Linguistic Distancing on Social Media: Indicators of Emotion Regulation Across Age Groups

Daniela Teodorescu[α*], **Saif M. Mohammad**[†], **Alona Fyshe**[α*‡]
[α]Alberta Machine Intelligence Institute, [*]Department of Computing Science, University of Alberta
[†]National Research Council Canada,
[‡]Department of Psychology, University of Alberta
{dteodore,alona}@ualberta.ca, saif.mohammad@nrc-cnrc.gc.ca

Abstract

Managing our emotional responses to events is key to emotional well-being, a process referred to as emotion regulation in psychology. Previous work has established that the degree to which we distance events is a type of emotion regulation. When we psychologically distance from events there can be markers in our language. These markers have been referred to as linguistic distancing. We build upon a previous metric to operationalize linguistic distancing, and explore how it changes across the lifespan. We explore this systematically by analyzing large amounts of social media text, a venue where people express their emotions. By investigating how distancing varies across age groups we can better understand how emotion regulation varies with age and provide initial benchmarks on social media data. We provide additional evidence further strengthening the hypothesis that linguistic distancing occurs in proportionally more instances with age. These findings align with past work in psychology which indicate improved well-being with older age. Better understanding how linguistic distancing changes with age is important because it functions as a marker of well-being and can inform effective health interventions. We provide a foundation for further exploring emotion regulation through linguistic distancing in text data.

Keywords: Linguistic Distancing, Age, Emotion Regulation, Social Media

1. Introduction

Our everyday emotional experiences are not static but rather dynamic and our emotions are constantly changing over time. The way in which our emotional experiences change over time creates an emotional trajectory, or what some refer to as an *emotion arc* (Mohammad, 2011; Reagan et al., 2016). Emotion regulation includes the processes by which "we influence which emotions we have, when we have them, and how we experience and express them" (Gross, 1998). More simply put, emotion regulation is the way in which we manage our emotions.

Emotion regulation is key to overall well-being; dysregulation of emotions is associated with psychopathologies such as major depressive disorder and social anxiety disorder (Aldao et al., 2010). There are various ways (or strategies) of regulating emotions and some of these are seen as *maladaptive*. While such strategies provide momentary relief from distress, they are not effective in the long term. Maladaptive strategies include rumination (rethinking about memories or negative experiences), avoidance (of a situation or emotional experience), and suppression (of emotions or thoughts), and each of these has been shown to be significantly associated with anxiety, eating disorders, and substance abuse (Aldao et al., 2010). On the other hand, some strategies are considered *adaptive* and are seen as healthy ways of managing emotions, e.g., reappraisal (changing how one thinks about a situation to change emotional reactions), problem solving (taking actions to solve a problem), and acceptance (non-judgmental acceptance of emotions). Adaptive strategies are seen as aiding against psychopathology and use of such strategies are inversely associated with psychopathology (Aldao et al., 2010).

Given the importance of emotion regulation to well-being, understanding how it changes with age is important for providing appropriate support and interventions. The specific emotion regulation strategies used change throughout childhood, adolescence, young adulthood, and adulthood (Nook et al., 2020). For example, a shift has been found from emotion regulation strategies which deal with a situation behaviorally (e.g., escaping a situation), to more cognitive strategies such as seeking information (Brown et al., 1991; Altshuler et al., 1995). Although there are some conflicting findings on how exactly various strategies track with age, many studies point to older ages having "enhanced emotion regulation" (Urry and Gross, 2010), being better at regulating their emotions (Gross et al., 1997; Mikkelsen et al., 2024), and have better overall well-being than younger adults despite losses physically and cognitively (Urry and Gross, 2010; Stawski et al., 2008).

Another lens through which emotion regulation has been studied is psychological distancing. By creating distance from an event, one can better handle their emotions related to the event and view the event objectively (allowing for a reappraisal of the

Proceedings of Computational Affective Science (CAS) @ LREC 2026, pages 151–163
16 May 2026.

event). Adopting a distanced perspective has been shown to down-regulate negative affect (Kross and Ayduk, 2017), and it is a common technique used in Cognitive-Behavioral Therapy (Beck and Beck, 2020) and Dialectical Behavior Therapy (Linehan, 1993). When psychologically distancing from an event, one is viewing the event from a third-person perspective, and often psychological distancing appears in language as markers. For example, when distancing, there are less first-person pronouns and more past and future tense verbs rather than present tense verbs. These markers are described via *linguistic distancing* (Nook et al., 2017). Linguistic distancing has been shown to be associated with successful emotion regulation and reduced negative affect (Nook et al., 2017). Further, in client-therapist transcripts, less linguistic distancing was associated with worse internalizing symptoms (inward focused behaviors often occurring with anxiety and depression) (Nook et al., 2022).

Through linguistic distancing we can study emotion regulation in text; although so far only client-therapist transcripts have been analyzed for select age groups. While linguistic distancing has been explored across childhood, adolescence, and young adults (Cohen et al., 2022; Nook et al., 2020), we do not know how it changes over adulthood which is what we address in our work. While one approach to measuring linguistic distancing across ages is through longitudinal research, that is very difficult as it would require text data from the same people across all the decades. Instead, we conduct cross-sectional research, which has its own benefits, but appropriate conclusions must be drawn that are different from what could be drawn from longitudinal research. Additionally, linguistic distancing could be influenced by the era in which people grew up; e.g., those in a certain age group may use more distancing due to world events, culture, social norms, etc. than other generations. Therefore, we explore how linguistic distancing changes in *contemporary* times using a platform which many of use to communicate: social media.

