Applied Psycholinguistics **25** (2004), 135–143
Printed in the United States of America
DOI: 10.1017.S0142716404001079

Adult age differences in judgments of semantic fit

DEBORAH M. LITTLE, KRISTEN J. PRENTICE,
and ARTHUR WINGFIELD
Brandeis University

ADDRESS FOR CORRESPONDENCE
Dr. Deborah M. Little, Center for MR Research (MC 707), Outpatient Care Center, Suite 1A, University of Illinois at Chicago, Chicago, IL 60612-8097. E-mail: little@uic.edu

ABSTRACT
A total of 130 young adults and 173 older adults gave goodness of fit ratings for sets of word alternatives for 123 English sentence frames. Word alternatives were semantically plausible in all cases but varied in the likelihood of their occurrence within the sentence frames. Adding support to the general belief that vocabulary and language knowledge are well preserved in normal aging, our results showed generally high agreement in goodness of fit ratings by the young and older adults. Differences that did occur appeared to be due to cohort differences in word usage and interests rather than underlying cognitive function. We suggest that this similarity in semantic judgments may help to explain why adult age differences in comprehension and memory for sentences are generally smaller than age differences for other examples of episodic memory.

Perhaps no single question in cognitive aging has attracted as much research effort as the source of older adults' decline in episodic memory (Kausler, 1994). In the domain of verbal learning, one frequently raised factor is the importance of the quick formation of useful semantic and temporal associations during the course of learning and the reinstatement of these associations at the time of retrieval (e.g., Kahana, Howard, Zaromb, & Wingfield, 2002; Kahana & Wingfield, 2000; Wingfield, Lindfield, & Kahana, 1998). Although age differences in rates of learning word lists are reduced when the lists are composed of semantically related items or items that are prototypical of a single category, a residual age difference typically remains (Kausler, 1994; Wingfield & Kahana, 2002). This remaining difference has raised the question of whether older adults might have different prototypicality judgments or different patterns of semantic relations between words. Changes in either of these factors could produce an age-related disadvantage relative to their younger counterparts.

One approach to this question has been to examine relative frequency distributions of associative responses given by young and older adults in tests of free association (e.g., Perlmutter, 1979; Riegel & Birren, 1966; Tresselt & Mayzner, 1964) and in the generation of responses that fit various conceptual categories (Howard, 1980). Although older adults tend to show a general slowing in word retrieval (Bowles & Poon, 1985; Burke, MacKay, Worthley, & Wade, 1991), vocabulary

© 2004 Cambridge University Press 0142-7164/04 $12.00

https://doi.org/10.1017/S0142716404001079 Published online by Cambridge University Press

knowledge is generally found to remain intact (Bowles, 1993; Bowles, Obler, & Poon, 1989; Nicholas, Barth, Obler, Au, & Albert, 1997). Consistent with this, many studies have shown similarities across young and older adults in word associations and category-member associations (Bowles, Williams, & Poon, 1983; Brown & Mitchell, 1991; Burke & Peters, 1986; Howard, 1980; Lovelace & Cooley, 1982). Age differences in association norms have been reported, but these differences appear predominantly for low-frequency responses (Bowles et al., 1983; Burke & Peters, 1986; Howard, 1980; Lovelace & Cooley, 1982).

Although the overall trend of the literature is against age differences for word and category member associations, the question remains as to whether there are changes in the relationship between meaningful semantic phrases with increased age. There is also the question of whether such differences may contribute to the consistent finding in the literature of small, but still significant, age differences observed in recall of meaningful sentences (see Wingfield & Stine–Morrow, 2000, for a review). The magnitude of these age differences are larger for sentences with greater syntactic complexity, an effect due in part to such sentences having a heavier demand on working memory resources (Caplan & Waters, 1999; Kemper, 1986; Light, 1991; Norman, Kemper, Kynette, Cheung, & Anagnopoulos, 1991; Wingfield, Peelle, & Grossman, in press). These findings, in combination with the above noted reduction in age-related changes for semantically related materials, raise the question as to whether semantic predictability, like syntactic complexity, would also have a differential effect on older adults' memory of sentence materials.

