the method. We start with a broad view of words that are typically used at different ages. We then zoom into more detailed topics, including word use as a function of both age and gender. Finally, we provide an example of how the method could be used to test hypotheses based on developmental theory and research by investigating the occurrence of the positivity effect in this sample and modality.

Method

Participants and Measures

Data were collected from the myPersonality application (Kosinski & Stillwell, 2011) on Facebook, although our method could be applied to other big data sources as well. Facebook was first released in 2004 to connect students and alumni from Harvard University, and quickly spread to other universities, professions, and the full public. It now includes over a billion active users (Facebook.com, 2012). Users are prompted with a space to freely share thoughts, opinions, photographs, links, and more (i.e., the status update). Facebook includes the option of adding applications, which allow users to enhance their experience beyond simply posting updates or photographs to their profile. The myPersonality application offers various personality-type tests, which users can complete and receive a report on, for instance, how extraverted or neurotic they are.

Upon first accessing the application, participants agree to the anonymous use of their test scores for research purposes. About 25% of users have also optionally allowed access to their Facebook status updates, linked by a random identification number to the myPersonality test scores. For the current investigation, we included 74,859 English-speaking users who had at least 1,000 words across their status updates,[1] with age and gender information available. Detailed location, socioeconomic status, and other demographic information was unavailable, but based upon language preferences, about 85% were from the U.S. or Canada, 14% were from the United Kingdom or other European English speaking countries, and 1% was from other locations globally. Altogether, participants contributed about 20 million status updates and 286 million words, equivalent to the words included in 363 copies of the King James Bible.

Participants self-reported gender (62% female). Upon registration, age was reported either as exact date of birth, or as current age in years. For users with date of birth information (n = 33,324), we calculated the interval between the birth date and the date of the first status update. For users for which we only had self-reported age (n = 41,535), we adjusted age to the average time interval across users between the date that the application was added and the date that statements were made by the users. Participants ranged in age from 13 to 64.[2]

Analytic Strategy: A Computational Linguistic Approach

[1] A minimal word criterion is needed to reduce noise from sparse responses. We tested 500, 1000, and 2000 word thresholds; correlations stabilized around 1000 words. Optimal cutoffs can be tested in future research.

[2] We chose to exclude the oldest users (age 65+) from our analyses, as sparse data (82 users) resulted in unstable correlation coefficients.

To examine relations between age and word use, we used a new open vocabulary technique, termed *differential language analysis* (Schwartz et al., in press). More details on the methodology are available at wwbp.org.

Briefly, "tokens" (single words) are extracted from the large sets of text using an algorithm based upon Pott's "happyfuntokenizing" (sentiment.christopherpotts.net/code-data/happyfuntokenizing.py), with modifications to identify additional social media specific language, such as emoticons (e.g. ":-)", " <3") and hashtags (e.g., "#SpidermanMovie"). The tokens are then automatically compiled into phrases, (i.e., sequences of two or three words that occur together more often than chance, such as *happy birthday* or *4th of July*), using a point-wise mutual information criteria (Church & Hanks, 1990; Lin, 1998). To focus on common language and maintain adequate power, words and phrases are restricted to those used by at least one percent of the sample. To adjust for differing lengths of text available per person, word counts are normalized by the individual's total number of words before processing, and are transformed using the Anscombe (1948) transformation to stabilize variance (i.e., to reduce the impact of an outlier who uses a single word much more than the rest of the sample).

Using an ordinary least squares linear regression framework, a linear function is fitted between independent variables (i.e., relative frequency of words or phrases) and dependent variables (e.g., age), adjusting for other characteristics (e.g., gender). The parameter estimate (β) indicates the strength of the relation. *P* values offer a heuristic for identifying meaningful correlations, but with millions of data points, tens of thousands of correlations may be significant at the $p < .05$ level. To minimize Type I errors, parameters are considered meaningful only if the *p* value is less than a two-tailed Bonferroni-corrected value of 0.001 (that is with

20,000 language features, a p value less than 0.001 / 20,000, or $p < .00000005$ is retained as important).[3]

An important component of our method is visualization, which we believe can aid the human mind in making sense of the many significant correlations. We present a series of analyses to demonstrate various features of our method that may be useful in different contexts. First, we used age as a categorical variable, similar to past research that has compared groups of young, middle, and older adults. Age was split into five, relatively equally sized groups, which we arbitrarily labeled as teenagers (age 13-18), emerging adults (age 19-22), young adults (age 23-29), early-middle adults (age 30-44), and middle-late adults (age 45-64). The 100 words or phrases most correlated with each age group (i.e., the words that most significantly distinguished that group from the rest of the sample) were combined into a word cloud using the advanced version of Wordle software (www.wordle.net/advanced). Contrary to more basic uses of this visualization technique, in these visualizations, the size of the words indicates the strength of the correlation between the word and group (β), and the intensity of the color is used to indicate the frequency of word use across posts. For example, in the top of **Figure 1,** the large phrase "like_about_you"[4] is light grey. The size indicates that it is relatively highly related to the teenager age group, whereas the color indicates that the phrase is relatively rarely used.

---

[3] The stringent Bonferroni correction is one approach for defining meaningful correlations. As a test of effect robustness, we cross-validated findings by examining the split-half reliability (Spearman ρ) between older data (range 01 Jan 2009 through 20 Jul 2010; $n_{posts} = 6{,}742{,}747$) and newer data (range 20 Jul 2010 through 07 Nov 2011; $n_{posts} = 7{,}924{,}568$), splitting the data by the mean date a message was posted. Words were adequately stable across the age groups, with some variation by age: overall: ρ = .86; age 13-18: ρ = .91; age 19-22: ρ = .77; age 23-29: ρ = .99; age 30-44: ρ = .89; age 45-64: ρ = .88.