Social media is a platform where we frequently communicate and express our feelings. Platforms such as Reddit and X provide a space where we can freely express our thoughts at any time of day, connect with others, and possibly receive support. Given the important role social media plays in everyday communication, as researchers we make use of large amounts of data to analyze linguistic distancing in a systematic way across adulthood. We explore how linguistic distancing varies across age groups on social media. We make use of a social media dataset annotated with the age of the author at the time of posting to answer the following research questions:

- How does linguistic distancing vary across age groups?

  We hypothesize that linguistic distancing increases with age as people become better at regulating emotions based on literature in psychology.
- How do the various *dimensions* of linguistic distancing (i.e., social distancing, temporal distancing, passive voice and abstractness) vary across age groups? Do some dimensions change more than others across adulthood?

By answering these questions, we provide important findings on how linguistic distancing tracks across age groups on social media. Through linguistic distancing, our findings allow us to better understand emotion regulation on a larger scale across age, enabling further study of this phenomenon in psychology and the social sciences. We make our code publicly available.[1]

2. Related Work

Below we describe past work examining the relationship between emotion regulation, linguistic distancing and how they change across the lifespan. Afterwards, we describe markers of distancing in language.

2.1. Emotion Regulation, Linguistic Distancing & Age

Appropriately regulating emotions is key to mental health and well-being. Vast amounts of literature point to dysregulation of emotions being tied with psychopathology such as major depressive disorder and anxiety disorders (Gross and Muñoz, 1995; Sheppes et al., 2015; Aldao et al., 2010). The various ways in which we manage our emotions are related with health outcomes: adaptive strategies such as cognitive reappraisal are significantly associated with positive indicators of mental health, whereas maladaptive strategies such as suppression of emotions are significantly associated with negative indicators of mental health (Hu et al., 2014). Emotion regulation also changes throughout the lifespan. Infants have minimal to no control over their emotions, relying on their caregiver for support (Bowlby, 1969). Infants begin to develop approaches such as changing their eye gaze to focus on other objects as a way of signaling autonomous emotion regulation (Derryberry and Rothbart, 1988). As they become toddlers, they begin to understand language which opens a whole new world in terms of emotion regulation. Now, they can begin to understand instructions from their parents (Thompson, 1991) and they can

[1] https://github.com/dteodore/Age--Linguistic_Distancing

begin to express how they feel and socialize with others, learning to differentiate between positive and negative emotions (Gross and Muñoz, 1995). In adolescence, more sophisticated strategies develop, such as reappraisal as well as maladaptive strategies (Gross and Muñoz, 1995; Hall et al., 1993; Félix-Ortiz et al., 1995). While overall findings point to older age being associated with better overall well-being, control of emotions, and enhanced emotion regulation (Urry and Gross, 2010; Gross et al., 1997), there are some discrepancies in exactly how emotion regulation changes with age, such as which strategies do we use more vs. less. For example, some work found *suppression* decreased across the life span (John and Gross, 2004; De France and Hollenstein, 2019), whereas others found *suppression* increased with age (Brummer et al., 2014). De France and Hollenstein (2019) found adolescents used *distraction* and *suppression* more than young adults, and young adults used more *rumination*, whereas Sütterlin et al. (2012) found no differences in *rumination* across ages. John and Gross (2004) found *reappraisal* increased across the lifespan, whereas De France and Hollenstein (2019) found no difference. Given the conflicting findings, we use social media data to examine how *reappraisal* changes across the lifespan.

We specifically study reappraisal through linguistic distancing. Linguistic distancing has been studied across children, adolescents, and young adults when performing a reappraisal task, however no differences across ages was found (Nook et al., 2020). Although there were changes in the exact strategies of cognitive reappraisal employed across ages: *changing circumstances* and *distancing* increased across age; *changing consequences* decreased across age; and adolescences used more *challenging reality* and less *problem-solving* compared to other ages. Linguistic distancing has also been studied specifically in adolescents: more linguistic distancing was associated with lower levels of hopelessness and higher levels of perceived agency, whereas greater use of linguistic distancing predicted fewer depressive symptoms in follow-ups (Cohen et al., 2022). While these studies provide findings for ages 10–23 (Nook et al., 2020) and 13–16 (Cohen et al., 2022), we are interested in changes across a vast range of ages i.e., 13–70's. Further, we investigate the natural use of linguistic distancing as it appears in everyday communication (e.g., social media), whereas past findings in psychology were based on self-reports and questionnaires (e.g., Emotion Regulation Questionnaire (ERQ) (Gross and John, 2003), Cognitive Emotion Regulation Questionnaire (CERQ) (Garnefski and Kraaij, 2007)) or reappraisal tasks. While past work has created a corpus for detecting coping strategies in text (Troiano et al., 2024), which is very intertwined with emotion regulation (coping strategies take direct action to manage distress), we study naturally occurring text (rather than text from role-playing scenarios) and are interested in this phenomenon across ages.

2.2. Distancing in Language

Distancing can occur through various dimensions, such as in terms of time, space and socially (Construal Level Theory (Liberman and Trope, 2008; Trope and Liberman, 2010)). Therefore, past operationalizations of linguistic distancing include measures of *social distancing* and *temporal distancing*. Use of more first-person singular pronouns e.g., "I", "me", "mine" represents a more immersed perspective and less *social* distancing, which means that one is performing less psychological distancing. On the other hand, using more second- and third-person pronouns e.g., "she", "they", "them" represents using more *socially* distanced language, and therefore one is doing more psychological distancing. Using more present tense verbs represents less *temporal distancing* and less psychological distancing, whereas using more past- and future-tense signals more *temporal* distancing and therefore more psychological distancing (Nook et al., 2022). We use these definitions of *social* and *temporal* distancing in our work. The bidirectional relationship between emotion regulation and linguistic signatures of psychological distancing has been shown (Nook et al., 2017). Regulating emotions through psychological distancing has been associated with increased linguistic markers of social and temporal distance. Likewise, using distanced language also regulated emotions and reduced negative affect (Nook et al., 2017). Self-distancing ("I" vs. "you" pronouns) and social withdrawal (indicated by more "they" pronouns), has been explored on Reddit for healthcare workers during the COVID-19 pandemic, however we are interested in linguistic distancing on a broader scale (i.e., for those beyond the healthcare field) and in day-to-day contexts outside of the pandemic (Ireland et al., 2022).