As indicated above, however, one might ask whether older adults' poorer sentence recall can also be partially explained by differences in sensitivity to the semantic fit of words within sentence frames. Were a difference such as this to exist, older adults might be less able than younger adults to use the contextual relations between words in a sentence to enhance the likelihood of recall (Potter, 1990; Wingfield, Tun, & Rosen, 1995). Relatively subtle age differences in semantic fit for words in context could thus have important consequences in memory for sentences. This question of sensitivity to the goodness of fit of a word in a sentence is not answered by the previous studies that have shown age similarities in word associations and category prototypicality ratings.

Our goal in the present study was to assess young and older adults' judgments of the goodness of fit of a particular word within a sentence context. Imagine, for example, that one were given the sentence, "The meeting was held in the _____ this afternoon." Given a word like *library*, *study*, or *hall*, one might rate any of these as a good fit to this sentence, in the sense that they are all words one might reasonably expect to encounter in such a sentence. Other words, such as *loft*, *booth*, or *closet*, are also possible but are much less likely. Given any of these alternatives for this sentence frame, most people would agree that they are certainly possible but much less probable as a place to hold a meeting. These differences in judgments can be referred to as their *goodness of fit* within the context of a sentence.

In this study we asked young and older adults to rate the goodness of fit for alternative words seen in a total of 123 English sentence frames. In all cases the word alternatives would be semantically plausible, although many would be of a low likelihood of occurrence within the specific sentence context. We had

https://doi.org/10.1017/S0142716404001079 Published online by Cambridge University Press

two goals in conducting this study. The first goal was to compare semantic fit judgments for words in a sentence context by young and older adults. Our second goal was to develop a set of stimuli that investigators might find useful in studies of adult age differences in comprehension of, and memory for, linguistic materials.

METHOD

Participants

Ratings were obtained from a total of 303 participants (130 young adults, 173 older adults). The young participants were university undergraduates (48 men, 82 women) between the ages of 17 and 25 years ($M = 18.8$). At the time of testing they had a mean of 12.9 years of formal education ($SD = 1.0$). The older participants were a group of healthy, community-dwelling adults (62 men, 111 women) between the ages of 60 and 89 years ($M = 74.3$). The older adult group had an average of 15.6 years of formal education ($SD = 2.6$), representing an average of 2.7 more years of formal education than the younger group, $t\ (294) = 11.43$, $p < .01$. All of the participants were native speakers of American English. (Responses were initially collected from 311 participants. The data reported here do not include 7 participants for whom English was not their native language and 1 participant who responded over 50% of the time with only the highest rating.)

Stimulus materials

The stimuli consisted of 123 sentence frames ranging from 8 to 10 words in length. Twelve versions of each sentence were prepared, that were identical in all regards except for a single target word. Across the 123 stimulus sentences the target word was always located between 2 and 5 words from the end of the sentence. An example sentence and target words are presented in Table 1. (The full set of 123 sentences and their target words are available at http://www.bio.brandeis.edu/memlab/norms.)

All of the sentences were created for this study. Some of the target words were taken from the Battig and Montague (1969; Uyeda & Mandler, 1980) norms in which participants were asked to give prototypicality ratings for the goodness of fit of words to various semantic categories. Additional word alternatives were generated by the authors. All of the words for each category were first chosen for their semantic relatedness to one another. The target sentence was then generated to fit the semantically related words in each category. Each of the words chosen to fit in each sentence frame belonged to the same syntactic class. The context established by the sentence produced the bias for the goodness of fit of the items within the sentence. It was intended that the participants would know the meanings of all of the target words. Over 90% of the exemplars were rated as having a high familiarity between 6 and 7 in the Hoosier Mental Lexicon database (Nusbaum, Pisoni, & Davis, 1984).