[4] Underscores (_) are used to connect multiword phrases in our visualizations; these characters are not present in the original text.

Second, we used age as a continuous variable and examined specific words as a function of age by plotting word occurrence frequency as a time series. It is important to note that we are capturing cross-sectional trends, which may simply reflect cohort differences, not change that occurs over time. The horizontal axis indicates age and the vertical axis represents the standardized percentage of times that participants used the word at each age. A first-order LOESS line, adjusted for gender, visualizes the data trends (Cleveland, 1979). We descriptively summarize the resulting trends.[5]

Third, our method can automatically generate categories or *topics* based on words that naturally cluster together, rather than relying on manually created categories. Topics were generated using Latent Dirichlet Allocation (LDA, Blei, Ng, & Jordan, 2003). Similar to latent class cluster analysis (Clogg, 1995), LDA assumes that messages contain distributions of latent *topics*, or groups of words. Words are grouped together, and an iterative process refines the factors, based on word co-occurrence across posts (e.g., the words *bill* and *rent* are more likely to appear in the same post than *rent* and *happy*). Before creating the clusters, the number of topics to create is determined, and stop words (i.e., very frequent words with low specificity such as “the”, “as”, and “no”) are removed. We produced 2,000 total topics.[6] Topic usage was then determined by combining the word frequency information for each age group with probabilities given from LDA. The words comprising the six most distinguishing topics for each age group were combined into word clouds. Then, using the continuous age variable, we

[5] Our age group word clouds are held to significance tests while the graphs are meant as more a more nuanced descriptive visualization of our data for which significance testing is more difficult to establish.

[6] Topic lists are available in a variety of formats on our website, http://wwbp.org/data.html

selected the dominant topic from each age group and plotted topic occurrence as a time series across the age spectrum.

In the regression equation, we adjusted for gender, but additional covariates can be added to the equation. Further, word occurrence on two variables can be considered. To illustrate, we generated word clouds as a function of both age and gender. Using the regression beta weights from models with features simultaneously regressed on age and gender, the 500 features (words/phrases) most positively correlated with each of the five age groups (i.e., the 100 words/phrases visualized in Figure 1, plus the next 400 most significant correlations) were selected. Features were then sorted by their correlations with gender. The 50 features most positively (for females) and negatively (for males) correlated with gender were combined into word clouds. The size of the word indicates the absolute size of the gender correlation (i.e., larger words are more strongly correlated with gender).

Finally, we demonstrate how our approach can be used to test substantial developmental theories by examining the aging positivity effect. We examined high and low arousal positive and negative emotion word use within each age group and the continuous pattern as a function of age (e.g., time series trends of “hate” versus “proud”), by testing a modified list of emotions from the Positive and Negative Affect Schedule (Watson, Clark, & Tellegen, 1988) and the 4d Measure of Affect (Huelsman, Nemanick, & Munz, 1998).

Results

Word Use as a Function of Age

Supporting the validity of the method, the most predominant preoccupations shifted across the age range, aligned with what could be considered on-time developmental tasks (e.g., Baltes, Reese, & Lipsitt, 1980; Baltes & Smith, 2004; Havinghurst, 1972). **Figure** 1 illustrates the most frequent words used by teenagers (age 13-18) and young adults (age 23-29).[7] Teenagers mentioned "homework", "school tomorrow", and "bieber" (i.e., Justin Bieber, a popular social icon at the time). Emerging adults (not shown, age 19-22) discussed "college", "studying", and "roommate". Young adults mentioned "at work", "apartment", and "wedding". Individuals over age 30 (not shown) frequently mentioned family and health concerns (e.g., "had_cancer").

Similarly, when words are plotted as a function of age (**Figure 2**),[8] age-appropriate concerns are evident. For instance, the words "school" and "college" peak during adolescence and early 20s, respectively. "Work" increases through the late teens and early 20s, is fairly stable through adulthood, and begins to decline in the older cohorts. "Health" and "family" concerns gradually increase. The words "boyfriend" and "girlfriend" peak during teenage and the early 20s. In the late 20s, "wedding" reaches a maximum, close to the U.S. median marriage age of 27.2 (U.S. Census Bureau FactFinder, 2012). "Husband" and "wife" increase monotonically.

Other patterns are intuitively meaningful. "Apartment" becomes a concern through the 20s then decreases, whereas "house" shows an inverse pattern, dipping in the early 20s and then increasing. "Sleep" peaks around age 20. Household tasks such as "laundry" and "cleaning" increase after college. "Exercise" gradually increases, but different activities are seemingly relevant for different age cohorts; the "gym" is prevalent in the 20s and then declines, whereas "walk" dips in the 20s and 30s and then increases. Interestingly, although

[7] See http://wwbp.org/age-wc.html for word clouds for the other three age groups.

[8] We selected words that we found personally interesting or that colleagues asked about as we presented our method, but we provide these only as examples. We encourage readers to test other words at our website: http://wwbp.org/v2/age-plot.html

statements related to alcohol occur across the age range, words reflect a growing sophistication. The word “drunk” peaks at the age of 21 and then decreases. “Beer” remains high from the 20s into the early 40s, whereas “wine” monotonically increases.

Topical Language

Extending beyond single words, our method automatically creates topics that distinguish particular groups. Using differential language analysis, co-occurring words were clustered together to create 2,000 topics. **Figure 3** illustrates the four strongest topics for young adults (age 23-29) and middle-aged adults (age 45-64).[9] Again supporting the validity of the method, the most dominant categories point to common concerns shared by a particular age group. For example, the young adult topics reflect establishing life as an adult, including financial responsibilities (“bill”, “rent”, “owe”), moving out of the parents’ home (“lease”, “roommate”, “apartment”), starting to work (“job”, “interview”, “company”), and maintaining a social life (“beer”, “drinking”, “BBQ”). The dominant topics in the 45+ group include a political topic (“government”, “taxes”, “Obama”, “economy”, “benefits”) and a military topic (“freedom”, “veterans”, “lives”, “served”). Some topics reflect common concerns that distinguish teenagers from young adults, whereas other topics may reflect individual differences. Although in these analyses we compared different age cohorts, the DLA method could further be used within an age cohort to identify sub-group differences. For example, a major theme for some teenagers is scheduled classes (“English”, “history”, “chemistry”, “honors”), whereas a second theme reflects disengagement with school (“boring”, “sucks”).