When distancing from an event psychologically, this creates a more abstract mental representation of the event – leading to a high correlation between linguistic measures of psychological distancing and *abstractness* (Nook et al., 2025). Measures of linguistic distancing and *abstractness* have been found to be highly correlated; *abstractness* increased when cognitive reappraisal was used to regulate emotions; when people distanced their language, *abstractness* also increased; and lastly, increased *abstractness* when regulating emotions was correlated with regulation success (Nook et al., 2025). Given the established relationship between *abstractness* and linguistic distancing, we also in-

Age Group	#Posts	
	TUSC-City	Reddit
13–19	94,857	9,281,055
20–29	406,027	15,455,426
30–39	462,238	6,151,757
40–49	360,946	1,220,498
50–59	278,809	449,230
60–69	218,008	246,398
70–79	97,432	160,465

Table 1: The number of posts across the age groups in each subset of the *AgeCorpus*.

corporate a measure of *abstractness* in our metric of linguistic distancing.

Work in psychology has also pointed to the relationship between using passive voice and psychological distancing (Chan and Maglio, 2020a,b). *Passive voice* is where the object is placed before the verb such that the subject receives the action rather than performing it (e.g., passive voice: "The ball was thrown by the child" vs. active voice: "The child threw the ball").[2] When reading passages in passive voice participants reported feeling temporal, hypothetical, and spatial distance from the events in the passage resulting in a more abstract mental construal (Chan and Maglio, 2020b; Trope and Liberman, 2003). Past work has also studied passive voice in the context of assault, where distancing removed the action from the actor and places blame on the victim (Bohner, 2001). When using passive voice, the agency of the subject is taken away, creating distance from the action.

Given the past literature demonstrating the relationship between linguistic distancing with *social* distancing, *temporal* distancing, *abstractness*, and *passive voice*, we construct a measure of linguistic distancing which includes all four dimensions. Analyzing how each of these dimensions changes across adulthood is informative for understanding behaviors and tendencies which is of interest not only to psychologists, linguists, and sociologists, but is also linked with mental health.

3. Dataset: AgeCorpus

We perform our experiments on *AgeCorpus*, a social media dataset containing posts annotated with the author's age at the time of writing (Teodorescu et al., 2026). The dataset contains posts from both Reddit and X, making it a suitable dataset for exploring linguistic distance on social media platforms. Further, there are a large number of posts in the dataset, allowing us to aggregate results per age group. The dataset was collected by identifying instances where authors self-disclosed their age e.g., "I am X years old" or "Me [20F] and my best friend [21M]". Using various high quality pattern matching templates, the age of an author can be determined at a point in time based on the timestamp of the post (seed post). We show the pattern templates used in Appendix A. Using the age declared at the given timestamp in the seed post, the author's age can be determined for any other post based on the difference of the timestamp to that of the seed post's. The dataset consists of thousands of users each with hundreds of posts spanning from 2010–2022 for Reddit and 2020–2021 for X. We show the exact number of posts per age group in Table 1. The X component of the dataset consists of *TUSC-Country* and *TUSC-City* which differ in how they were collected. TUSC-City was collected using Twitter's free API for American and Canadian cities during 2020–2021. Whereas TUSC-Country was collected using Twitter's Academic API for tweets from the US and Canada during 2015–2021. Our results on the two subsets are very similar so we report those on TUSC-City since it is the larger subset and show the results on TUSC-City in the Appendix (Section C). We use this dataset for our experiments described in the next section.

[2] https://www.grammarly.com/blog/grammar/passive-voice/

4. Experiments

In the sections below we describe our methods for exploring how linguistic distancing varies across age groups.

4.1. Computing Linguistic Distancing in Text

We build on Nook et al. (2022)'s work on computing linguistic distancing in text. Specifically, the authors operationalize linguistic distancing as being composed of two components: *social* distancing and *temporal* distancing. *Social* distancing is measured through the proportions of pronouns present in the text. Fewer first-person singular pronouns are associated with proportionally more instances of distancing. Here, we analyze the proportions of pronouns present in an instance or a post. More specifically, *social* distancing is computed per post. Let $f_{1\text{p,sg}}, f_{1\text{p,pl}}, f_{2\text{p}}, f_{3\text{p,sg}}, f_{3\text{p,pl}}$ denote the corresponding frequencies of first-person singular, first-person plural, second-person, third-person singular and third-person plural pronouns respectively. Then the social distancing ($social$) component of linguistic distancing (LD) per post or instance (t) is computed as:

$$LD_{social}(t) = \frac{f_{2\text{p}} + f_{1\text{p,pl}} + f_{3\text{p,sg}} + f_{3\text{p,pl}}}{f_{2\text{p}} + f_{1\text{p,pl}} + f_{3\text{p,sg}} + f_{3\text{p,pl}} + f_{1\text{p,sg}}}$$

Temporal distancing is captured through the use of verb tense. More past and future tense verbs

Dataset	Metric	df1	df2	F-statistic	P-value	Effect Size (*est* ω^2)
Reddit	Linguistic	6	1203317.10	17703.63	$p<.001$	0.003
TUSC-City	Linguistic	6	539267.68	2830.26	$p<.001$	0.009

Table 2: The degrees of freedom (for the numerator and denominator), F-statistic, p-value, and effect size in the one-way Welch's ANOVA test for differences in linguistic distancing between age groups.

