Age-related patterns in lexical bundle usage: Evidence from a corpus of vernacular Japanese

Kevin Heffernan (Kwansei Gakuin University, Japan) and Yo Sato (Satsuma Language Services, England)

Studies on cognition and aging report slower cognitive performance with increased age (Deary et al., 2010; Nyberg et al., 1996; Pauls et al., 2013; Peelle et al., 2010). However, we do not fully understand the causes of reduced performance. On one hand, these researchers argue that reduced performance is a consequence of reduced processing speed, which in turn is a consequence of aging of the brain. On the other hand, Ramscar and colleagues argue that reduced processing speed simply reflects lifelong learning (Ramscar et al., 2014). These two arguments make predictions about lexical bundle usage. Motivated by these two stances—the aging hypothesis and the learning hypothesis—we ask the question: Does lexical bundle usage show age-related patterns? We review these stances in more detail, and then posit several hypotheses based on them. We then present our investigation of the relationship between lexical bundle usage and age.

These two stances make different predications about age-related trends in lexical bundle usage. Lexical bundles are frequently occurring sequences of words, such as ‘have a look at’ and ‘don’t you think that’. Due to their high frequency, speakers store lexical bundles as a cohesive package in declarative memory. Declarative knowledge includes our semantic knowledge of language, and our episodic knowledge of facts and events. Declarative knowledge also encompasses idiomatic knowledge of proverbs, collocations, and lexical bundles (Ullman, 2016; van Lancker Sidtis, 2012), although the exact nature of this knowledge is still unclear. Importantly, the high level of cohesion within a lexical bundle leads to its faster processing compared to less frequent expressions (Siyanova-Chanturia et al., 2011).

The advocates of the aging hypothesis argue that our reduced processing performance is a consequence of biological changes. For example, Peelle et al. (2009) examined brain activity in older and younger adults when processing sentences with two levels of syntactical complexity. They found that when processing the complex sentences, older adults produced less neural activity in the areas of the brain associated with syntactic processing, and showed less overall coordinated activity between different areas. Morrison and Baxter (2012) present evidence for similar age-related trends in cognitive activity in both language-speaking humans and non-language-speaking animals such as rhesus monkeys. They claim that “[t]he cognitive processes that are mediated by the hippocampus… are those that are most vulnerable to ageing.” The hippocampus “underlies the rapid linking (binding, associating) of different bits of knowledge or experiences” (Ullman, 2016), that is, our episodic knowledge. These researchers argue that the human brain ages, which has consequences for language usage in old age. In contrast to the aging hypothesis, Ramscar et al. (2014) argue that older adults’ reduced cognitive response speed is not evidence for cognitive decline. They instead posit that the reduced response patterns are a result of the added volume of information a person acquires through day-to-day experiences. Choosing between or recalling items such as memories or lexical information becomes more difficult as the number of stored items increases. Using computer simulations of adult learners acquisition of vocabulary over the adult lifespan, they demonstrated that as an individual's experience grows, his or her knowledge will increase, and that this will in turn raise the processing costs in his or her cognitive system. Ramscar et al. (2014, p.34) conclude that "there is no neurobiological evidence for any declines in the processing capacities of healthy older adults."

If older speakers struggle with the reduced performance of an aged declarative memory, then we expect less utilization of lexical bundles with increased age. On the other hand, if the underlying cause of older speakers' slower processing is larger volume, then we expect there to be no correlation between age and lexical bundle usage, since by definition lexical bundles are few in number. Another possibility is that older speakers may rely on lexical bundles—which are processed faster—more than younger speakers to compensate for their slower cognitive processing. Furthermore, if increased lexical knowledge due to older speakers' greater life experience plays a role, then we expect that older speakers will use a greater variety of lexemes. However, we do not expect a corresponding increase in lexical bundles as we assume that adult do not continue to learn lexical bundles throughout their lifetime, again because there is a very small number of them.

We used the Corpus of Kansai Vernacular Japanese, a collection of 152 conversational interviews between students (the interviewers) and family members or close acquaintances (the interviewees). The average length of an interview is 55 minutes. The corpus also contains personal information about each interviewee, such as age and education level. Each interview was transcribed, parsed into morphemes and tagged with part of speech information. We limit our investigation to the interviewee (hereafter, speaker) data, which consists of 885,027 morphemes. We divided the speakers into four age groups: students (N = 49, age range 15 to 23 years old), younger adults (N = 32, age range 24 to 39 years old), middle-aged adults (N = 34, age range 40 to 59 years old), and older adults (N = 37, age range 60 to 79 years old). In order to determine the lexical bundles, we took an equally-sized sample of 171,000 morphemes (the size of the smallest group) from each age group. Following Biber et al. (2004), we compiled a list of the four-morpheme sequences that occurred at least 40 times per million morphemes (or in our case, at least 27 times in our balanced sample). This methodology yielded 155 bundles, of which we removed 12 bundles that we judged to be specific to our data, such as *ni-sun-de-ru* 'I live in.'