[9] See http://wwbp.org/age-wc.html for the other age groups.

As illustrated in **Figure 4**, we plotted the strongest topic for each age group as a time series across the age range. Each topic peaks at its respective period. Teenagers show a dominant use of social media slang, abbreviations, and emoticons. School, work, and family become a dominant concern for emerging adults, young adults, and adults, respectively. The most dominant topic for middle-aged adults (age 45-64), suggests positive relationships (i.e., a combination of “friends”, “family”, “thankful”, “wonderful”, etc.).

How do our automatic categories compare to manually created lexica? We calculated word frequency in six of the LIWC categories (Pennebaker & Francis, 1999). Replicating Pennebaker and Stone (2003), older individuals used a great number of positive words and future tense words, and younger adults used a greater number of negative words and first person pronouns (**Figure 5a**). Aligned with our topic results (see **Figure 4**), the family category monotonically increased (**Figure 5b**). The work category was more like the school category plotted in **Figure 4**. This is perhaps not surprising, as the LIWC category includes both school-related words such as “homework”, “campus”, and “exam” and work-related words such as “worker”, “business”, and “office”. Our automatic categories allow greater sensitivity to age-related educational and occupational stages of life than the closed approach based upon manually constructed categories.

Age and Gender Co-occurrence

Greater differentiation is evident by examining words occurrence based on two variables. **Figure 6** plots words and phrases as a function of both age and gender. For example, women in their 20s were more likely to use the words “shopping”, “excited”, and “can’t_wait”, whereas men in their 20s were more likely to use the words “himself”, “beer”, and “iphone”.

Older women used words such as “thank you” and “beautiful”; older men mentioned political type words (e.g., “president”, “obama”, “government”). Teenage women used emoticons such as “<3”, “:(“ and “:)”; and men in their early 20s used more swear words.

An Applied Example of Testing Psychological Theories: The Aging Positivity Effect

The patterns above provide support for the validity of the differential language analysis instrument and highlight features that may be valuable for research questions. Finally, we tested whether our approach can be used to test psychological theories. We selected emotions that represented high arousal positive affect (e.g., excited, energetic, vigorous), low arousal positive affect (e.g., serene, proud, grateful), high arousal negative affect (e.g., hate, angry, distressed), and low arousal negative affect (e.g., bored, weary, dull) and examined word frequency across the age range. In line with the exploratory open vocabulary approach, we selected five words that were significantly different at different ages (“hate”, “bored”, “excited”, “proud”, and “grateful”). **Figure 7** plots the time series for each word as a function of age. Providing some support for the positivity effect, both high and low arousal negative emotion words (“hate” and “bored”, respectively) decreased across the age range, high-arousal positive emotion (“excited”) showed a similar decline after peaking in the 20s, whereas low-arousal positive words (“grateful”, “proud”) gradually increased. Similarly, words such as “sad”, “angry”, and “energetic” decreased over time (not shown). However, other positive and negative emotions demonstrated inconsistent trends. For example, “anxious” increased through the 20s and then remained level and “calm” was level across the age range.

Most research on age and emotion assesses multiple positive and negative emotions and then combines the emotions based on valence, frequency, and/or intensity. As indicated in

Figure 5a, the LIWC positive and negative emotion categories linearly increased and decreased, respectively. Do such categories naturally appear in the data? We manually examined the previously generated topics that reflected emotion. High arousal was seemingly represented in emoticons and net-speak, which were more prevalent in the younger ages. However, no clear emotion topics appeared; topics were over-inclusive of other non-emotion words.

Discussion

Computational social science has arrived. Taking advantage of the vast amount of data available through social media, techniques developed in computational linguistics, and developmental theory from psychology, we introduced a novel instrument for studying human development. We highlighted different features of the method, including finding words that distinguish groups based on a characteristic (e.g., age, gender); patterns of word use as a function of age, cohort, or time; and data-driven topics. We descriptively reviewed some of the most prominent results, and our comprehensive lists of words and categories and an interactive graph for plotting words as a function of age are available on our website for deeper exploration by the research community. The tool can be used both for exploratory analyses to discover unexpected variations for different age cohorts within different subgroups of the population, as well as to test or better characterize specific theories. We provided one example with the aging and emotion positivity effect, but we hope that other researchers will bring their own hypotheses to the data and test specific research questions.

While this study focused primarily on common trends across age, the same methodology can be used to explore heterogeneity within a developmental period, or to explore characteristics beyond age that differentiate groups of people. Many characteristics influence word use in social media, including age (as we found here), personality (Kern et al., 2013), gender, socioeconomic status, cognitive differences, and culture. Educational opportunities or social experiences, for example, may influence the development of interests, values, or motivation, which in turn may be expressed through language. Coupling our methodology with carefully constructed comparison groups could reveal differences that are not fully captured using traditional approaches.

Categories can provide a meaningful organizational structure for language. For example, when we see that young adults frequently mention “laundry”, we can think of this word as an indicator of a broader category of “housework”. Such categories can be manually developed from theories and understanding of development, or we can automatically distinguish clusters. Complementing top-down approaches that group words into conceptual categories (e.g., the LIWC dictionaries; Pennebaker & Francis, 1999), our approach allows categories to arise from the data. In essence, there is an implicit lexicon present in social media, and our method captures pieces of that lexicon.