are associated with proportionally more instances of distancing. We compute *temporal* distancing per post. Let $f_{\text{past}}, f_{\text{present}}, f_{\text{future}}$ denote the corresponding frequencies of past-tense, present-tense and future-tense verbs. Then the temporal distancing ($temp$) component of linguistic distancing (LD) per post or instance (t) is computed as:

$$LD_{temp}(t) = \frac{f_{\text{past}} + f_{\text{future}}}{f_{\text{past}} + f_{\text{future}} + f_{\text{present}}}$$

We further build on the metric for linguistic distancing by including a measure of *abstractness* and *passive voice*. We include *abstractness* in our measure of linguistic distancing as more abstract mental representations of an event are linked with more successful emotion regulation, and more abstract language is highly correlated with more distanced language (Nook et al., 2025). More details on this are described in Section 2. To quantify *abstractness* we use a word list with ratings of concreteness for forty-thousand words (Brysbaert et al., 2014).[3] Each word is rated on a 5-point scale from abstract to concrete which we invert to be consistent with other measures. We calculate the *abstractness* score (LD_{abs}) per post (t) as the average of the abstractness score of each word at index i ($abs(t_i)$) of the instance with length n:

$$LD_{abs}(t) = \frac{1}{n}\sum_{i=1}^{n} abs(t_i)$$

Lastly, we determine whether a text contains *passive voice* or not using the PassivePy tool (Sepehri et al., 2023). PassivePy uses POS tagging, dependency parsing and rule-based matching for sentence structures/forms in order to determine whether passive voice is present in the text or not. We integrate this binary score of passive voice ($pass(t)$) per instance (t) into our metric for linguistic distancing.

$$LD_{pass}(t) = pass(t)$$

Each of these measures are on different scales. Therefore, we standardize all the scores for each dimension individually such that each dimension has scores with a mean of 0 and a standard deviation of 1.[4] Then we compute the linguistic distancing (LD) score per post (t) as the average of the four standardized ($stand()$) dimensions:

$$LD(t) = \frac{\begin{array}{c} stand(LD_{temp}(t)) + stand(LD_{social}(t)) \\ + stand(LD_{abs}(t)) + stand(LD_{pass}(t)) \end{array}}{4}$$

Linguistic Distancing Per Age Group: The next step is to aggregate scores per age groups. For our analyses we consider the following age groups: 13–19 (teens), 20–29 (twenties), 30–39 (thirties), 40–49 (forties), 50–59 (fifties), 60–69 (sixties), 70–79 (seventies). We group instances based on the age of the author at the time of writing. Then we compute the linguistic distancing score per group as the average of all the scores of instances within the particular group.

4.2. Statistical Analyses

To compare whether there is a statistically significant difference in linguistic distancing measures across age groups, we perform a one-way ANOVA test. Before performing inferential tests, we ensure that our data met ANOVA's prerequisites. Namely, we tested that the data are independent of each other, each independent variable is approximately normally distributed, and each group has roughly the same variance (homoscedasticity). Next, we use a qq-plot of the residuals to verify that each score is normally distributed for each group. In addition to the normality assumption, we can assume the estimate of the mean is normally distributed due to the large sample size (law of large numbers) according to the central limit theorem. To verify whether each group has approximately the same variance, we performed Levene's test and the result was significant ($p < 0.001$) for Reddit and TUSC-City, meaning that the age groups do not have equal variance in the Reddit and TUSC-City subsets of the dataset. We can assume each data point is independent of others since each post is generally independent of one another. While it is possible that while browsing social media, a post may influence others to some degree, this is largely not the case. Therefore, we will perform a one-way Welch's ANOVA for TUSC-City and Reddit. Results for Levene's tests can be found in Appendix B. After we determined our data met the assumptions, we performed the one-way ANOVA.

[3] We opt for this method of determining abstractness since Nook et al. (2025) use the LIWC Lexicon (Pennebaker et al.), which is not freely available.

[4] Standardizing is the same as computing the z-score.

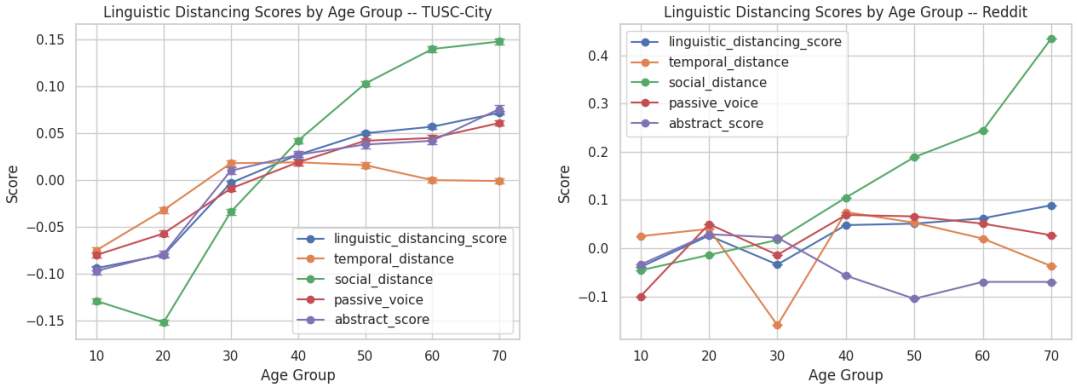

Figure 1: **Linguistic Distancing** scores for TUSC-City, and Reddit subset of the *AgeCorpus* (blue line). Individual components of linguistic distancing across age groups are shown: temporal distance (orange line), social distance (green line), passive voice (red line) and abstractness (purple line). Error bars represent the standard error of the mean.