Before the final exemplars and sentences were presented to participants, a pilot study was conducted to ensure that none of the exemplars would appear to be

https://doi.org/10.1017/S0142716404001079 Published online by Cambridge University Press

Table 1. *Example sentence frame ("After the battle, doctors removed the _____ in his leg.)" and target words*

	Overall		Young Adults		Older Adults	
Sentence and Exemplars	Mean	*SD*	Mean	*SD*	Mean	*SD*
Shrapnel	7.59	2.15	6.30	2.95	8.35	1.00
Bullet	7.33	1.93	8.00	1.31	7.00	2.13
Pellet	5.63	2.34	5.10	1.73	6.22	2.86
Slug	5.59	3.04	2.82	1.54	7.28	2.42**
Shell	5.20	2.61	5.55	2.02	4.93	3.05
Arrow	4.91	2.41	5.73	2.20	4.09	2.43
Buckshot	3.92	2.26	2.90	1.45	2.64	2.50
Mortar	3.39	2.74	3.67	2.81	3.19	2.76
Grenade	3.38	2.48	3.30	2.45	3.45	2.62
Dart	2.57	1.63	2.00	1.12	3.00	1.86
Stone	2.04	1.14	2.75	1.14	1.38	0.65*
Bomb	1.29	0.55	1.46	0.66	1.09	0.30

Note: Following each exemplar we show the overall mean rating averaged over both the young and older raters (1 = *a least expected word*, 9 = *a most expected word*) and the standard deviations. This is followed by the means for each of the two age groups.
$^{*}p < .01$. $^{**}p < .00004$. The asterisks indicate a significant difference between the mean ratings for the young and older participants.

anomalous within the sentence frames. In this pilot study a group of 36 participants (age range = 22–47, $M = 32$ years) was asked to rate the words' goodness of fit within the sentences, and to give a rating of "does not fit" if the target word did not seem appropriate to the sentence context either syntactically or semantically. Any target word that received this rating from more than one participant was removed and replaced. None of the pilot subjects were used in the final study.

Procedures

Our interest was to obtain participants' estimates of how well a word "fit" within a sentence context. It was thus important to us that the wording of the instructions be understandable to the participants and likely to elicit this judgment. The phrasing that best met these criteria was to ask the participants to judge how predictable the target word would be in the sentence context. Each participant was shown each of the 123 sentences with the following instructions:

> On the following four pages you will find a list of sentences. In each sentence, one word is underlined. Your job is to judge how predictable the underlined word is in the sentence. Next to each sentence you will find a series of numbers 1 through 9. This is the scale you are going to use to judge how predictable the underlined word is. If the underlined word is the one that you would most expect to see in that sentence, please circle the "9" on the scale. If the underlined word is one that you would least expect to see in that sentence, please circle the "1" on the scale. The more predictable

https://doi.org/10.1017/S0142716404001079 Published online by Cambridge University Press

the word, the larger the number you will circle on the scale. Please circle only one number per sentence.

For example, consider the following two sentences: (1) She put the birthday candles on the cake. (2) She put the birthday candles on the floor. The underlined word in sentence (1), "cake," is highly predictable. That is, you would expect the word "cake" to appear at that point in the sentence, and you would therefore circle one of the higher numbers on the scale. Conversely, the underlined word in sentence (2), "floor," is less predictable. That is, you would probably not expect to see that word in the sentence, and you would therefore circle one of the lower numbers on the scale.

Although in this example the target word was the final word in the sentence, participants understood that in the test sentences target words would not be sentence-final words.

Each participant saw only one version of each sentence, with different participants seeing each of the 12 word alternatives for that sentence. With a total of 303 participants, we thus had ratings for each of the 12 words in each frame from an average of 11 young adults and 15 older adults.

RESULTS

Table 1 shows one example sentence and the data collected. Overall mean ratings averaged over both the young and older raters (1 = *a least expected word*, 9 = *a most expected word*) for each target word in each sentence frame are listed in decreasing order of goodness of fit. (Standard deviations are given in parentheses.) This is followed by the separate group means for the young and older participants. Single asterisk presented at the end of each row indicate a significant difference between the mean ratings for the young and older participants, with a liberal threshold of $p < .01$ and no correction for multiple comparisons. To control for multiple comparisons we applied a Bonferroni adjustment and a threshold of $p < .00004$ was established. Comparisons meeting this more conservative criterion are indicated by two asterisks.

