ISBN: 978-90-76203-88-1

Effects of Aging and Predictability on Language Processing

Cornelia Moers

MPI
Series

127

Max Planck Institute
for Psycholinguistics

Series

CORNELIA
MOERS

THE NEIGHBORS WILL TELL YOU WHAT TO ________.

EFFECTS OF AGING AND PREDICTABILITY ON LANGUAGE PROCESSING

INVITATION

You are hereby invited to the defense of my doctoral thesis entitled

**THE NEIGHBORS WILL TELL YOU WHAT TO EXPECT:
Effects of Aging and Predictability on Language Processing**

on
WEDNESDAY, OCTOBER 4, 2017
at 14:30 in the aula of the Radboud University Nijmegen, Comeniuslaan 2.

After the defense there will be a reception at the Max Planck Institute, Wundtlaan 1.

CORNELIA MOERS
connymoers@gmail.com

Paranymphs:
CAITLIN DECUYPER
Caitlin.Decuyper@mpi.nl

THORDIS NEGER
Thordis.Neger@mpi.nl

Max Planck Institute
for Psycholinguistics

Series

The Neighbors Will Tell You What To Expect: Effects of Aging and Predictability on Language Processing

Cover design: Constanze Fleischmann, info@designstanze.de

ISBN: 978-90-76203-88-1

Printed and bound by Ipskamp Printing, Nijmegen

The research reported in this thesis was supported by a scholarship from the Max Planck International Research Network on Aging (MaxNetAging).

The Neighbors Will Tell You What To Expect: Effects of Aging and Predictability on Language Processing

Proefschrift
ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op woensdag 4 oktober 2017
om 14.30 uur precies

door
Cornelia Moers
geboren op 13 mei 1984
te Olpe, Duitsland

Promotor:

Prof. dr. A.S. Meyer

Copromotor:

Dr. E. Janse

Manuscriptcommissie:

Prof. dr. M.T.C. Ernestus

Prof. dr. E. Belke (Ruhr-Universität Bochum, Duitsland)

Dr. S.L. Frank

CONTENTS

CHAPTER 1

GENERAL INTRODUCTION

General Introduction

> *“You shall know a word by the company it keeps.”*
> (John Rupert Firth, 1957)

The words that we hear, read, or speak usually occur in the context of other words, which shape our processing and interpretation of the current language input. Hence, the processing of the word “knife” is faster and easier in the context of the predictable sentence “The chef chopped the vegetables with a sharp…”, compared to the more neutral context “The girl does not want to use the …”. Another factor that influences the processing of a word is language experience. Whether a word has been processed oftentimes or whether it has been used barely determines how much effort an individual person needs to process the word. As such, one might find it much easier to process the frequent word “knife” compared to the infrequent word “machete”.

In fact, the two factors – context-dependent processing and language experience – go hand in hand. With increasing language experience we acquire probabilistic knowledge about the contexts in which certain words are likely to occur. Consequently, native speakers typically have strong intuitions about the words that occur close or immediately next to each other. Young infants already start to exploit co-occurrence patterns in their language input. In adulthood, people may use the statistically salient patterns to generate predictions about the words that are likely to occur in order to facilitate their lexical processing.

By definition, language experience increases over the life course. Therefore, one might assume that older adults are advantaged over younger adults in their processing of context information. However, in what ways older adults differ from younger adults in probabilistic and predictive language processing is not straightforward. For example, older adults have been reported to show *larger* as well as *smaller* predictability effects during language processing than younger adults. Therefore, one of the key questions that the research in this thesis tackles is how aging and language experience influence the processing of word co-occurrence information. As such, the thesis combines three recent lines of research on human cognition: research on how language users perform prediction, how probabilistic knowledge influences language processing, and how cognitive aging and experience affect language processing. Crucially, the empirical research in this thesis is relevant and innovative in that it combines the three areas, leading to a deeper understanding of the interactions between the mechanisms in language processing over the life span. Before outlining the individual thesis chapters, I will briefly summarize the current state of research in each area.

Predictions in Language Processing

During language processing, people can anticipate likely upcoming words. Knowledge about the contexts in which certain words commonly occur supports generating predictions about upcoming linguistic input (for a recent discussion on predictive language processing see e.g., Huettig, 2015; also Federmeier, 2007). Predictive processing may facilitate language tasks, for example conversations in which listeners may anticipate the interlocutor’s words in order to efficiently process and integrate them into their own utterance planning (e.g., Barthel, Sauppe, Levinson, & Meyer, 2016; Magyari & de Ruiter, 2012). A well-established finding in prediction research is that predictable words receive some processing benefit relative to unpredictable words. Hence, words that are highly predictable from their context can be recognized faster in speech comprehension and reading and can be named faster in speech production tasks relative to less predictable words (e.g., Bell et al., 2003; Frisson, Rayner, & Pickering, 2005; Griffin & Bock, 1998; McDonald & Shillcock, 2003; Van Berkum et al., 2005). Some researchers even hypothesize that the brain is essentially a predictive machine (e.g., Clark, 2013), and that the use of probabilistic knowledge gained through experience is one of the basic mechanisms for fast and efficient processing of information across many domains (Chater & Oaksford, 2008). Elaborate language models assume that language acquisition, language comprehension and language production are linked by the mechanism of prediction. On these accounts, language comprehenders are thought to use their speech production system internally in order to generate representation of likely upcoming words and adapt representations (i.e., learn) from prediction errors (e.g., Dell & Chang, 2014; Pickering & Garrod, 2007 & 2013). Yet, many questions about predictive processing remain. One example is whether prediction is always mandatory during language processing or whether people optionally engage in predictive processing strategies. Moreover, if predictive processing were optional, it is unclear which conditions (such as task conditions, or language user abilities) might encourage setting up predictions. Another key issue, which is also addressed in this thesis (cf. Chapter 5), is about the type of predictions, as different types of context information may be differentially used in predictive processing. The type of context information may range from simple co-occurrence statistics of words to complex semantic information provided in a complete sentence (e.g., Hahn, 2012; Shaoul, Baayen, & Westbury, 2014; Smith & Levy, 2011; van Petten, 2014), and may thus be more or less difficult to use in order to generate predictions.

The influence of probabilistic knowledge on lexical processing

Probabilistic knowledge influences the speed and efficiency of lexical access. Numerous studies report that lexical processing is faster for frequent words and word combinations than for infrequent linguistic units (e.g., Arnon & Cohen Priva, 2013; Hauk & Pulvermüller, 2004; Janssen & Barber, 2012; Jescheniak & Levelt, 1994; Kliegl, Grabner, Rolfs, & Engbert, 2004; Navarrete, Basagni, Alario, & Costa, 2006; Revill & Spieler, 2012; Sereno, Rayner, & Posner, 1998; Spieler & Balota, 2000). One probabilistic measure of how likely a word is to occur in a particular context is Transitional Probability (TP). TP reflects the likelihood of a word given the right or left neighboring word and can be obtained from frequency counts in large-scale corpora. Language users are sensitive to this kind of probabilistic information. Hence, TPs influence reading times in silent reading, such that contextually likely words are fixated for shorter periods of time and skipped more often than words with lower likelihood of occurrence

(Frisson, Rayner & Pickering, 2005; McDonald & Shillcock, 2003). In speech production, TPs influence articulatory effort and precision, such that high-TP words are acoustically more reduced than low-TP words (e.g., Bell et al., 2003), which has been termed *Probabilistic Reduction* (e.g., Jurafsky, Bell, Gregory, & Raymond, 2001).