To understand within-person variability and the influence of natural environments and contexts, intensive momentary assessments of thoughts and feelings are needed (Bolger & Laurenceau, 2013; Hoppmann & Riediger, 2009). Momentary reports often can be quite different than the remembered self that is typically assessed in questionnaires (Conner & Barrett, 2012). Facebook status updates are designed to be a self-descriptive text modality that elicits affective content, at the very time that the thought occurs (Kramer, 2010). Social media essentially enables in-the-moment responses at a larger level than ever before (Kietzmann, Hermkens, McCarthy, & Silverstre, 2011).

In this study, it is important to note that we presented cross-sectional comparisons across different age cohorts. The differences in the use of emotion might be due to cohort-related differences rather than to age differences per se. Language changes, and words go in and out of favor over time, as new interests and activities occur. For example, the word "fail" became popular online for a certain demographic within the last five years or so, but it has now gone out of favor, either from overuse or because it is used by a broader demographic. With cross-sectional data, it is impossible to distinguish cohort, time, and developmental effects (Donaldson & Horn, 1992). In building our method, we collapsed words across all times that a user posted, but a next step is to consider longitudinal and dynamic patterns over time. Future research should examine age-related trends longitudinally. Given that social media sources such as Facebook and Twitter include message time stamps, users' written expressions in social media represent an expanding longitudinal dataset of large parts of the population who are growing up and growing older online.

In line with prior studies on word use and individual characteristics (e.g., Fast & Funder, 2008; Pennebaker & Stone, 1999), we limited the current presentation to English speakers. As the myPersonality application presents personality tests in English, most of the participants were primarily English speaking. However, the differential language analysis approach is not limited to English. Whereas closed vocabulary approaches such as LIWC require careful translation, one advantage of using an open vocabulary approach is that translation is unnecessary. Some languages may be more challenging to work with, but words distinguishing user characteristics can be determined, as long as sufficient data are available.

Massive social media data can be used to test psychological theories in alternative contexts. For example, we found some support for the aging positivity effect using single words, such that negative affect words declined with age, high arousal positive affect declined, and low arousal positive affect increased. Theoretically generated categories such as the LIWC positive and negative emotion categories supported these trends, but only positive and negative valence, not high versus low arousal, could be distinguished. We did not find clear emotion topics in the automatically generated topics. This may be an artifact of the clustering, or it may be that single words are more informative than categories for emotions. For example, Grühn and colleagues (2010) examined discrete emotions across the lifespan (age 18 to 78) and found that fear, hostility, guilty, sadness, self-assurance, shyness, fatigue, and surprise linearly declined; positive affect, joviality, serenity, and surprise followed a u-shaped pattern. In a second study, across multiple cultures, aging related to less anger, sadness, and fear, and increased happiness and emotional control (Gross et al., 1997). Our method can allow such distinctions to be replicated with many more observations.

The focus on big data does not imply that small studies following a group of individuals over time lack importance. To the contrary, the carefully designed, prospective studies often used by developmental psychologists can help distinguish cohort-related versus developmental effects, and allow a better understanding of long-term processes. For example, teenagers were especially likely to use emoticons (e.g., ":)", "<3", ":p") and net speak (e.g., "lol", "tmrw", "jk"); this could reflect certain characteristics of youth, or may be a cohort related effect. There may be educational and socioeconomic status (SES) differences in word use, although recent research by the Pew Research Center finds that social media use is spread fairly evenly across different SES and educational groups (Brenner, 2012). In our sample, we were unable to test word differences in older age, as only 82 individuals were age 65 or older. As the population matures and becomes increasingly connected online, further consideration of how big data fit within the developmental and aging literature are warranted. In addition, although a growing percentage of the population has used some form of social media at some point, individuals vary in the information they are willing to share online (Karl, Peluchette, & Schlaegel, 2010). Especially as online privacy concerns increase (TRUSTe U.S. Consumer Confidence Index, 2013), future research will need to consider biases that any online sample entails. Whereas the tools from computer science can help make sense of data, developmental and social psychologists can play an important role in noting the limitations of any particular dataset.

In conclusion, this study adds a tool into the developmental methodology toolbox. Our method is meant to complement, not replace, existing developmental methods. Using only a hammer and nails, one might build a structure that stands, but only by using a suite of tools does this structure become a house. Likewise, each design and statistical method has its own strengths and limitations, by creatively combining findings and methods across studies, the full structure of development can emerge.

References

Anscombe, F. J. (1948). The transformation of poisson, binomial and negative-binomial data. *Biometrika, 35*, 246-254. Doi: 10.2307/2332343

Baltes, P. B., Reese, H. W., & Lipsitt, L. P. (1980). Life-span developmental psychology. *Annual Review of Psychology, 31*, 65-110. Doi: 10.1146/annurev.ps.31.020180.000433

Baltes, P. B., & Smith, J. (2004). Lifespan psychology: From developmental contextualism to developmental biocultural co-constructivism. *Research in Human Development, 1*, 123-144. Doi: 10.1207/s15427617rhd0103_1

Biss, R. K., & Hasher, L. (2012). Happy as a lark: Morning-type younger and older adults are higher in positive affect. *Emotion, 12*, 437-441. Doi: 10.1037/a0027071

Blei, D. M., Ng., A. Y., & Jordan, M. I. (2003). Latent direichlet allocation. *Journal of Machine Learning Research, 3*, 993. http://jmlr.org/papers/volume3/blei03a/blei03a.pdf

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York: Guilford Press.