The primary question of interest was whether older and younger adults would rate items similarly. In order to answer this question we first rank ordered the young adults' ratings for the 12 alternative words for each of the 123 sentence frames and then plotted the mean rating for young and older adults as a function of the younger adults' mean rank order. This produced a plot of younger and older adults' ratings for the same words across all the 123 sentence frames. These functions are presented in Figure 1 where the similarity between young and older adults' ratings is clearly visible. Overall there was a high correlation between young and older adults' ratings of the goodness of fit of the various words in the sentence frames, $r\ (1248) = .77$, $p < .001$.

This tendency was confirmed with 123 tests on the rank order (Mann–Whitney U) by younger and older adults for the ratings of each exemplar for the 123 sentences.[1] The rank order tests demonstrated significant differences ($p < .05$) for only seven of the sentence frames. These seven sentences and the exemplars that produced these age differences are available in the full set of materials available on the website. One example of these ordering differences occurred for the

https://doi.org/10.1017/S0142716404001079 Published online by Cambridge University Press

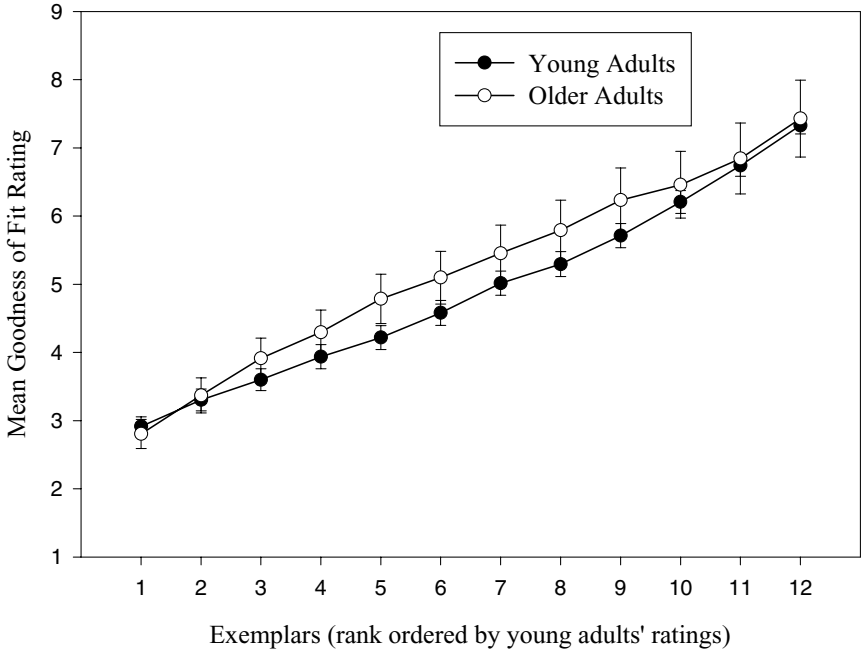


Figure 1. Mean goodness of fit ratings given by the young and older participants for each of the 12 exemplars and ranked from lowest to highest, averaged across all sentences. The error bars represent 1 standard error.

sentence frame "The doctor treated her _____ with medication." Both younger and older adults rated "fever" and "hepatitis" as the two most prototypical but differed on their rankings of all other items. Specifically, younger adults rated "scurvy," "rubella," and "malaria" as the three least likely items, whereas older adults rated the "mumps," "cholera," and "polio" as the least likely items.

Figure 1 also shows that within the general similarity in ratings, there was a tendency for the older adults to assign higher goodness of fit ratings than the young adults. This tendency was confirmed by a one-way analysis of variance, $F\ (1, 302) = 50.30$, $p < .05$. As can be seen by the relative sizes of the error bars in Figure 1, there was also a tendency for the older adults to show greater inter-participant variability than the young adults.