5. Results

5.1. RQ1: How Does Linguistic Distancing Vary Across Age Groups?

In Figure 1, we show the trend of linguistic distancing across age groups. We see that regardless of the dataset (i.e., X or Reddit), linguistic distancing (blue line) occurs more *frequently* with age. This upwards trend is consistent across age groups (except with a dip in the 30's on Reddit). This result is also reflected in the statistical analyses. We show in Table 2 the F-statistic and p-value for the ANOVA tests. For all datasets, there is a significant increase in linguistic distance with age (Reddit: $F(6, 1203317.10) = 17703.63$, $p < 0.001$,$\eta^2 = 0.003$; TUSC-City: $F(6, 539267.68) = 2830.26$, $p < 0.001$, $\eta^2 = 0.009$). The effect size on the X dataset is near 0.01 which is considered small, and the effect size on the Reddit dataset is even smaller.[5]

Discussion: Past literature has found that overall well-being, lower negative affect, higher positive affect and improved emotion regulation skills is associated with age (Urry and Gross, 2010; Charles et al., 2023; Stone et al., 2010; Isaacowitz, 2022). Our findings show that with age, there are proportionally more instances of distancing. Given the relationship between linguistic distancing and emotion regulation, our findings support more use of reappraisal with age and in turn better well-being in the older ages. The drop in the 30's Reddit data is interesting as past work has found a dip in well-being during midlife (e.g., 35's–50's) (Blanchflower and Oswald, 2008; Blanchflower and Graham, 2021; Stone et al., 2010), however, there are some theories opposing the U-shape of happiness (Galambos et al., 2020).

5.2. RQ2: How does Social Distancing, Temporal Distancing, Abstractness, and Passive Voice Change with Age?

In Figure 1, we see the trend for each of the dimensions of linguistic distancing across age groups. Every dimension (i.e., *temporal* distance, *social* distance, *abstractness*, *passive*) occurs more frequently with age, although some have a steeper slope. *Social* distancing has the steepest slope across datasets based on the graph. This means that the rate at which *social* distancing increases with age is higher than the rate for other dimensions. In addition, *social* distancing also starts off at a lower point on the y-axis than all the other dimensions, allowing more room to grow with age. However, on Reddit *abstractness*, *temporal*, and *passive* scores plateau/decrease after age 40 and onwards, and likewise with *temporal* distancing on X. This could be due to the nature of the platforms and the topics discussed on each e.g., often people describe situations on Reddit to obtain advice and therefore talk in a more present-tense, less abstract manner. Reddit also has a dip in linguistic distancing in the 30's which is primarily driven by a decrease in temporal distance. This means that individuals are using more present tense verbs.

Discussion: While some of these dimensions have not been studied directly with age, they measure concepts that have been shown to change with age. For example, more *social* distancing is associated with fewer personal-pronouns used. Various literature point to the negative relationship between age and first-person singular pronouns (Schwartz et al., 2013; Pang et al., 2025; Pennebaker and Stone, 2003). The trends we see across X and

[5] Effect size of 0.01 is small; 0.06 is medium; and 0.14 is large. Effect sizes help us quantify whether differences are practically meaningful in the real-world.

Reddit support this pattern. In terms of *temporal* distancing, higher levels mean that fewer present tense verbs are used which is also supported by past work which points to more future-tense and fewer past-tense verbs with age (Pennebaker and Stone, 2003). While there are few works studying *passive voice* with age, older adults were found to produce fewer passive sentences than younger adults in Korean (Sung et al., 2024). Also, older adults (i.e., 70's) had more difficulty producing and understanding abstractness on tests (Albert et al., 1990). For abstractness, no statistical difference was found for the ability to define abstract words across participants aged 19–89, but the use of abstract vs. concrete words in everyday language was not examined (Pezzuti et al., 2021). While some work examines how each of these components vary across childhood or in terms of acquisition (e.g., linguistic distancing: Nook et al. (2020); Cohen et al. (2022); passive voice: Baldie (1976); Horgan (1978)), fewer work examines how they change across adulthood. Further, past work examining linguistic distancing does not consider the contributions of *temporal* vs. *social* distancing to linguistic distancing across age groups (Nook et al., 2017). Therefore, we provide new findings that lay the foundation for future exploration into what drives changes in linguistic distancing with age.

6. Conclusion

We study how linguistic distancing changes across adulthood to better understand emotion regulation with age. We use large social media datasets to systematically analyze how does linguistic distancing vary across teens to 70's on X and Reddit. We construct an interpretable measure of linguistic distancing based on four dimensions: *temporal* distancing, *social* distancing, *passive* voice and *abstractness*. We found that linguistic distancing occurs in proportionally more posts with age. When exploring how each of the dimensions changed across age groups, we found that *social* distancing has the highest rate of increase. We also saw that *temporal* distancing plateaus and decreases in later ages. Our work provides an initial foundation for understanding how linguistic distancing varies in text across age groups and supports further exploration into measuring linguistic distancing on social media. Future work could examine how such measures change across platforms discussing diverse topics (e.g., subreddits) and contrast results with LLM measures of linguistic distancing. Future work could also explore linguistic distancing in different languages and cultures, as psychological distancing and distancing in language may differ based on culture or contexts.

Ethical Considerations

Our research interest is to study how linguistic distancing changes across adulthood at the aggregate/group level. This has applications in emotional development psychology and in public health (e.g., overall well-being and mental health). However, personal well-being is complex, private, and central to an individual's experience. Additionally, each individual expresses themselves differently through language, which results in large amounts of variation.

Our work on studying linguistic distancing should not be construed as detecting how people feel; rather, we draw inferences on the language used.

The inferences we draw in this paper are based on aggregate trends across large populations. We do not draw conclusions about specific individuals or momentary states of well-being.