Brenner, J. (2012). *Pew internet: Social networking (full detail). Pew Research Center.* Retrieved from http://pewinternet.org/Commentary/2012/March/Pew-Internet-Social-Networking-full-detail.aspx

Carstensen, L. L., & Mikels, J. A. (2005). At the intersection of emotion and cognition: Aging and the positivity effect. *Current Directions in Psychological Science, 14*, 117-121. Doi: 10.1111/j.0963-7214.2005.00348.x

Carstensen, L. L., Pasupathi, M., Mayr, U., & Nesselroade, J. R. (2000). Emotional experience in everyday life across the adult life span. *Journal of Personality and Social Psychology, 79*, 644-655. Doi: 10.1037/0022-3514.79.4.644

Carstensen, L. L., Turan, B., Scheibe, S., Ram, N., Ersner-Hershfield, H., Samanez-Larkin, G. R., Brooks, K. P., & Nesselroade, J. R. (2011). Emotional experience improves with age: Evidence based on over 10 years of experience sampling. *Psychology and Aging, 26*, 21-33. Doi: 10.1037/a0021285

Charles, S. T., Reynolds, C. A., & Gatz, M. (2001). Age-related differences and change in positive and negative affect over 23 years. *Journal of Personality and Social Psychology, 80*, 136-151. Doi: 10.1037/0022-3514.80.1.136

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computer Linguistics, 16*, 22–29. http://acl.ldc.upenn.edu/J/J90/J90-1003.pdf

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*, 829-836. Doi: 10.1080/01621459.1979.10481038

Clogg, C.C. (1995). Latent class models. In G. Arminger, C.C. Clogg & M.E. Sobel (eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311-359). New York: Plenum Press.

Conner, T. S., & Barrett L. F. (2012). Trends in ambulatory self-report: The role of momentary experience in psychosomatic medicine. *Psychosomatic Medicine, 74*, 327-337. Doi: 10.1097/PSY.0b013e3182546f18

Diehl, M., Hay, E. L., & Berg, K. M. (2011). The ratio between positive and negative affect and flourishing mental health across adulthood. *Aging & Mental Health, 15*, 882-893. Doi: 10.1080/13607863.2011.569488

Donaldson, G., & Horn, J. L. (1992). Age, cohort, and time developmental muddles: Easy in practice, hard in theory. *Experimental Aging Research, 18*, 213-222. Doi: 10.1080/03610739208260360

Facebook.com (2012). Fact sheet. Retrieved from http://newsroom.fb.com/content/default.aspx?NewsAreaId=22

Fernandez-Ballesteros, R., Fernandez, V., Cobo, L., Caprara, G., & Botella, J. (2010). Do inferences about age differences in emotional experience depend on the parameters analyzed? *Journal of Happiness Studies, 11*, 517-521. Doi: 10.1007/s10902-009-9169-y

Garry, J., & Lohan, M. (2011). Mispredicting happiness across the adult lifespan: Implications for the risky health behaviour of young people. *Journal of Happiness Studies, 12*, 41-49. Doi: 10.1007/s10902-009-9174-1

Griffin, P. W., Mroczek, D. K., & Spiro, A. III. (2006). Variability in affective change among aging men: Longitudinal findings from the VA Normative Aging Study. *Journal of Research in Personality, 40*, 942-965. Doi: 10.1016/j.jrp.2005.09.011

Gross, J. J., Carstensen, L. L., Tsai, J., Skorpen, C. G., & Hsu, A. Y. C. (1997). Emotion and aging: Experience, expression, and control. *Psychology and Aging, 12*, 590-599. Doi: 10.1037/0882-7974.12.4.590

Grüehn, D., Kotter-Grüehn, D., & Röcke, C. (2010). Discrete affects across the adult lifespan: Evidence for multidimensionality and multidirectionality of affective experiences in young, middle-aged and older adults. *Journal of Research in Personality, 44*, 492-500. Doi: 10.1016/j.jrp.2010.06.003

Havighurst, R. J. (1972). *Developmental tasks and education* (3rd ed.). New York: McKay.

Hoppmann, C. A., & Riediger, M. (2009). Ambulatory assessment in lifespan psychology: An overview of current status and new trends. *European Psychologist, 14*, 98-108. Doi: 10.1027/1016-9040.14.2.98

Huelsman, T. J., Nemanick, R. C. Jr., & Munz, D. C. (1998). Scales to measure four dimensions of dispositional mood: Positive energy, tiredness, negative activation, and relaxation. *Educational and Psychological Measurement, 58,* 804-819. Doi: 10.1177/0013164498058005006

Isaacowitz, D. M., & Blanchard-Fields F. (2012). Linking process and outcome in the study of emotion and aging. *Perspectives on Psychological Science, 7*, 3-17. Doi: 10.1177/1745691611424750

Karl, K., Peluchette, J., & Schlaegel, C. (2010). Who's posting Facebook faux pas? A cross-cultural examination of personality differences. *International Journal of Selection and Assessment, 18*, 174–186. Doi: 10.1111/j.1468-2389.2010.00499.x

Kessler, E.-M., & Staudinger, U. M. (2009). Affective experience in adulthood and old age: The role of affective arousal and perceived affect regulation. *Psychology and Aging, 24,* 349-362. Doi: 10.1037/a0015352

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons 54*, 241-251. Doi: 10.1016/j.bushor.2011.01.005

Kosinski, M. & Stillwell, D. J. (2011). myPersonality Research Wiki. *myPersonality Project.* Retrieved from http://mypersonality.org/wiki

Kramer, A. D. I. (2010, April). *An unobtrusive behavioral model of "gross national happiness".* CHI '10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA. Retrieved from http://dmrussell.net/CHI2010/docs/p287.pdf

Kunzmann, U., Little, T. D., & Smith, J. (2000). Is age-related stability of subjective well-being a paradox? Cross-sectional and longitudinal evidence from the Berlin Aging Study. *Psychology and Aging, 15*, 511-526. Doi: 10.1037/0882-7974.15.3.511

Lawton, M. P. (2001). Emotion in later life. *Current Directions in Psychological Science, 10,* 120–123. Doi: 10.1111/1467-8721.00130

Lin, D. (1998, August). *Extracting collocations from text corpora*. First Workshop on Computational Terminology, Montreal, Canada. Retrieved from www-rohan.sdsu.edu/~gawron/mt_plus/readings/sim_readings/collocations_lin_98.pdf

Mroczek, D. K. (2001). Age and emotion in adulthood. *Current Directions in Psychological Science, 10*, 87-90. Doi: 10.1111/1467-8721.00122

National Science Foundation (2012). Core techniques and technologies for advancing big data science & engineering. National Science Foundation (Solicitation #12-499). Retrieved from: www.nsf.gov/funding/pgm_summ.jsp?pims_id=504767

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6,* 101-115. Doi: 10.1037/1089-2699.6.1.101

Pennebaker, J. W., & Francis, M. E. (1999). *Linguistic Inquiry and Word Count: LIWC*. Mahwah, NJ: Erlbaum.

Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the life span. *Personality Processes and Individual Differences, 85*, 291-301. Doi: 10.1037/0022-3514.85.2.291

Pinquart, M. (2001). Age differences in perceived positive affect, negative affect, and affect balance*. Journal of Happiness Studies, 2,* 375–405. Doi: 10.1023/A:1013938001116

Scheibe, S., & Carstensen, L. L. (2010). Emotional aging: Recent findings and future trends. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences, 65*, 135-144. Doi: 10.1093/geronb/gbp132

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., …, & Ungar, L. H. (in press). Personality, gender, and age in the language of social media: The open vocabulary approach. *PLOS ONE.*

Stone, A. A., Schwartz, J. E., Broderick, J. E., & Deaton, A. (2010). A snapshot of the age distribution of psychological well-being in the United States. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 9985-9990. Doi: 10.1073/pnas.1003744107

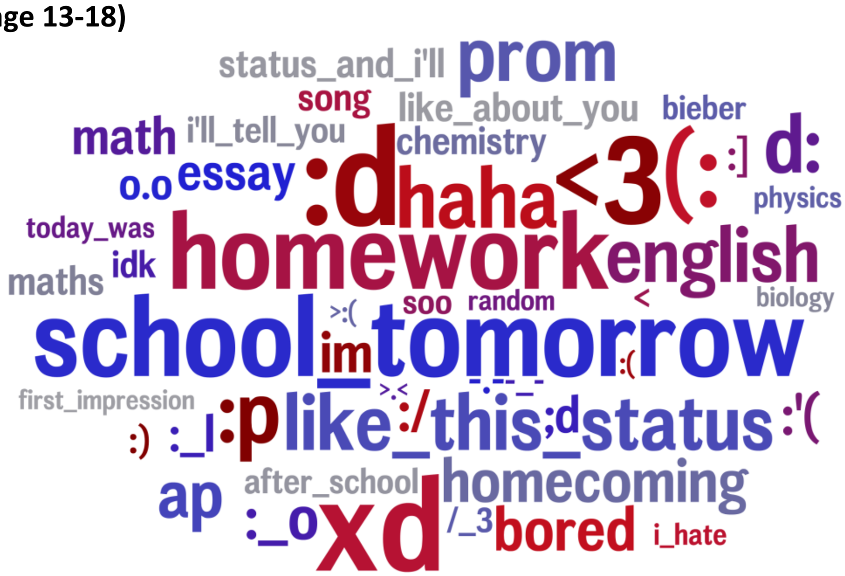
TRUSTe US Consumer Confidence Index. (2013). Retrieved from http://www.truste.com/us-consumer-confidence-index-2013/

U.S. Census Bureau FactFinder (2012). Median age at first marriage. Retrieved from http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_10_5YR_B12007&prodType=table

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology, 54,* 1063-1070. Doi: 10.1037/0022-3514.54.6.1063

Figure 1. The most common words used by teenagers (age 13-18) and young adults (age 23-29). Words are based on the strongest correlations between words/phrases and the age category, adjusted for gender. The size of the word or phrase indicates the strength of correlation (larger = stronger) and color indicates how frequently the word or phrase appears across user posts (black = frequent, gray = less frequent). Underscores (_) are used to connect multiword phrases; these characters are not present in the original text. See http://wwbp.org/v2/age-wc.html for the other age categories.

a) Teenagers (age 13-18)



b) Young adults (age 23-29)

Figure 2. Single word patterns as expressed across the range of ages.