One principle that can explain how probabilistic reduction in articulation occurs is interactivity between processing levels. Higher-level knowledge, such as probabilistic knowledge about the likelihood of a word given a neighboring word, may influence lower processing levels, such as articulation, via the level of lexical access. Take the example of speech production, for which we may assume that the processing levels are not discrete stages, but overlap and interact with each other (e.g., Goldrick & Blumstein, 2006). That means, if lexical access is more difficult, as for contextually unlikely words, the disruptions encountered during lexical access will cascade to the articulatory level, resulting in difficulties in articulatory processing (manifested in lengthening and/or distortion of pronunciation). As disruptions in lexical access are more pervasive in older adulthood (cf. Burke & Shafto, 2008), one question that remains is whether and how adult aging influences this interactivity between planning and acoustic realization.

Cognitive aging research

There is evidence that older adults are more variable in their cognitive performance than younger adults (cf. Salthouse, 2010; Ramscar et al., 2014). The greater cognitive variability in later adulthood affects language processing in various ways. Among the processes that seem to deteriorate in healthy older adults is word retrieval. Older adults more often report tip-of-the-tongue states and they are slower and less accurate in picture naming than younger adults (cf. Abrams & Farrell, 2011), which reflects a decline in the speed and efficiency of lexical processing. However, not all aspects of language worsen with age. Particularly, measures associated to language experience and stored or crystallized knowledge, such as vocabulary size and the richness of semantic and lexical networks, seem to be preserved or even increase with age (e.g., Ben-David, Erel, Goy, & Schneider, 2015; Verhaeghen, 2003).

With their life-long language experience, older adults may capitalize on their rich semantic and contextual knowledge in order to engage in predictive processing. Indeed, there is evidence for a shift to greater reliance on sentence context in older compared to younger adults. For example, if listening conditions are made equally difficult for younger and older listeners by presenting speech against age-appropriate noise levels, older adults use contextual information more than younger adults for speech comprehension (e.g., Pichora-Fuller, Schneider & Daneman, 1995; Sheldon, Pichora-Fuller, & Schneider, 2008). Older adults have also been shown to rely more heavily on context than younger adults during reading in order to adapt to difficulties with word recognition and lexical processing (Rayner et al., 2006). Yet, other studies report the reverse. Namely, older adults engage less than younger adults in the use of context in order to predict likely upcoming words during language comprehension (e.g., Federmeier, Kutas, & Schul, 2010). This discrepancy in findings may relate to the time course of context use. If older people need more time for semantic integration (e.g., Payne & Stine-Morrow, 2012) and hence for setting up predictions than younger adults, findings may depend on whether the task involves time pressure. Moreover, predictive processing may require high levels of cognitive capacity, such as working memory, which mainly younger adults or some older adults with well-preserved cognitive functioning can meet (cf. Federmeier et al., 2010; Janse & Jesse, 2014). The mixed results reported across studies call for further research into possible age-related changes in predictive processing and their underlying causes. This thesis focuses on the investigation of potential age differences in the processing of probabilistic co-occurrence patterns between words, which has not been studied before.

Thesis Outline

The empirical studies reported in the following chapters share the scientific objective of scrutinizing interactions of predictive processing, probabilistic knowledge and age-related changes in language use.

For the research reported in Chapter 2, a corpus study was conducted, which investigated word frequency and predictability effects on word durations in the domain of reading aloud. The main research questions for this study were whether inter-word predictability effects generalize to read aloud speech of different age groups and whether predictability effects change with increasing age and reading experience. From previous research it was not clear whether and how probabilistic acoustic reduction would change across the life span. Therefore, in my study frequency and TP effects were analyzed for three different age groups: children, adolescents, and older adults. Unfortunately, the Dutch JASMIN corpus, from which the speech samples were taken, did not include a reference group of university students or middle-aged adults. Therefore, I investigated probabilistic effects within each individual age group, as well as in an age-group comparison between older adults versus adolescents. The corpus study allowed me to conclude that age and differential language experience may indeed modify probabilistic processing, but group differences were small.

From the corpus study one could not be sure at which processing level age differences in probabilistic effects arise, as reading aloud involves language comprehension and language production. Furthermore, I did not have access to relevant background information about the speakers, for instance their vocabulary size, which had been shown to affect probabilistic effects. I therefore designed a follow-up experiment, in which I investigated age and individual differences in frequency and co-occurrence effects in silent reading versus reading aloud using an eye-tracking paradigm. The main research questions targeted two points: First, by comparing probabilistic effects between silent reading and reading aloud, I aimed to pinpoint whether participants made more use of predictability when solely perceiving words or when additionally producing them. Second, I planned to clarify whether age groups show differential frequency and predictability effects in either silent or oral reading, or both. Younger and older adults read sentences that contained Dutch noun-verb combinations (such as “muziek spelen” - to play music) varying in frequency and co-occurrence predictability (TP). In Chapter 3 of this thesis, I report the analysis of the speech production data from the reading aloud task to study potential age differences in frequency and TP effects on spoken word durations. The analysis was similar to the word duration analysis in the corpus study of Chapter 2, but now the study involved a group comparison of older adults versus university students, and more controlled reading materials. Consequently, a more complete picture of age-related differences in probabilistic reduction could be obtained.

Subsequently, Chapter 4 of this thesis reports the analysis of the eye-tracking data from the experiment reported in Chapter 3. Previous research showed that word frequency and predictability affect initial reading of words in a sentence as well as later processing of the sentence content. Moreover, in order to get a complete picture of age differences in reading – with aging potentially affecting initial and later reading processes – my analysis included several eye fixation measures to capture different reading stages.

The study described in Chapter 5 was conducted during a research visit to the Linguistic Department of Northwestern University. The study focused on age differences in contextually constrained speech production, and differed from the preceding chapters in two important aspects. First, in contrast to the previous chapters, the task involved picture naming, which necessarily includes conceptual (semantic) processing, relative to reading. Second, in this experiment I contrasted two different types of predictability: global sentence predictability versus local word co-occurrence predictability. The main research questions were: First, whether cloze predictability and transitional probability both affect experimentally controlled speech production. Second, whether younger and older adults differ in how much they benefit from global cloze predictability, or from local word co-occurrence predictability. In Chapter 3 and Chapter 4, in which I had solely investigated TP effects, there was no difference for younger versus older adults in their processing of local predictability. However, other researchers had found age-related changes in predictability effects when measuring sentence predictability using cloze probabilities (e.g., Federmeier, McLennan, De Ochoa, & Kutas, 2002). I conducted a contextually-constrained picture naming experiment, which was designed to test the hypothesis that age may interact with the type of information that people readily process, such that younger and older adults differ in their use of cloze predictability, but not so much in their processing of TP predictability. I also assessed participant's vocabulary size in order to test for a link between age, vocabulary knowledge and the size of predictability effects.

In Chapter 6 I summarize the findings of the empirical studies, discuss their broader theoretical implications and provide an outlook for future research.

References

Abrams, L., & Farrell, M.T. (2011). Language Processing in Normal Aging. In Guendouzi, J., Loncke, F., & Williams, M.J. (Eds.), *The Handbook of Psycholinguistic and Cognitive Processes: Perspectives in Communication Disorders* (pp. 49-74). New York: Psychology Press.