Limitations

The dataset used in this study relies on self-disclosure of age on social media. As with all scenarios involving self-disclosure, individuals may falsely report their age due to social pressures, or in order to relate to and engage with certain communities. Additionally, the templates used for pattern matching with age may not fully capture all possible ways to express age, so there may be age declarations not included in the dataset.

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Social Sciences and Humanities Research Council (SSHRC), the Digital Research Alliance of Canada (alliancecan.ca), and the Canadian Institute for Advanced Research (CIFAR). Compute resources were graciously supplied by Prairies DRI and the Digital Research Alliance of Canada. Alona Fyshe holds a Canada CIFAR AI Chair.

7. Bibliographical References

MS Albert, J Wolfe, and G Lafleche. 1990. Differences in abstraction ability with age. *Psychology and aging*, 5(1):94—100.

Amelia Aldao, Susan Nolen-Hoeksema, and Susanne Schweizer. 2010. Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, 30(2):217–237.

Jennifer L. Altshuler, Janice L. Genevaro, Diane N. Ruble, and Marc H. Bonstein. 1995. Children's knowledge and use of coping strategies during hospitalization for elective surgery. *Journal of Applied Developmental Psychology*, 16(1):53–76.

Brian J. Baldie. 1976. The acquisition of the passive voice. *Journal of Child Language*, 3(3):331–348.

J.S. Beck and A.T. Beck. 2020. *Cognitive Behavior Therapy, Third Edition: Basics and Beyond*. Guilford Publications.

David G Blanchflower and Carol L Graham. 2021. The u shape of happiness: A response. *Perspectives on psychological science : a journal of the Association for Psychological Science*, 16(6):1435—1446.

David G. Blanchflower and Andrew J. Oswald. 2008. Is well-being u-shaped over the life cycle? *Social Science Medicine*, 66(8):1733–1749.

Gerd Bohner. 2001. Writing about rape: Use of the passive voice and other distancing text features as an expression of perceived responsibility of the victim. *British Journal of Social Psychology*, 40(4):515–529.

John Bowlby. 1969. *Attachment and loss*. 79. New York: Basic Books.

Kirk Brown, Katherine Covell, and Rona Abramovitch. 1991. Time course and control of emotion: Age differences in understanding and recognition. *Merrill-Palmer Quarterly*, 37(2):273–287.

Laura Brummer, Lusia Stopa, and Romola Bucks. 2014. The influence of age on emotion regulation strategies and psychological distress. *Behavioural and Cognitive Psychotherapy*, 42(6):668–681.

Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior research methods*, 46(3):904—911.

Eugene Y. Chan and Sam J. Maglio. 2020a. The voice of cognition: Active and passive voice influence distance and construal. *Personality and Social Psychology Bulletin*, 46(4):547–558. PMID: 31390936.

Eugene Y. Chan and Sam J. Maglio. 2020b. The voice of cognition: Active and passive voice influence distance and construal. *Personality and Social Psychology Bulletin*, 46(4):547–558. PMID: 31390936.

Susan T Charles, Jonathan Rush, Jennifer R Piazza, Eric S Cerino, Jaqueline Mogle, and David M Almeida. 2023. Growing old and being old: Emotional well-being across adulthood. *Journal of personality and social psychology*, 125(2):455—469.

Katherine A. Cohen, Akash Shroff, Erik C. Nook, and Jessica L. Schleider. 2022. Linguistic distancing predicts response to a digital single-session intervention for adolescent depression. *Behaviour Research and Therapy*, 159:104220.

Kalee De France and Tom Hollenstein. 2019. Emotion regulation and relations to well-being across the lifespan. *Developmental psychology*, 55(8):1768—1774.

D Derryberry and MK Rothbart. 1988. Arousal, affect, and attention as components of temperament. *Journal of personality and social psychology*, 55(6):958—966.

María Félix-Ortiz, Ricardo Muñoz, and Michael D. Newcomb. 1995. The role of emotional distress in drug use among latino adolescents. *Journal of Child & Adolescent Substance Abuse*, 3(4):1–22.

Nancy L Galambos, Harvey J Krahn, Matthew D Johnson, and Margie E Lachman. 2020. The u shape of happiness across the life course: Expanding the discussion. *Perspectives on psychological science : a journal of the Association for Psychological Science*, 15(4):898—912.

Nadia Garnefski and Vivian Kraaij. 2007. The cognitive emotion regulation questionnaire. *European Journal of Psychological Assessment*, 23:141–149.

James J. Gross. 1998. The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, 2(3):271–299.

James J Gross and Oliver P John. 2003. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of personality and social psychology*, 85(2):348—362.

James J. Gross and Ricardo F. Muñoz. 1995. Emotion regulation and mental health. *Clinical Psychology: Science and Practice*, 2(2):151–164.

JJ Gross, LL Carstensen, M Pasupathi, J Tsai, CG Skorpen, and AY Hsu. 1997. Emotion and aging: experience, expression, and control. *Psychology and aging*, 12(4):590—599.

Sharon M Hall, Ricardo F Muñoz, Victor I Reus, and Karen L Sees. 1993. Nicotine, negative affect, and depression. *Journal of consulting and clinical psychology*, 61(5):761.

Dianne Horgan. 1978. The development of the full passive. *Journal of Child Language*, 5(1):65–80.

Tianqiang Hu, Dajun Zhang, Jinliang Wang, Ritesh Mistry, Guangming Ran, and Xinqiang Wang. 2014. Relation between emotion regulation and mental health: a meta-analysis review. *Psychological reports*, 114(2):341—362.

Molly Ireland, Kaitlin Adams, and Sean Farrell. 2022. Tracking mental health risks and coping strategies in healthcare workers' online conversations across the COVID-19 pandemic. In *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*, pages 76–88, Seattle, USA. Association for Computational Linguistics.

Derek M Isaacowitz. 2022. What do we know about aging and emotion regulation? *Perspectives on psychological science : a journal of the Association for Psychological Science*, 17(6):1541—1555.