We calculated the lexical bundle usage rate for each speaker as the portion of his or her speech contained within lexical bundles. The average rate is 4.2% (SD = 1.78), and ranged from 0.9% to 10.6%. We also calculated the type-token ratio for both lexical bundles and lexical morphemes (common nouns, verbs, and adjectives) for each speaker. The lexical bundle type-token ratios ranged from

23.0% to 95.0%, with an average of 58.0% (SD = 13.8). The speakers produced 381,009 lexical morphemes (42.3% of the data). The lexeme type token ratios ranged from 17.1% to 45.5%, with an average of 29.6% (SD = 5.0).

In order to determine the relationship between age and our three measurements (lexical bundle usage rate, lexical bundle type-token ratio, and lexeme type-token ratio), we conducted three one-way ANOVA tests with a Bonferroni-corrected p value of 0.017. Means (with standard deviations in parenthesis) for each group are presented in Table 1. There are significant differences in the lexical bundle usage rates ($F(3, 148) = 8.39$, $p < .001$). However, the posthoc Tukey test showed that the significant differences were between the older adults and the other groups (all comparisons $p < .01$). None of the other groups differed significantly with each other. In contrast, the ANOVA tests on the lexical bundle and lexeme type token ratios did not reveal significant differences between any of the groups.

These results support the aging hypothesis more than the learning hypothesis. First, the reduced usage of lexical bundles is consistent with the claim that our declarative memory ages. Second, similar to Peelle et al. (2009), we found between-group differences for the older adults only. The middle-aged adults did not show usage patterns that differed from the students, in spite of the fact that the middle-aged adults have two to three times more life experience than the students. But if our declarative memory ages, then why do the older adults not show reduced type token ratios? In the case of lexical bundles, there are so few of them—compare 143 lexical bundles to the 15,298 lexemes used by these speakers—that even an aged memory may easily store them all without loss. More importantly, we need to make a distinction between two types of declarative memory: episodic memory and semantic memory. Previous studies have demonstrated that the episodic memory is prone to aging but semantic memory is not (Nyberg et al., 1996). If lexical bundles rely more on episodic memory than on semantic memory, then we are able to account for the seemingly contracting claims of the aging hypothesis and the learning hypothesis. The aging hypothesis applies primarily to linguistic functions in which episodic memory plays an important role, whereas the learning hypothesis applies to linguistic functions in which semantic memory plays and important role. This distinction between episodic memory and semantic memory, and the implications for language processing, needs to be explored in future research.

Table 1: Mean rates and standard deviations for the three measures by age group

	Students	Younger adults	Middle-aged adults	Older adults
N	49	32	34	37
LB rate (%)	4.28 (1.96)	4.24 (1.53)	4.32 (1.51)	2.73 (1.80)
LB type token rate (%)	55.36 (14.0)	59.42 (12.34)	57.12 (14.91)	60.89 (13.65)
Lexeme type token ratio (%)	29.71 (3.89)	31.47 (5.61)	29.53 (4.99)	28.07 (5.25)

Note. LB = lexical bundle.

References

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at …: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, 25(3), 371-405. doi:10.1093/applin/25.3.371

Deary, I. J., Johnson, W., & Starr, J. M. (2010). Are Processing Speed Tasks Biomarkers of Cognitive Aging? *Psychology and Aging*, 25(1), 219-228. doi:10.1037/a0017750

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, 31(1), 80-93. doi:10.1080/23273798.2015.1047459

Morrison, J. H., & Baxter, M. G. (2012). The ageing cortical synapse: hallmarks and implications for cognitive decline. *Nature Reviews Neuroscience*, 13(4), 240-250. doi:10.1038/nrn3200

Nyberg, L., Bäckman, L., Erngrund, K., Olofsson, U., & Nilsson, L.-G. (1996). Age differences in episodic memory, semantic memory, and priming: relationships to demographic, intellectual, and biological factors. *Journal of Gerontology*, 51(4), 234-240.

Peelle, J. E., Troiani, V., Wingfield, A., & Grossman, M. (2009). Neural Processing during Older Adults' Comprehension of Spoken Sentences: Age Differences in Resource Allocation and Connectivity. *Cerebral Cortex*. doi:10.1093/cercor/bhp142

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The Myth of Cognitive Decline: Non-Linear Dynamics of Lifelong Learning. *Topics in Cognitive Science*, 6(1), 5-42. doi:10.1111/tops.12078

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Language, Memory and Cognition*, 37(3), 776-784.

van Lancker Sidtis, D. (2012). Two-Track Mind: Formulaic and Novel Language Support a Dual-Process Model. In M. Faust (Ed.), *The Handbook of the Neuropsychology of Language* (pp. 342-367). Malden, MA and Oxford: Wiley-Blackwell.

Ullman, M. T. (2016). The Declarative/Procedural Model: A Neurobiological Model of Language Learning, Knowledge, and Use. In G. Hickok & S. L. Small (Eds.), *The Neurobiology of Language* (pp. 953-968). Amsterdam: Elsevier.