Arnon, I., & Cohen Priva, U. C. (2013). More than Words: The Effect of Multi-word Frequency and Constituency on Phonetic Duration. *Language and Speech*, 56(3), 349-371. doi: 10.1177/0023830913484891

Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The Timing of Utterance Planning in Task-Oriented Dialogue: Evidence from a Novel List-Completion Paradigm. *Frontiers in Psychology*, 7, 1858. doi: http://doi.org/10.3389/fpsyg.2016.01858

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, 113(2), 1001-1024. doi: http://dx.doi.org/10.1121/1.1534836

Ben-David, B. M., Erel, H., Goy, H., & Schneider, B. A. (2015). “Older is always better”: Age-related differences in vocabulary scores across 16 years. *Psychology and Aging*, 30(4), 856-862. doi: 10.1037/pag0000051

Burke, D.M. & Shafto, M.A. (2008). Language and aging. In F.I.M. Craik & T.A. Salthouse (Eds.), *The Handbook of Aging and Cognition* (pp. 373-443). New York: Psychology Press.

Chater, N. & Oaksford, M. (2008). The probabilistic mind: Where next? In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 501–14). Oxford: Oxford University Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences,* 36(3), 181–253. doi: 10.1017/s0140525x12000477

Dell, G. S., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society London,* 369(1634), 20120394. doi: 10.1098/rstb.2012.0394

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, 39(2), 133-146. doi: 10.1017/S0048577202001373

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491-505. doi: 10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115(3), 149-161. doi: http://dx.doi.org/10.1016/j.bandl.2010.07.006

Firth, J.R. (1957). *A Synopsis of Linguistic Theory, 1930-1955*. Studies in linguistic analysis, Oxford: Blackwell.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of Contextual Predictability and Transitional Probability on Eye Movements during Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 31(5), 862-877. doi: 10.1037/0278-7393.31.5.862

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, 21, 649-683. doi: 10.1080/01690960500181332

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and levels of processing in spoken word production. *Journal of Memory and Language*, 38(3), 313-338. doi: http://dx.doi.org/10.1006/jmla.1997.2547

Hahn, L. (2012). Measuring local context as context–word probabilities. *Behavior Research Methods*, 44(2), 344-360. doi: 10.3758/s13428-011-0148-y

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, 115(5), 1090-1103. doi: http://dx.doi.org/10.1016/j.clinph.2003.12.020

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, 1626, 118-135. doi: 10.1016/j.brainres.2015.02.014.

Janse, E., & Jesse, A. (2014). Working memory affects older adults' use of context in spoken-word recognition. *Quarterly Journal of Experimental Psychology*, 67(9), 1842-1862. doi: 10.1080/17470218.2013.879391

Janssen, N., & Barber, H. A. (2012). Phrase Frequency Effects in Language Production. PLoS ONE, 7(3): e33202. doi: 10.1371/journal.pone.0033202

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 824-843. doi: 10.1037/0278-7393.20.4.824

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In J. L. Bybee & P.J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229-254). Amsterdam: John Benjamins.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16, 262-284. doi: 10.1080/09541440340000213

Magyari, L., & de Ruiter, J. P. (2012). Prediction of Turn-Ends Based on Anticipation of Upcoming Words. *Frontiers in Psychology*, 3, 376. doi: http://doi.org/10.3389/fpsyg.2012.00376

McDonald, S. A., & Shillcock, R. C. (2003). Eye Movements Reveal the On-Line Computation of Lexical Probabilities During Reading. *Psychological Science*, 14(6), 648-652. doi: 10.1046/j.0956-7976.2003.psci_1480.x

Navarrete, E., Basagni, B., Alario, F. X., & Costa, A. (2006). Does word frequency affect lexical selection in speech production? *The Quarterly Journal of Experimental Psychology*, 59(10), 1681-1690. doi: 10.1080/17470210600750558

Payne, B. R., & Stine-Morrow, E. A. (2012). Aging, parafoveal preview, and semantic integration in sentence processing: testing the cognitive workload of wrap-up. *Psychology and Aging*, 27(3), 638-649. doi: 10.1037/a0026540

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, 97(1), 593–608. doi: 10.1121%2F1.412282

Pickering, M. J. & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences,* 11(3), 105–110. doi: http://dx.doi.org/10.1016/j.tics.2006.12.002

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329-347. doi: 10.1017/s0140525x12001495

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

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, 21(3), 448-465. doi: 10.1037/0882-7974.21.3.448

Revill, K. P., & Spieler, D. H. (2012). The effect of lexical frequency on spoken word recognition in young and older listeners. *Psychology and Aging,* 27(1), 80-87. doi: 10.1037/a0024113

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society, 16*(5), 754-760. doi: 10.1017/s1355617710000706

Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *Neuroreport*, 9(10), 2195-2200.

Shaoul, C., Harald Baayen, R., & Westbury, C. F. (2014). N-gram probability effects in a cloze task. *The mental lexicon,* 9(3), 437–472. doi: 10.1075/ml.9.3.04sha

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *Journal of the Acoustical Society of America*, 123, 489–499. doi: 10.1121%2F1.2783762

Smith, N. J., & Levy, R. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pp. 1637-1642.

Spieler, D. H., & Balota, D. A. (2000). Factors influencing word naming in younger and older adults. *Psychology and Aging,* 15(2), 225-231. doi:10.1037/0882-7974.15.2.225

Van Berkum, J. J. A., Brown, M. C., Zwitserlood, P., Kooijman, V. & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443–467. doi: 10.1037/0278-7393.31.3.443

Van Petten, C. (2014). Examining the N400 semantic context effect item-by-item: Relationship to corpus-based measures of word co-occurrence. *International Journal of Psychophysiology*, 94(3), 407-419. doi: http://dx.doi.org/10.1016/j.ijpsycho.2014.10.012

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging,* 18(2), 332-339. doi: 10.1037/0882-7974.18.2.332

CHAPTER 2

EFFECTS OF WORD FREQUENCY AND TRANSITIONAL PROBABILITY ON WORD READING DURATIONS OF YOUNGER AND OLDER SPEAKERS

This chapter has been adapted from:
Moers, C., Meyer, A.S., & Janse, E. (2016). Effects of Word Frequency and Transitional Probability on Word Reading Durations of Younger and Older Speakers. *Language and Speech*, 60(2), 289-317. doi: 10.1177/0023830916649215

Abstract

High-frequency units are usually processed faster than low-frequency units in language comprehension and language production. Frequency effects have been shown for words as well as word combinations. Word co-occurrence effects can be operationalized in terms of transitional probability (TP). TPs reflect how probable a word is, conditioned by its right or left neighboring word. This corpus study investigates whether three different age groups – younger children (8-12 years), adolescents (12-18 years) and older (62-95 years) Dutch speakers – show frequency and TP context effects on spoken word durations in reading aloud, and whether age groups differ in the size of these effects. Results show consistent effects of TP on word durations for all age groups. Thus, TP seems to influence the processing of words in context, beyond the well-established effect of word frequency, across the entire age range. However, the study also indicates that age groups differ in the size of TP effects, with older adults having smaller TP effects than adolescent readers. Our results show that probabilistic reduction effects in reading aloud may at least partly stem from contextual facilitation that leads to faster reading times in skilled readers, as well as in young language learners.