Oliver P. John and James J. Gross. 2004. Healthy and unhealthy emotion regulation: Personality processes, individual differences, and life span development. *Journal of Personality*, 72(6):1301–1334.

E. Kross and O. Ayduk. 2017. Chapter two - self-distancing: Theory, research, and current directions. volume 55 of *Advances in Experimental Social Psychology*, pages 81–136. Academic Press.

Nira Liberman and Yaacov Trope. 2008. The psychology of transcending the here and now. *Science (New York, N.Y.)*, 322(5905):1201—1205.

Marsha M Linehan. 1993. *Skills training manual for treating borderline personality disorder.* Guilford press.

Mai Bjørnskov Mikkelsen, Mia Skytte O'Toole, Emma Elkjaer, and Mimi Mehlsen. 2024. The effect of age on emotion regulation patterns in daily life: Findings from an experience sampling study. *Scandinavian journal of psychology*, 65(2):231—239.

Saif Mohammad. 2011. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 105–114, Portland, OR, USA. Association for Computational Linguistics.

Erik C. Nook, Hayoung E. Ahn, Jessica L. Schleider, and Leah H. Somerville. 2025. Emotion regulation is associated with increases in linguistic measures of both psychological distancing and abstractness. *Affective Science*, 6(1):63–76. Publisher Copyright: © The Author(s) 2024.

Erik C. Nook, Thomas D. Hull, Matthew K. Nock, and Leah H. Somerville. 2022. Linguistic measures of psychological distance track symptom levels and treatment outcomes in a large set of psychotherapy transcripts. *Proceedings of the National Academy of Sciences*, 119(13):e2114737119.

Erik C Nook, Jessica L Schleider, and Leah H Somerville. 2017. A linguistic signature of psychological distancing in emotion regulation. *Journal of experimental psychology. General*, 146(3):337—346.

Erik C. Nook, Constanza M. Vidal Bustamante, Hyun Young Cho, and Leah H. Somerville. 2020. Use of linguistic distancing and cognitive reappraisal strategies during emotion regulation in children, adolescents, and young adults. *Emotion*, 20(4):525–540. Publisher Copyright: © 2020 American Psychological Association.

Dandan Pang, Sharath C. Guntuku, Garrick Sherman, Tingting Liu, Sunny Rai, Young-Min Cho, Rian You, Arnold B. Bakker, and Lyle H. Ungar. 2025. Understanding gender and age differences in language use: cross-cultural insights from weibo and facebook. *Humanities Social Sciences Communications*, 12(1):1667. Copyright - © The Author(s) 2025. This work is published under http://creativecommons.org/licenses/by-nc-nd/4.0/ (the "License"). Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License; Last updated - 2025-11-07; SubjectsTermNotLitGenreText - United States–US; China.

James W Pennebaker and Lori D Stone. 2003. Words of wisdom: language use over the life span. *Journal of personality and social psychology*, 85(2):291—301.

James W Pennebaker et al. Linguistic inquiry and word count: Liwc 2001.

Lina Pezzuti, James Dawe, and Anna Maria Borghi. 2021. Does mastering of abstract words decline with age? *Educational Gerontology*, 47(12):527–542.

Andrew J. Reagan, Lewis Mitchell, Dilan Kiley, Christopher M. Danforth, and Peter S. Dodds. 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5(1):31.

H. Andrew Schwartz, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Stephanie M. Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin E. P. Seligman, and Lyle H. Ungar. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS ONE*, 8(9):1–16.

Amir Sepehri, Mitra Sadat Mirshafiee, and David M. Markowitz. 2023. Passivepy: A tool to automatically identify passive voice in big text data. *Journal of Consumer Psychology*, 33(4):714–727.

Gal Sheppes, Gaurav Suri, and James J Gross. 2015. Emotion regulation and psychopathology. *Annual review of clinical psychology*, 11:379—405.

Robert S Stawski, Martin J Sliwinski, David M Almeida, and Joshua M Smyth. 2008. Reported exposure and emotional reactivity to daily stressors: the roles of adult age and global perceived stress. *Psychology and aging*, 23(1):52—61.

Arthur A. Stone, Joseph E. Schwartz, Joan E. Broderick, and Angus Deaton. 2010. A snapshot of the age distribution of psychological well-being in the united states. *Proceedings of the National Academy of Sciences*, 107(22):9985–9990.

Jee Eun Sung, Eunha Jo, Sujin Choi, and Jiyeon Lee. 2024. Coordinating words and sentences: Detecting age-related changes in language production. *Journal of Speech, Language, and Hearing Research*, 67(1):211–220.

Stefan Sütterlin, Muirne C. S. Paap, Stana Babic, Andrea Kübler, and Claus Vögele. 2012. Rumination and age: Some things get better. *Journal of Aging Research*, 2012(1):267327.

Daniela Teodorescu, Jan Wahle, and Saif M. Mohammad. 2026. Age and affect in language: How emotion expression on social media varies across adulthood.

Ross A. Thompson. 1991. Emotional regulation and emotional development. *Educational Psychology Review*, 3(4):269–307.

Enrica Troiano, Sofie Labat, Marco Antonio Stranisci, Rossana Damiano, Viviana Patti, and Roman Klinger. 2024. Dealing with controversy: An emotion and coping strategy corpus based on role playing. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 1634–1658, Miami, Florida, USA. Association for Computational Linguistics.

Yaacov Trope and Nira Liberman. 2003. Temporal construal. *Psychological review*, 110(3):403—421.

Yaacov Trope and Nira Liberman. 2010. Construal-level theory of psychological distance. *Psychological review*, 117(2):440—463.

Heather L. Urry and James J. Gross. 2010. Emotion regulation in older age. *Current Directions in Psychological Science*, 19(6):352–357.