We conducted one further analysis on these data, which was to determine, on a word by word and sentence by sentence basis, any significant differences in the absolute ratings given by the young and older adults. Out of a total of 1248 comparisons, the younger and older adults rated items significantly different ($p < .01$) in 101 cases (8% of the comparisons). After the Bonferroni adjustment was applied ($p < .00004$) the number of significant comparisons was reduced to 19 cases (1.5%). These cases, as noted previously, are indicated with the full data set on the website.

https://doi.org/10.1017/S0142716404001079 Published online by Cambridge University Press

CONCLUSIONS

When taken together the age groups show great similarity. There were remarkably few differences in the ratings of individual words. In addition, as previously indicated there were only a small number of sentence frames for which the young and older adults showed an ordering difference ($n = 7$) that reflected what we believe to be notable but reasonable cohort differences. What we believe to be cohort differences most times did not significantly alter the ordering of exemplars but instead produced smaller effects on single exemplars within a sentence frame. One example is observable for the sentence frame, "When the music played, she remembered her first _____ lesson." Whereas both age groups' ratings ranked *ballet* highest (young = 6.55, older = 7.67), the young adults gave *salsa* a mean rating of 4.91 and the older adults gave it a mean rating of only 1.40. In the same sentence, *polka* had the second highest rating for the older adults (7.50), whereas it was rated seventh highest by the young adults (5.00).

As we indicated earlier, prior studies have shown general similarities in the patterns of word associations and category prototypicality ratings given by young versus older adults (Bowles et al., 1983; Burke & Peters, 1986; Howard, 1980; Lovelace & Cooley, 1982). As we also indicated, however, such findings do not speak to the question of whether there might nevertheless be differences between young and older adults in their sensitivity to the goodness of fit of words within the contexts of meaningful sentences.

Our analysis of goodness of fit ratings for words in sentence contexts showed generally good agreement between the 130 young adults and 173 older adults tested. Indeed, the cases where the young and older adults disagreed on ratings were few in number and most probably reflected differences in word usage and interests rather than any cognitive shifts from young to older adulthood. A tendency was observed for older adults to give overall higher goodness of fit ratings than the young adults, but in these cases the general orderings were similar for the two age groups. Such findings reinforce the view that functions having to do with vocabulary and language knowledge remain stable in older adulthood (Wingfield & Stine–Morrow, 2000). The results also suggest that the age differences often demonstrated for memory for sentence materials are not due to age differences in sensitivity to the semantic fit of the component words.

Investigators wishing to use these stimuli in experimental studies with young and older adults can avoid the small number of cases where age differences appeared. Other investigators may be particularly interested in exploring these few differences.

ACKNOWLEDGMENTS

This research was supported by NIH Grants AG04517 and AG15852 from the National Institute on Aging. The first two authors gratefully acknowledge NIA Training Grant AG00204 to Brandeis University. We also acknowledge support from the W. M. Keck Foundation and thank Jennifer Silber for help in data collection and entry and Karen Kemtes and Patricia Tun for helpful comments on the manuscript. Deborah Little is now at the Center for MR Research, College of Medicine, University of Illinois at Chicago. Kristen Prentice is now at

https://doi.org/10.1017/S0142716404001079 Published online by Cambridge University Press

the Maryland Psychiatric Research Center, University of Maryland. All stimuli are available at http://www.bio.brandeis.edu/memlab/norms or from the first author by request.

NOTE

1. The older adults tended to use the extreme ratings much more often than the intermediate ratings, compared to the young adults, who used all ratings equally. This trend was confirmed with a chi-square value of 1310.09 ($p < .001$). In fact, the use of extreme rating points on the scale by the older adults was twofold greater than any of their intermediate rating points.

REFERENCES

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology, 80*, 1–46.

Bowles, N. L. (1993). Semantic processes that serve picture naming. In J. Cerella, J. Rybash, W. Hoyer, & M. L. Commons (Eds.), *Adult information processing: Limits on loss* (pp. 303–326). San Diego, CA: Academic Press.

Bowles, N. L., Obler, L. K., & Poon, L. W. (1989). Aging and word retrieval: Naturalistic, clinical, and laboratory data. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 244–264). Cambridge: Cambridge University Press.