Effects of Word Frequency and Transitional Probability on Word Reading Durations of Younger and Older Speakers

Introduction

During the course of our lives we acquire probabilistic knowledge on how often certain linguistic units occur. These units may vary in size and probabilistic knowledge therefore consists of knowledge about the frequency of phonemes, syllables, words, phrases, and sentence types. Knowledge about how often words occur and co-occur gradually builds up with language experience. Consequently, frequency and predictability effects may change with age (e.g., Gollan, Montoya, Cera, & Sandoval, 2008; Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006). Only a few studies have investigated age-related changes in probabilistic reduction effects, however. Probabilistic reduction is the acoustical reduction of frequent or predictable words in speech. In particular, to the best of our knowledge, no study has investigated word occurrence and co-occurrence effects in the demanding task of reading aloud for readers of different ages. There is some evidence that probabilistic reduction occurs in reading aloud, but it is unclear whether these effects hold for different age groups. Age groups may differ both in the comprehension and production processes needed for reading aloud. Consequently, probabilistic reduction effects may differ in size if children or older adults are reading out loud, compared to the students typically studied in previous research. The Dutch JASMIN corpus contains read aloud speech by younger children and adolescents, as well as older adults. The corpus therefore provides an excellent starting point for exploring how co-occurrence frequency modulates reading durations with increasing language experience. Below, we will discuss the skill of reading aloud in more detail, then review relevant literature on frequency and predictability effects, and describe how and why these probability effects may differ between age groups.

Reading aloud is a complex skill

Reading research has shown that reading aloud is a complex and demanding task. The visual recognition of words, the planning of speech and the actual articulation of the read words need to be synchronized efficiently to enable fluent oral reading (e.g., Breznitz & Berman, 2003). Furthermore, efficient reading draws on general cognitive capacities and limits the resources available for other tasks, for instance in dual-task situations (Kemper et al., 2014). Importantly, despite more than 100 years of reading research (e.g., Anderson & Swanson, 1937; Buswell, 1921; Huey, 1908), knowledge about the successful interplay between the sub-components involved in reading aloud is scarce. Reading researchers have proposed detailed models about the processes of visual perception, word identification, semantic integration and speech production. However, these models typically explain only one or two subcomponents of the reading process (cf. Rayner & Reichle, 2010). Specifically, many studies have explained and modelled single-word naming, using well-known computational models such as, for instance, the Dual-Route Cascaded model (DRC; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) or the Parallel Distributed Processing model (PDP; Seidenberg & McClelland, 1989). So far, these models do not account for inter-word context effects. In reading aloud a tight synchronization between the progress of visual word identification and subsequent articulation needs to be kept, such that the voice does not lag behind the eyes too much. The eyes will usually be slightly ahead of the voice in reading aloud (about 500 milliseconds, cf. Inhoff, Solomon, Radach, & Seymour, 2011). One reason for this may be that the execution of motor commands is a relatively slow process. Thus, the additional articulation of words in reading aloud needs more time than word recognition per se. Furthermore, readers may need to look ahead in order to generate an appropriate intonation contour during reading aloud. Additionally, readers may be able to speak much faster than they do when reading aloud, but they choose not to for communication or coordination reasons. One goal of the present study is to contribute to a better understanding of the reading aloud process by providing descriptive evidence about the way the task of reading aloud is accomplished by readers in different age groups and varying linguistic experience.

Frequency and predictability effects in tasks related to reading aloud

Generally, word frequency is a robust predictor of processing speed across different language tasks. In language production, pictures with high-frequency names are named faster than pictures with low frequency names (e.g., Jescheniak & Levelt, 1994; Navarrete, Basagni, Alario, & Costa, 2006) and naming single printed words is faster for high-frequency words than for low-frequency words (e.g., Spieler & Balota, 2000). In language comprehension, there is a large body of evidence that shows that high-frequency words are recognized faster than low-frequency words (e.g., Allen, Smith, Lien, Grabbe, & Murphy, 2005; Hauk & Pulvermüller, 2004; Revill & Spieler, 2012; Sereno, Rayner, & Posner, 1998). In silent reading of sentences, numerous eye-tracking studies have shown that high-frequency words are fixated for shorter periods of time and skipped more often than low-frequency words (e.g., Hand, Miellet, O'Donnell, & Sereno, 2010; Kliegl, Grabner, Rolfs, & Engbert, 2004; see also Rayner, 1998).

Importantly, frequency effects in language production and comprehension also occur at the level of processing word combinations and multi-word sequences (e.g., Arnon & Cohen Priva, 2013). The dependency of a word on its immediate context can be operationalized as transitional probability (TP): TP reflects an estimate of how likely a word occurs, given its right or left neighboring word. In silent reading, student readers spend more time reading words that are less predictable from their local context than on highly predictable words (for TP effects in reading see McDonald & Shillcock, 2003a, 2003b; Wang, Pomplun, Chen, Ko, & Rayner, 2010; for more general studies on predictability effects in reading see Hand et al., 2010; Rayner & Clifton, 2009; Whitford & Titone, 2014; for self-paced reading times see Smith & Levy, 2011; for a general discussion see Smith & Levy, 2013). Note that Frisson, Rayner, and Pickering (2005) argue that TP effects are not independent from regular predictability effects (as measured by a cloze task), but are rather part of sentence-based predictability effects. Probabilistic information in the form of word-to-word contingency statistics is available from a user's experience with language. TP effects are therefore indicators of predictive processing in silent reading, as they reflect the ease or difficulty with which an upcoming word can be processed given local context.

Transitional probability effects have also been demonstrated in language production, such that frequently co-occurring words are pronounced with less effort and undergo more articulatory reduction than less frequent combinations (e.g., Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Pluymaekers, Ernestus, & Baayen, 2005b). These studies analysed corpora of conversational speech and lead to the formulation of the Probabilistic Reduction Hypothesis (Jurafsky, Bell, Gregory, & Raymond, 2001), which states that more probable items (e.g., words, phrases, and syntactic constructions) are more reduced. Probabilistic reduction processes may involve segment deletion (e.g., Jurafsky et al., 2001), weakening of vowels (Hanique, Schuppler, & Ernestus, 2010), shortening of syllables (Aylett & Turk, 2004) or overall word duration (e.g., Bell et al., 2009; Moers, Janse, & Meyer, 2015) and have been confirmed for a variety of word classes, word forms, and syntactic constructions (e.g., Bell et al., 2003; Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999; Jurafsky et al., 2001; Pluymaekers, Ernestus, & Baayen, 2005a; see also: Gahl & Garnsey, 2004; Kuperman & Bresnan, 2012; Shriberg & Stolcke, 1996; Tily et al., 2009). According to Bell et al. (2009) probabilistic reduction effects are found for both content and function words, but the two word classes are differentially affected by forward TP (FTP) and backward TP (BTP), and only content words are influenced by word frequency.

In sum, more frequent and more predictable linguistic units are easier to process than infrequent or less predictable units, and these frequency and predictability effects occur in language comprehension as well as production. As reading aloud involves both comprehension and production, probabilistic facilitation should also be seen in this task. Even though probabilistic effects in a clear reading aloud style may be somewhat smaller than in more spontaneous speech or less formal reading (Baker & Bradlow, 2009; Hanique & Ernestus, 2011), multiple studies have replicated Lieberman's (1963) finding of probabilistic acoustic reduction effects in reading aloud (Clopper & Pierrehumbert, 2008; Gahl & Garnsey, 2004). Crucially, though, the question remains whether probabilistic reduction effects in reading aloud hold for different age groups. Age-related differences in both comprehension and production processes may influence probabilistic reduction, such that probabilistic reduction (and/or lengthening for unlikely words) may vary to a great extent in children or older adults compared to the students typically studied in previous research.