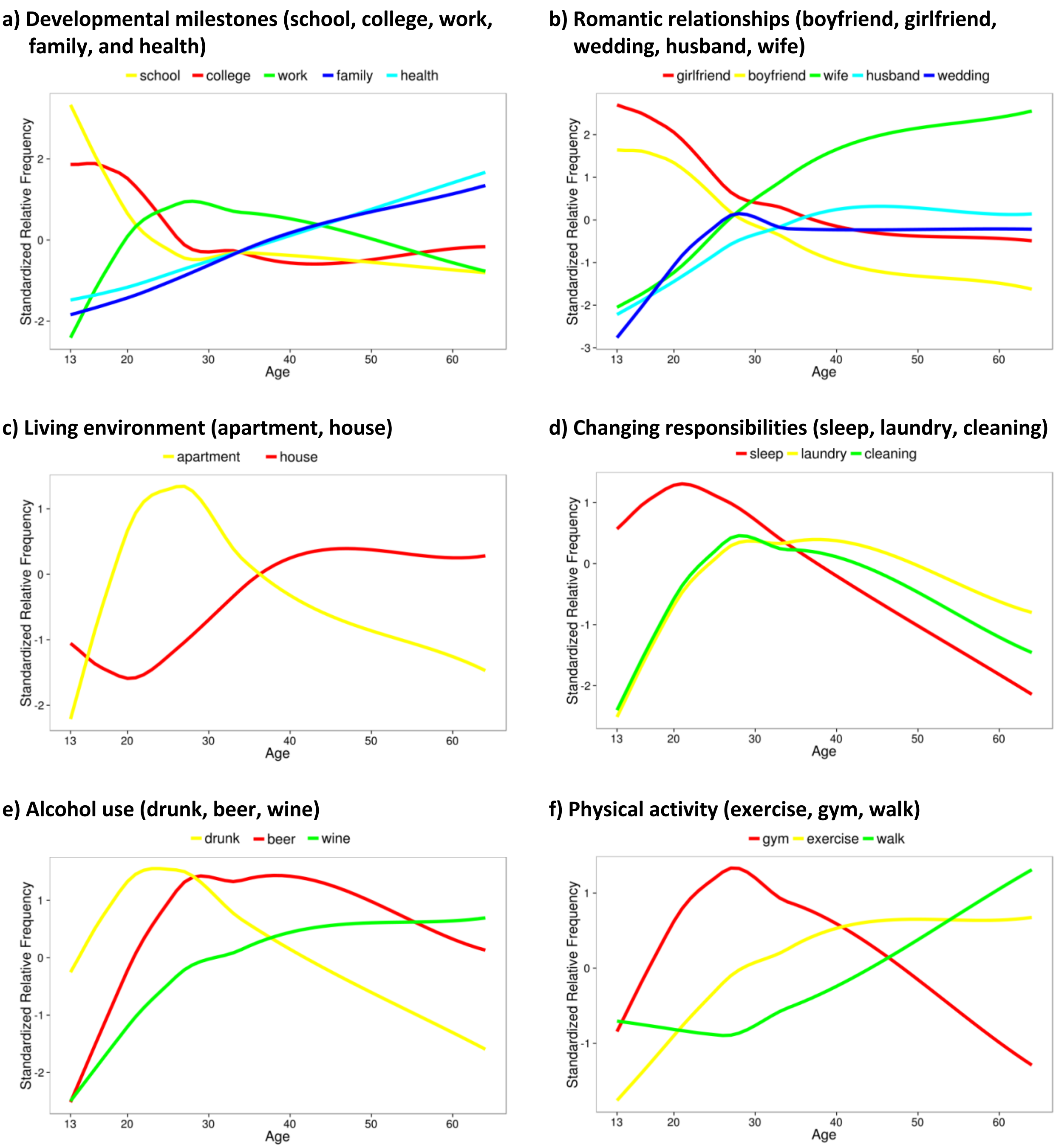


Figure 3. Four of the strongest topics for young adults (age 23-29) and middle-aged adults (age 45-64). See http://wwbp.org/v2/age-wc.html for the other three groups.

a) Young adults (age 23-29)

b) Middle-aged adults (age 45-64)

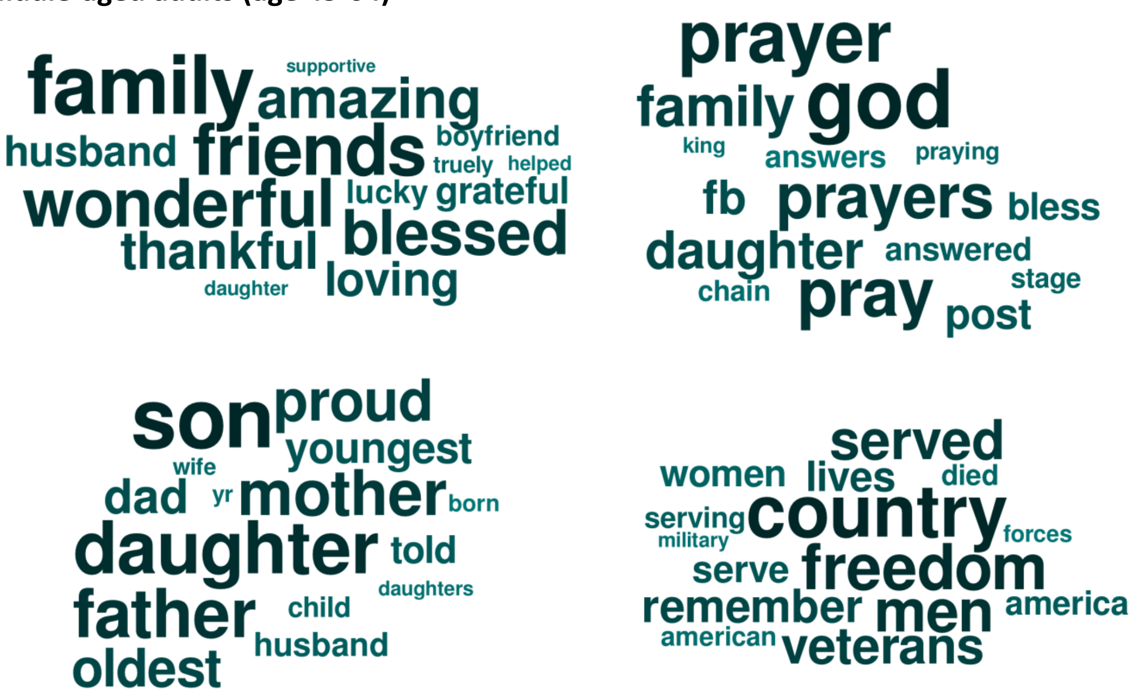


Figure 4. The dominant topic from each age group (listed from top to bottom by age: 13-18, 19-22, 23-29, 30-44, and 45-64) as a time series of occurrence across the age spectrum. The strongest words comprising each topic are listed.