A. Patterns to Match Age Declarations

In Table 3, we show the pattern templates used to match age declarations in seed posts in the *AgeCorpus*.

B. Levene's Test of Homogeneity of Variance

Levene's test indicated that the assumption for homogeneity of variance was not violated for the effect of age group on linguistic distancing for TUSC-Country, but was violated for Reddit and TUSC-City subsets of the *AgeCorpus*. We show the results in Table 4.

C. TUSC-Country Results

In the following Sections we show the number of posts in the TUSC-Country version of the dataset and the linguistic distancing results on this subset of the *AgeCorpus*.

C.1. *AgeCorpus*: TUSC-Country

In Table 5, we show the number of posts in TUSC-Country, TUSC-City and Reddit subsections of the *AgeCorpus* for comparison.

C.2. Linguistic Distancing Results

In Table 6, we show the results from the ANOVA tests for the difference in linguistic distancing across age groups. We include the rows for the Reddit and TUSC-City subsets of the *AgeCorpus* dataset as a comparison.

In Figure 2, we show the trends of linguistic distancing across age groups (blue line), as well as the trends for *temporal* distancing, *social* distancing, *passive voice* and *abstractness*. Largely the trends for TUSC-Country follow those of TUSC-City, with an increasing proportion of distancing occurring with age. Generally, all dimensions of distancing grow similarly across age groups, with social distancing having the steepest slope.

Regex	Example
[I] `\bI(?:\s+am|'m)\s+(\d{1,2})\s+years?\s+old\b` *I'm 30 year old*	[I]*I am 25 years old*
[I]`\bI(?:\s+am|'m)\s+(\d{1,2})` `(?=\s*(?:$|[,.!?;:\-]|(?:and|but|so|yet)\s))` *I'm 30, and ...*	[I]*I am 25.*
[I]`\bI(?:\s+was|\s+am|'m)\s+born\s+in\s+` `(19\d{2}|20(?:0\d|1\d|2[0-4]))\b` *I am born in 2005*	[I]*I was born in 1998*
[I]`\bI(?:\s+was|\s+am|'m)\s+born\s+in\s+'(\d{2})\b`	*I was born in '98* *I'm born in '05*

[I] `\bI\s+was\s+born\s+on\s+`
`(?:\d{1,2}(?:st|nd|rd|th)?\s+)?`
`(?:January|February|March|April|May|June|`
`July|August|September|October|November|December|`
`Jan|Feb|Mar|Apr|May|Jun|Jul|Aug|Sep|Sept|Oct|Nov|Dec)`
`\s+ (?:\d{1,2}(?:st|nd|rd|th)?,?\s+)?`
`(19\d{2}|20(?:0\d|1\d|2[0-4]))\b`
I was born on March 15th, 1998
	[I]*I was born on 15 March 1998*

[I]`\bI\s+was\s+born\s+on\s+\d{1,2}` `[/\-]\d{1,2}[/\-](19\d{2}|20(?:0\d|1\d|2[0-4]))\b` *I was born on 15-03-1998*	[I]*I was born on 03/15/1998*

Table 3: Regexes used to identify users from both the Reddit and X datasets.

Dataset	Metric	df1	df2	F-statistic	P-value
Reddit	Linguistic	6	32964822	10269.65	$p<.001$
TUSC-City	Linguistic	6	1918310	86.13	$p<.001$
TUSC-Country	Linguistic	6	11,940	1.12	$p=.348$

Table 4: The degrees of freedom, F-statistic, and p-value in Levene's test of Homogeneity of Variances for linguistic distancing and each dimension of linguistic distancing across the subsets of the *AgeCorpus*.

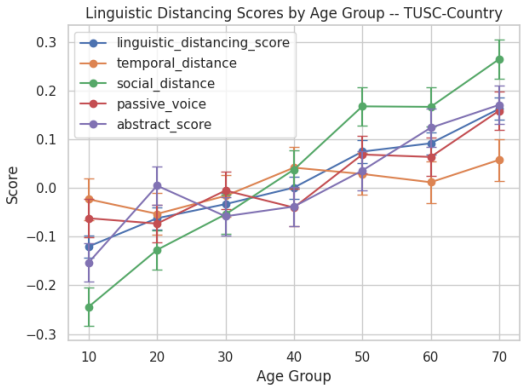

Figure 2: **Linguistic Distancing** scores for TUSC-Country subset of the *AgeCorpus* (blue line). Individual components of linguistic distancing across age groups are shown: temporal distance (orange line), social distance (green line), passive voice (red line) and abstractness (purple line). Error bars represent the standard error of the mean.

Age Group	**#Posts**		
	Country	City	Reddit
13–19	551	94,857	9,281,055
20–29	3,076	406,027	15,455,426
30–39	3,001	462,238	6,151,757
40–49	1,808	360,946	1,220,498
50–59	1,541	278,809	449,230
60–69	1,551	218,008	246,398
70–79	419	97,432	160,465

Table 5: The number of posts across the age groups in each subset of the *AgeCorpus*. "Country" refers to TUSC-Country and "City" refers to TUSC-City subsets of the dataset. We use this data in our experiments.

Dataset	Metric	df1	df2	F-statistic	P-value	Effect Size (*est* ω^2)
Reddit	Linguistic	6	1203317.10	17703.63	$p<.001$	0.003
TUSC-Country	Linguistic	6	11940	29.66	$p<.001$	0.015
TUSC-City	Linguistic	6	539267.68	2830.26	$p<.001$	0.009

Table 6: The degrees of freedom (for the numerator and denominator), F-statistic, p-value, and effect size in the one-way ANOVA test for differences in linguistic distancing between age groups. Welch's ANOVA was performed for Reddit and TUSC-City.