Age-related differences in frequency and predictability effects

The transfer of frequency and predictability effects to reading aloud in different age groups might not be straightforward given that the synchronization of the different processes makes reading aloud complex and possibly demanding. This may particularly be the case for groups that are learning how to read and who have not yet optimized the individual components (children), or for people possibly facing some cognitive decline (e.g., older adults). In this exploratory study we investigate how probabilistic reduction effects develop across different ages. The JASMIN corpus enables us to study reading aloud in three different groups–children, adolescents, and older adults–and therefore provides an excellent starting point for answering this question.

Frequency and predictability effects are part of probabilistic knowledge which starts to build up in childhood and when acquiring a new language (e.g., Chater & Manning, 2006; Saffran, Aslin, & Newport, 1996; Thompson & Newport, 2007). Word frequency has been shown to affect reading speed already in young learning readers. School children rely heavily on frequency knowledge when optimizing lexical processing in either silent reading (see Joseph, Nation, & Liversedge, 2013) or reading aloud (Valle, Binder, Walsh, Nemier, & Bangs, 2013; Vorstius, Radach, & Lonigan, 2014). Moreover, age-of-acquisition of a word is an important variable in reading, both for children and across adulthood (e.g., Coltheart, Laxon, & Keating, 1988; Davies, Arnell, Birchenough, Grimmond, & Houlson, 2017; Morrison, Hirsh, Chappell, & Ellis, 2002; Zevin & Seidenberg, 2004), but note that this variable is correlated with frequency (for a recent discussion see Ambridge, Kidd, Roland, & Theakston, 2015). Importantly, in sentence reading, the context in which a word may occur plays a role, making single-word factors such as frequency and age-of-acquisition less important. School children are sensitive to the contextual diversities (i.e., the number of textbooks in which a given word appears), showing distributional knowledge that goes beyond mere exposure of single words (Perea, Soraes, & Comesana, 2013). Furthermore, Calfee and Drum (1986) showed that words are read aloud more quickly when placed in context rather than in isolation, which implies that semantic clues provided by context benefit reading in children. Yet, to the best of our knowledge, no results have been published on local predictability effects for word combinations in children or adolescents in either silent or oral reading. Thus, it is unclear whether TPs between words affect word pronunciation beyond word frequencies in childhood and adolescence. In our study we analysed read speech samples of young children (8-12 years) and adolescents (12-18 years). Based on findings of consistent frequency effects in previous studies, we predicted that we should find reliable word frequency effects in both age groups. With more reading experience, the learner's processing span may often go beyond single words (cf. Blythe & Joseph, 2011), such that we expect word reading times in children and adolescents to be influenced by predictability. We predict this holds for both the predictability of a target from the preceding neighboring word and from the following neighboring word, as children might expand their visual and predictive processing span to the left and the right. Furthermore, as frequency and predictability effects are based on distributional knowledge which grows with experience, we investigate whether probabilistic effects on word durations change in size with child age.

The interdependence of language experience and probabilistic effects is also relevant for aging research. Distributional knowledge about how often words occur and co-occur is based on language experience and hence subject to constant change (Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014). Older adults have years of experience with the words in their native language. Probabilistic effects can be expected to be stronger in older compared to younger readers, if only because corpus-based probabilistic measures better approximate the more experienced reader's expectations about which words are frequent or predictable. On this account, one would expect stronger associations between word durations and corpus measures with increasing age.

So far, however, existing studies have not yielded consistent results and led to two contrasting hypotheses with regard to age-related (or experience-related) changes in frequency effects. Note, however, that neither type of accounts provides a detailed specification of the mechanisms underlying these changes. The first type of accounts assume that more language experience and vocabulary growth with increasing age (cf. Ben-David, Erel, Goy, & Schneider, 2015; Hartshorne & Germine, 2015) yield more lexical competition between words. Having a larger vocabulary typically involves knowing more low-frequency words. Low-frequency words will thus have more low-frequency competitors and the target activation takes longer. This increases the difference in activation speed between high- and low-frequency words for those with increasing language experience (Balota & Ferraro, 1996; Revill & Spieler, 2012; Spieler & Balota, 2000). In line with this account, older adults show stronger word frequency effects than younger adults in eye fixation data in silent reading (Rayner et al., 2006).

However, "lexical entrenchment" accounts for bilingual processing (Diependaele, Lemhöfer, & Brysbaert, 2012; Gollan et al., 2008; and see e.g., Andrews & Hersch, 2010 for experience differences in native language processing) make the opposite prediction: that is, increased language experience should lead to decreased frequency effects because language experience enhances the "entrenchment" of lexical representations. This entrenchment implies that lexical memory representations become more precise with increasing language experience. Higher precision is associated with a better integration of (or better links between) orthographic, phonological, and semantic information in memory (cf. Perfetti, 1997). In an activation-based model lexical precision and information integration in memory can be implemented as the strength of between-word lexical inhibition. That is, for an inexperienced reader, it is easy to activate high-frequency target words and inhibit these targets' competitors, because their competitors are relatively weak. On the other hand, it is hard to activate low-frequency words because their competitors are relatively strong. Hence, there will be large frequency effects. For an experienced reader, high lexical precision results in less interference from co-activated representations, such that the difference between high and low-frequency words diminishes. If we apply these accounts to age-related changes in first language processing, increased language experience with age should lead to decreased frequency effects.

In line with probabilistic effects becoming smaller with advanced adult age, multiple studies have shown that older adults engage less in anticipation of upcoming words in language comprehension than younger adults (e.g., DeLong, Groppe, Urbach, & Kutas, 2012; Federmeier, Kutas, & Schul, 2010). Again, it is unclear whether decreased predictive processing in older adults will generalize to effects of TP. That is, integrating information across longer distances (e.g., entire sentences) in order to make inferences about upcoming words may be more costly with advanced age than taking advantage of local word-to-word statistics due to age differences in working memory capacity (cf. Wingfield, Alexander, & Cavigelli, 1994). Furthermore, in the reading study by Rayner and colleagues (2006) older adults showed stronger word frequency effects than younger adults in eye fixation data, but predictability effects in that same study (predictability measured by a cloze task) were similar in size in the two age groups. A priori, it is unclear whether age differences in TP effects will pattern with age differences in word frequency effects or in predictability effects. As such, the Rayner study does not yield unambiguous predictions for our study. Therefore, we investigate whether older seniors will differentially use TP cues compared to younger seniors and in comparison to adolescents, particularly in the complex task of reading aloud with its high demands on synchronization of different processes.

The current study

The present study is a corpus-based investigation of how probability measures affect word durations for readers of different ages when reading sentences aloud. Local predictability was measured as a combined predictor of FTP (probability of a target word given the preceding word) and BTP (probability of a target word given the following word) (cf. Jurafsky et al., 2001). We combined the two types of TPs for two reasons: first, because our aim was to investigate age-related changes in inter-word predictability (both FTP and BTP), rather than to evaluate whether one type of TP was more influential than the other; and second, because FTP and BTP were highly correlated. The Dutch JASMIN corpus contains reading aloud data of a sample of children, of adolescents and a sample of older adults. This corpus allows us to explore the following research questions:

(1) Do predictability effects, as found in silent reading and language production, generalize to read aloud speech of different age groups? This question will be answered in the following sections (Section A-C) by investigating whether we find simple effects of TP, over and above simple word frequency effects, within and across age groups.

(2) Do frequency and predictability effects change in young readers with increasing age? This question will be addressed in Section A (“Effects in children and adolescents”) by investigating age interaction effects for frequency and TP within young school children (aged 8-12) and within adolescents (aged 12-18). Younger children read different texts chosen according to their reading proficiency. Adolescents were expected to all have similar levels of reading proficiency and therefore all read the same text. Taking these

properties of the materials into account, we investigated effects mainly within each group, rather than across groups.