Bowles, N. L., & Poon, L. W. (1985). Aging and retrieval of words in semantic memory. *Journal of Gerontology, 40*, 71–77.

Bowles, N. L., Williams, D., & Poon, L. W. (1983). On the use of word association norms in aging research. *Experimental Aging Research, 9*, 175–177.

Brown, A. S., & Mitchell, D. B. (1991). Age differences in retrieval consistency and response dominance. *Journal of Gerontology, 46*, 332–339.

Burke, D., MacKay, D., Worthley, J., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language, 30*, 542–579.

Burke, D. M., & Peters, L. (1986). Word associations in old age: Evidence for consistency in semantic encoding during adulthood. *Psychology and Aging, 1*, 283–291.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22*, 77–126.

Howard, D. V. (1980). Category norms: A comparison of the Battig and Montague (1969) norms with the responses of adults between the ages of 20 and 80. *Journal of Gerontology, 35*, 225–231.

Kahana, M. J., Howard, M., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency and lag-recency effects in free recall. *Journal of Experimental Psychology: Learning, Memory & Cognition, 28*, 530–540.

Kahana, M. J., & Wingfield, A. (2000). A functional relation between learning and organization in free recall. *Psychonomic Bulletin & Review, 7*, 516–521.

Kausler, D. M. (1994). *Learning and memory in normal aging*. San Diego, CA: Academic Press.

Kemper, S. (1986). Imitation of complex syntactic constructions by elderly adults. *Applied Psycholinguistics, 7*, 277–288.

Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology, 42*, 333–376.

Lovelace, E. A., & Cooley, S. (1982). Free associations of older adults to single words and conceptually related word triads. *Journal of Gerontology, 37*, 432–437.

Nicholas, M., Barth, C., Obler, L. K., Au, R., & Albert, M. L. (1997). Naming in normal aging and dementia of the Alzheimer's type. In H. Goodglass & A. Wingfield (Eds.), *Anomia: Neuroanatomical and cognitive correlates* (pp. 166–188). San Diego, CA: Academic Press.

Norman, S., Kemper, S., Kynette, D., Cheung, H. T., & Anagnopoulos, C. (1991). Syntactic complexity and adults' running memory span. *Journal of Gerontology, 46*, 346–351.

https://doi.org/10.1017/S0142716404001079 Published online by Cambridge University Press

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). *Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words* (Research on Spoken Language Processing Progress Report 10, pp. 357–376). Bloomington, IN: Speech Research Laboratory.

Perlmutter, M. (1979). Age differences in the consistency of adults' associative responses. *Experimental Aging Research, 5*, 549–553.

Potter, M. C. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory & Language, 29*, 633–654.

Riegal, K. F., & Birren, J. E. (1966). Age differences in verbal associations. *Journal of Genetic Psychology, 106*, 153–170.

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

Uyeda, K. M., & Mandler, G. (1980). Prototypicality norms for 28 semantic categories. *Research Methods & Instrumentation, 12*, 587–595.

Wingfield, A., & Kahana, M. J. (2002). The dynamics of memory retrieval in older adulthood. *Canadian Journal of Experimental Psychology, 56*, 187–199.

Wingfield, A., Lindfield, K. C., & Kahana, M. J. (1998). Adult age differences in the temporal characteristics of category free recall. *Psychology and Aging, 13*, 256–266.

Wingfield, A., Peelle, J. E., & Grossman, M. (in press). Speech rate and syntactic complexity as multiplicative factors in speech comprehension by young and older adults. *Journal of Aging, Neuropsychology, & Cognition.*

Wingfield, A., & Stine–Morrow, E. A. L. (2000). Language and speech. In F. I. M. Craik & T. A. Salthouse (Eds.), *Handbook of aging and cognition* (2nd ed., pp. 359–416). Mahwah, NJ: Erlbaum.

Wingfield, A., Tun, P. A., & Rosen, M. J. (1995). Age differences in veridical and reconstructive recall of syntactically and randomly segmented speech. *Journal of Gerontology: Psychological Sciences, 50B*, P257–P266.

https://doi.org/10.1017/S0142716404001079 Published online by Cambridge University Press