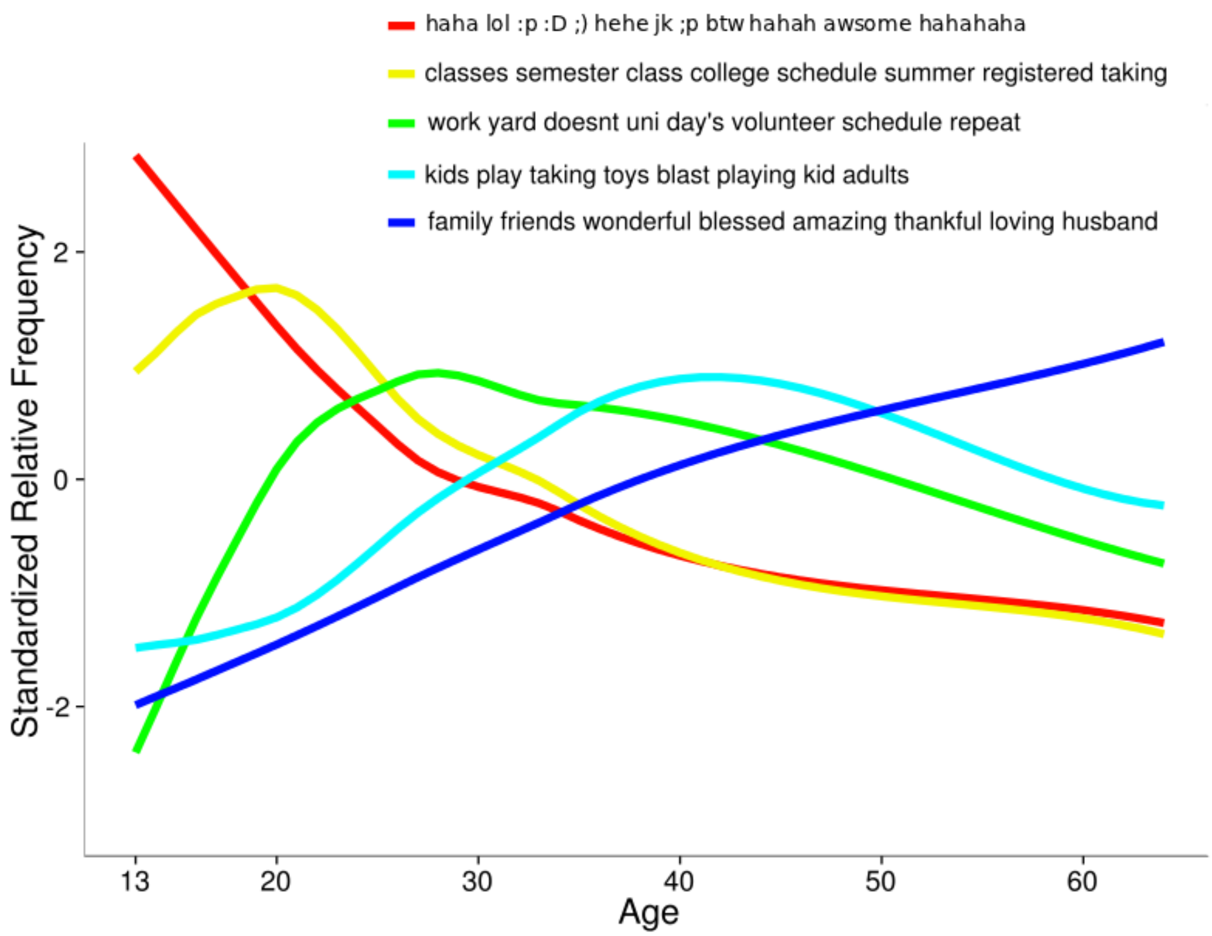


Figure 5. Occurrence of LIWC categories as a function of age. Figure A replicates age related findings related to positive emotion (posemo), negative emotion (negemo), first person pronouns (I), and future tense words (future) by Pennebaker and Stone (2003). Figure B tests two additional LIWC categories that conceptually align with our dominant topics: work and family.

a) b)

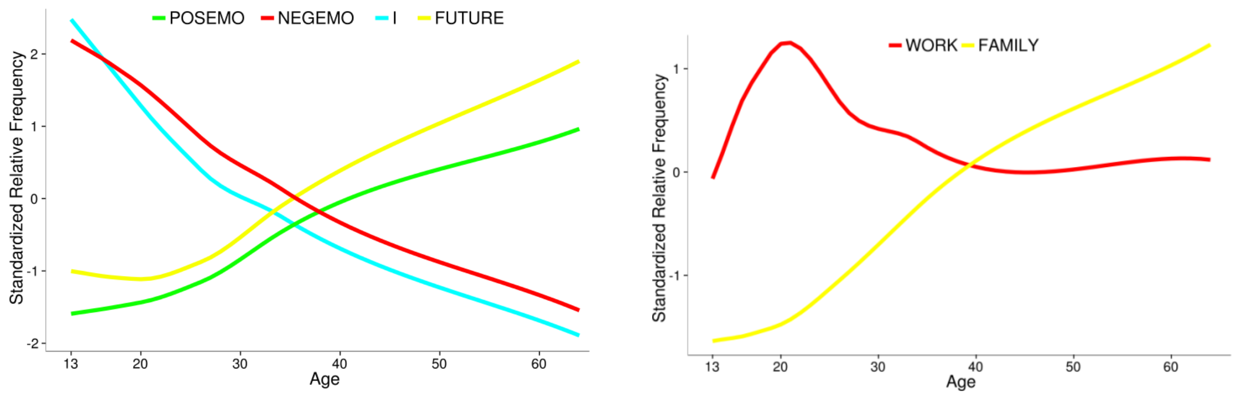


Figure 6. Words and phrases as a function of both age and gender. The 500 words/phrases most correlated with each age group were selected, and then sorted by their correlations with gender. The 50 features most positively and negatively correlated with gender were plotted as a word cloud. Size reflects the absolute size of the gender correlation (larger = stronger correlation with gender).

Figure 7. Testing the aging positivity effect. Low and high arousal positive and negative emotion words, plotted as a time series as a function of age.