(3) Do frequency and predictability effects change in size in read aloud speech among readers of advanced age? This question is answered in Section B ("Effects within older readers") by investigating TP-by-age interactions among a sample of readers aged from 62 to 95 years.

(4) Do older adults differentially benefit from predictability compared to adolescents? We investigate this question in Section C ("Age group comparison"), by comparing predictability effects in older adults (aged above 62) to the effects in adolescents (aged 12-18), who all read the same text. Using this age-group comparison data sample, we additionally investigated whether predictability effects were different for content versus function words. This analysis is reported in Section D ("Content versus function words").

General methods

Samples and materials

The study reports the effects of relative word frequency, local predictability and various control variables on word durations in reading aloud. Samples from three different groups–children, adolescents, and older adults–were drawn from the JASMIN-CGN corpus in Dutch (henceforth: JASMIN, for detailed information on the corpus see Cucchiarini, Driesen, Van Hamme, & Sanders, 2008), which is an extension to the larger Corpus of Spoken Dutch (Corpus Gesproken Nederlands, CGN; e.g., Oostdijk, 2002). As the JASMIN corpus does not contain recordings of native speakers of Dutch between the ages of 18 and 62, we cannot report any analysis for younger or middle-aged adults. Our first group consists of 61 school children (31 female, 30 male) aged 8 to 12 (M = 10.34, standard deviation (SD) = 1.41). Seven children were excluded in advance because their reading was extremely disfluent. Children of different ages read different texts matched to their reading proficiency level (Levels 1-9), as defined by the Dutch primary school reading programme Veilig Leren Lezen ('Learning to read safely'). The second group consists of 41 adolescents (21 female, 20 male) aged 12 to 18 years (M = 14.07, SD = 1.69; none were excluded). The third group comprises 64 older adults from the total of 68 older adults (41 female, 23 male) in the age range of 62 to 95 years (M = 78.66, SD = 8.41). Four older adults were excluded because their speech was very disfluent or because they did not read the entire text. Adolescents and older adults (group 2 and group 3) read aloud the same text, which contained phonetically rich sentences describing local traditions or facts about The Netherlands (e.g., how birthdays are celebrated). The recording sessions for the children and adolescents took place in schools. Recordings of the older adults were made in the participants' homes. For all three groups, recordings were made in a quiet room using high-quality recording equipment. Texts were presented on a computer screen and the speakers were asked to read out loud the sentences at their normal reading speed. At the end of the session speakers received a small gift for participating in the recording (such as a cinema ticket or book voucher). For each recording transcriptions on various levels are available within the corpus. Orthographic transcriptions were provided by one trained transcriber and then checked by a second trained transcriber. They are consistent with Dutch spelling and pronunciation conventions, as described in the JASMIN documentation. Furthermore, phoneme-level annotations were generated automatically with HMM-based alignment (Viterbi), adopting the CGN conventions and base lexica for native Dutch speakers. These annotations and transcriptions were used to compute the dependent and independent variables with the help of automated Perl or Praat scripts (Praat 5.3.; Boersma & Weenink, 2012). Furthermore, the transcriptions included part-of-speech tags (POS tags), which we used to differentiate between content and function words.

Computation of control variables and frequency measures

Several factors have been shown to influence word duration in addition to the predictability variables described above. Consequently, appropriate control variables were added. The data selection procedure will briefly be described in the following section. As a first step, words that occurred in disfluent surroundings – displaying either self-repairs, hesitations or restarts – were excluded from the analyses. The position of the word in the phrase is another variable taken into account, as previous research found effects of phrase-initial or phrase-final lengthening (e.g., Beckman & Edwards, 1990; Fougeron & Keating, 1997). Thus, in a second step, all items occurring in speech chunk-initial or chunk-final position were excluded, also because no TP values can be computed for these words. A speech chunk, or interpause stretch, was defined as a stream of fluent speech between pauses of more than 200 milliseconds (ms) (cf. Trouvain, 2003; Trouvain & Grice, 1999).

Two factors that will affect the duration of a target word are word length and local speech rate. Words with few letters are usually read faster than words with more letters. Secondly, faster speech rate automatically leads to shorter word durations. Speech rate and word length were combined into one control variable, expected word duration. To derive this variable, we multiplied the local speech rate of the speech chunk the word occurred in (average millisecond-per-phoneme over the speech chunk) with the number of phonemes of the target word. Broad phonetic transcriptions were used to count the number of phonemes of target words. These were part of the CGN annotations, in which an automatic speech aligner selected the best-matching standard pronunciation variant from the CGN lexicon (e.g., it is acceptable in Dutch to pronounce verbs such as “lopen” (to walk) without the final phoneme /n/; both pronunciation variants with and without final /n/ were part of the lexicon the HMM-based aligner could select from). Thus, the control variable reflects how long (in milliseconds) a target word is expected to be given its actual word length and given the local speech rate of a reader (speech rate: M = 85.49 and SD = 16.38 ms per phoneme; word length: M = 3.1 and SD = 1.59 phonemes).

Frequency variables are the predictors most relevant for our research question. All probability measures were computed on the basis of the CGN lexicon (e.g., Oostdijk, 2002) as one needs to consider a reasonably large collection of the language in order to obtain representative word frequency estimates. Since the CGN is one of the biggest corpora available for Dutch–including several components such as read speech and conversational speech–and JASMIN serves as an extension to this collection, we used CGN lexica for counting the number of times a word or a word pair occurs. The probability measures examined in this study are a word's frequency and the contextual probability of a target word with its right or left neighbor. The calculations of these measures are described with the following formulae:

- Relative frequency: $P(W_i)= F(W_i) / N$
- Forward TP*: $P(W_i|W_{i-1})) = F(W_{i-1}W_i) / F(W_{i-1})$
- Backward TP*: $P(W_i|W_{i+1})) = F(W_iW_{i+1}) / F(W_{i+1})$

Consider the phrase 'Let me know' with the middle word 'me' being a target word, for which we measured word duration in milliseconds. 'Me' would be paired with 'let' for the computation of FTP, and thereby one can estimate how likely 'me' occurs given 'let'. BTP would be an estimate of how likely 'me' is to occur before 'know'.

For relative frequency, a script counted how many times the specific target word (W_i) occurs in the CGN lexicon and divided this count by the number of word types in the lexicon (N). We

---

[*] Common terms found in previous studies are *forward transitional probability* (McDonald & Shillcock, 2003b), *conditional probability of a target word given the previous word* (Bell et al., 2003) or *previous conditional probability* (Bell et al., 2009) for the TP with the previous word; and *backward transitional probability* (McDonald & Shillcock, 2003b), *conditional probability of a target word given the next word* (Bell et al., 2003) or rather *following conditional probability* (Bell et al., 2009) for the TP with the following word.

used type number (compared to the more conventional token number) for the denominator, because the exact number of words in the CGN corpus was unclear from the frequency lists we obtained. For FTP, a script first combined the target word with the preceding word, checked how many times this word pair occurs in the CGN lexicon, obtained the overall frequency of the preceding word, and computed the ratio of these two counts. Analogous computations were carried out for BTP: the joint frequency of the target and the following word was obtained and divided by the number of time the following word occurred in the CGN lexicon. Word pairs for which no frequency of occurrence could be calculated were excluded from the analysis (which happened rarely as the text consisted of easy sentences with rather frequent word pairs). No smoothing strategy was adopted for low TP items. Altogether, the dataset consisted of 60690 observations (tokens) of 777 different words (types). For each observation all of the dependent and independent variables listed in Table 1 were calculated. As explained in the data selection procedure, target words occurred in a fluent speech chunk, and their position was neither chunk-initial nor chunk-final. Range, mean, and standard deviation for each continuous variable are given in Table 2.

Table 1. Overview of variables and their description, their type and transformation (if applicable).

Variable Name and Description	Type	Transformation
Word Duration (in milliseconds)	dependent	natural log
Expected Duration (speech rate * word length)	control	natural log
Frequency (relative frequency of target word)	predictor	natural log
Relative Bigram Frequency with left word	control	natural log
Relative Bigram Frequency with right word	control	natural log
ForwardTP (TP with previous word)	predictor	natural log
BackwardTP (TP with following word)	predictor	natural log
Age (in years)	predictor	none

Table 2. Descriptive statistics for each variable (minimum (Min), maximum (Max), mean, and standard deviation (SD)).

Variable	**Min**	**Max**	**Mean**	**SD**
Word Duration (in ms)	30.00	1358.00	239.25	142.27
Expected Duration (in ms)	49.54	1423.61	260.01	139.24
Frequency	0.000029	1.51	0.43	0.47
Bigram Frequency left	0.0000007	0.0186	0.00106	0.0029
Bigram Frequency right	0.0000007	0.0186	0.00086	0.0027
ForwardTP	0.0000040	1.00	0.04	0.08
BackwardTP	0.0000038	1.00	0.08	0.17

Statistical techniques

Regression techniques were applied to investigate effects of the independent and control variables on word durations. We fitted linear mixed-effects models in R version 3.1.1 (R Development Core Team, 2014), using the packages LanguageR (Baayen, 2011) and lme4 (Bates, Maechler, & Bolker, 2014). All models contained subjects (speakerID) and items (word type) as random effects. We always added by-subject random slopes for frequency and predictability and by-item random slopes for age (or age group). The models thus adhere to the standard of a design-driven maximal random effects structure for hypothesis testing and generalizability (Barr, Levy, Scheepers, & Tily, 2013). However, if including a random slope prevented the model from converging, we removed it from the model (in that case we report which random effects were included). Note that, in principle, the models could include by-item slopes for predictability because some words (but not all) occurred in more than one context (particularly content words had no by-word variability). However, because of problems with convergence we decided to exclude them.

P-values were obtained by likelihood-ratio tests, comparing nested models that differed only in the absence or presence of the fixed effect under investigation while keeping all other fixed and random effects constant. As all predictor and control variables were theoretically motivated, no model-stripping procedure was applied. Instead, all predictors and their interactions were kept in the reported models regardless of their significance. Continuous variables (i.e., word duration, expected word duration, relative frequency, bigram frequencies, and TPs) were log-transformed to normalize their distribution (see Table 1). Furthermore, all continuous variables were centred on their mean. Given recent reports on the undesired effects of residualizing (Wurm & Fisicaro, 2014; York, 2012), we did not remove collinearity between frequency, bigram frequency, and TP by residualizing them (r = 0.74 for logfrequency and logbackwardTP, r = 0.74 for logfrequency and logforwardTP, r = 0.86 for logfrequency and combined "predictability", r = 0.63 for combined "predictability" and combined "bigram frequency", in the entire dataset). As our main interest was in the additional effect of predictability (both forward and backward) after controlling for word (and bigram) frequency, we combined BTP and FTP into one predictability predictor (by z-transforming BTP and FTP and adding the two transformed variables together). In order to check for spurious effects due to collinearity between the frequency and predictability predictors, we fitted models that solely investigated the effect of frequency (and its interaction with age group), and models that solely investigated the effect of predictability (and its interaction with age group) in addition to the "complete" models including both variables. Only the complete models will be reported in the tables hereafter. Improvements of model fit for the simple versus complete models were determined by means of likelihood-ratio tests (comparing nested models) and by evaluations of their respective Akaike information criteria (AICs). For all models reported below the comparisons showed that the complete models were significantly better models than simple frequency-only models or TP-only models. Importantly, in almost all models (except in the function-word model reported in Section D), the frequency effect became weak and insignificant in the complete model with frequency and TP, while it was significant in the frequency-only model. This suggests that there was little variance that could be uniquely explained by word frequency. Consequently, there was no straightforward evidence to conclude that there were any clear frequency effects in our data, except in function words.

Additionally, we compared each complete model (with both frequency and TP) to a model in which a combined bigram frequency variable replaced the combined TP variable, following, for example, the work of Arnon and Cohen Priva (2013) on multi-word frequencies. For this comparison, bigram frequencies of a target with the left and right word were z-transformed and added together into one measure. Similarly to the complete TP model, the bigram model included a by-subject random slope, and a bigram-by-age interaction. All predictability models reported here outperformed the bigram models as indicated by lower AICs for the predictability models, compared to the bigram models. Full correlation matrices for the datasets used in Sections A to C can be found in Appendix B.

A – Effects in children and adolescents

As mentioned before, the JASMIN corpus includes samples of younger school children (N= 61, aged 8 to 12 years) and adolescents (N = 41, aged 12 to 18 years). The focus of this section is to investigate whether frequency and TP effects increase with growing reading experience. We will first turn to the young children to investigate whether TPs between words already play a role in young reading learners, who all read different texts matched to their reading proficiency. We continue with the adolescents' data, who all read the same text.

Results for young children

The statistical results reported below are based on 17989 observations of 647 different target words for the young children (aged 8-12 years). Statistical modelling followed the procedure described in the general methods section. A summary of the complete model is given in Table 3. The model showed significant simple effects of expected word duration, age and predictability. Thus, higher speech rate (operationalized as shorter expected word durations) led to shorter word durations. The simple age effect indicated that the older the child was, the faster words were read. Children were sensitive to TPs, such that higher-predictability words had shorter word durations than lower-predictability words. Word frequency was not significant. Furthermore, neither the age-by-predictability, nor the age-by-frequency interactions were significant in the model reported in Table 3. From this analysis, one would conclude that the influence of probabilistic variables on reading durations is stable across childhood. Note though, that the frequency effect and the age-by-frequency interaction were significant in the frequency-only model (high-frequency words were shorter than low-frequency words; older children had larger frequency effects than younger children). Similarly, both the predictability simple effect and the predictability-by-age interaction were significant in the predictability-only model (high-predictability words had shorter duration than low-predictability words; older children had larger predictability effects than younger children). From these simple models, one would conclude that there is indeed an increased use of probabilistic variables across child age, but in light of the co-variation of TP and frequency it is not clear which of the two change.

Table 3. Summary of the linear mixed-effects model, including estimates, standard errors (SEs), t-values and levels of significance, fitted for children (N = 61).

Effect	**ß**	**SE**	**t**	**p <**
logExpectedWordDur	0.8656	0.0079	109.2	***
logFrequency	0.0051	0.0040	1.3	n.s.
Age	-0.0114	0.0041	-2.8	*
Predictability	-0.0488	0.0026	-18.8	***
logFrequency*Age	-0.0021	0.0018	-1.1	n.s.
Predictability *Age	-0.0020	0.0016	-1.2	n.s.

*** p < 0.0001, * p < 0.05, n.s. not significant

Results for adolescents

The statistical results reported below are based on 17614 observations of 252 different words for the adolescent group (aged 12-18 years). A summary of the model is given in Table 4. The model converged with random intercepts for speakers and words, by-word random slopes for age, and by-speaker random slopes for predictability. The model showed significant simple effects of expected word duration and predictability. That is, a higher speech rate led to shorter word durations, compared to a slower speech rate. Additionally, a higher predictability of the target from the previous and following words led to shorter word durations. No simple age effect occurred, showing that reading speed (as indicated by spoken word durations) did not increase over the teenage years. This confirms that the adolescent group was more homogeneous in their reading skills than the primary school children. There was no simple effect of word frequency in the complete model, although there was one in a frequency-only model. Again, this indicates that there is not enough unique variance for frequency to explain, if entered together with its covariate predictability in a single model, because predictability is the stronger probabilistic variable in our data set. Neither the frequency by age nor TP by age interactions reached significance in the complete model, suggesting that probabilistic effects do not increase in reading across adolescence. Models in which TP or frequency was the only probabilistic variable confirmed this lack of an interaction with age.

Table 4. Summary of the linear mixed-effects model, including estimates, standard errors (SEs), t-values and levels of significance, fitted for adolescents (N = 41).

Effect	**ß**	**SE**	**t**	**p <**
logExpectedWordDur	0.8535	0.0079	108.4	***
logFrequency	-0.0009	0.0060	-0.2	n.s.
Age	-0.0005	0.0028	-0.2	n.s.
Predictability	-0.0722	0.0030	-23.9	***
logFrequency*Age	0.0013	0.0008	1.6	n.s.
Predictability *Age	-0.0007	0.0013	-0.5	n.s.

*** $p < 0.0001$, n.s. not significant

Discussion

Predictability effects reliably affected spoken word durations in reading aloud in children as well as adolescents, which speaks to our first research question. The results show that effects of TPs, as found in silent reading and conversational production, transfer to reading aloud of child readers, over and above word frequency effects. Because it turned out that frequency explained insufficient unique variance, we cannot draw any firm conclusions about the existence of frequency effects. Considering our second research question, data of the younger readers (aged 8 to 12) provided some (though limited) evidence that frequency and/or predictability effects build up with age and thus increase with language experience. The data do not allow strong conclusions on which of the two is driving this interaction, though. However, probabilistic effects do not increase further with age during adolescence (i.e., in the group aged 12 to 18 years). As noted above, the adolescents seem to be more homogeneous than the young children in their reading abilities, given the stability of the reading rate and the stability of the frequency and predictability effects across adolescent age. However, another obvious explanation of why the adolescent group may be more homogeneous than the children may be that all adolescents read the same text while children read different (reading-level appropriate) texts. Also note that the model for adolescents was based on a smaller participant sample and fewer words than that of the younger children, which may have affected the reliability of frequency estimates.

B – Effects within older readers

In this section we specifically investigate the effects of frequency and predictability variables on word durations in older adults' word reading times. Among the older adults (N = 64, aged 62 to 95) we investigate whether frequency and TP effects on word durations decrease with increasing age. As mentioned in the introduction, predictability differences among older adults may index age-related cognitive limits, which could complicate the synchronization of processes needed for reading aloud. The same methods and statistical techniques were applied as described in the general methods section. The results reported in this analysis are based on 25087 observations (tokens) of overall 305 different target words (types).

Results

All ß-estimates, standard errors, t-values, and levels of significance for fixed effects are summarized in Table 5. The fitted model included random intercepts for speakers and words, by-word random slopes for age, and by-speaker random slopes for predictability. The by-speaker slope for frequency was left out due to problems with convergence. The linear regression model showed significant effects of expected word duration and age. Hence, higher speech rate (operationalized as shorter expected word duration) led to shorter word duration. Furthermore, within this age group, the older a reader was the longer the word durations became. The simple effect of predictability showed that the more predictable a target word was from its context, the more it was reduced. This effect was confirmed in a model that included predictability only (leaving word frequency out). Frequency had no effect in the complete model, but was significant in a frequency-only model, such that frequent words were more reduced than infrequent words. None of the interactions between age and the probabilistic variables (frequency, and predictability) were significant.

Table 5. Summary of the linear mixed-effects model, including estimates, standard errors (SEs), t-values and levels of significance, fitted for older adults (N = 64).

Effect	**ß**	**SE**	**t**	**p <**
logExpectedWordDur	0.8035	0.0071	113.1	***
logFrequency	-0.0094	0.0051	-1.8	n.s.
Age	0.0011	0.0005	2.2	*
Predictability	-0.0535	0.0027	-20.2	***
logFrequency*Age	-0.0001	0.0001	-0.5	n.s.
Predictability *Age	0.0002	0.0003	0.9	n.s.

*** $p < 0.0001$, * $p < 0.05$, n.s. not significant

Discussion

The effects obtained in these models confirm our hypotheses concerning the relation between predictability and acoustic reduction. Our first research question was whether we would find predictability effects in reading aloud in this group, over and above word frequency effects. Indeed, local predictability shows facilitatory effects on reading durations within an older adult group. Consequently, our results underscore the link between high probability of occurrence and faster pronunciation, as stated in the Probabilistic Reduction Hypothesis (Jurafsky et al., 2001), even for reading aloud. Again, we cannot conclude from our data that frequency had an overall effect, due to its covariation with TP. In answer to our third research question, the non-significant interactions between age and frequency or predictability suggest that these facilitatory effects are stable effects within older adulthood. Additionally, older seniors generally had longer word durations than younger seniors. This age-related slowing could be due to slower reading comprehension or slower execution of motor commands, for instance. We will come back to this in the general discussion.

C – Age group comparison

This section investigates the effects of frequency and predictability on word durations in reading aloud for older adults compared to adolescents (for a comparison of all three groups – children, adolescents, and older adults – see Appendix A). The sample for this age group comparison consists of 41 adolescents (aged 12 to 18 years, cf. Section A) and 64 older adults in the age range of 62 to 95 years (cf. Section B). As noted before, adolescents and older adults read aloud the same text. The results reported below are based on 25087 observations for the older adults and 35603 observations for the younger group (i.e., 60690 observations in total) of overall 325 different words. Statistical modelling followed the procedure described in the statistical techniques section, except for replacing the continuous age variable by a categorical age group factor, which was contrast coded (0.5 for older adults, and -0.5 for the younger group).

Results

A summary of the model fitted for the age group comparison, including ß-coefficients, standard errors, t-values and significance levels, can be found in Table 6. The model included random intercepts for speakers and words, by-word random slopes for age group, and by-speaker random slopes for predictability (the by-speaker frequency slope was left out due to a perfect correlation with the by-speaker intercept and subsequent convergence errors). The model yielded significant simple effects of expected word duration and age group. Thus, word durations that were expected to be longer (on the basis of word length and speech rate) were indeed longer. Older adults read words overall more slowly than adolescents. Furthermore, predictability showed a strong facilitatory effect on word duration, with higher predictability leading to shorter durations. Additionally, predictability interacted with age group in that older adults had smaller-sized predictability effects compared to adolescent readers. The same interaction effect occurred in a predictability-only model (leaving word frequency out). Word frequency had no effect in the complete model reported in Table 6, but was significant in a frequency-only model. Similarly, age group and word frequency did not interact in the complete model, but did in the frequency-only model, such that word frequency effects were smaller for older adults compared to adolescents. In sum, we do not have firm evidence for age-related changes in frequency effects, but we can conclude that predictability effects change across age groups.

Table 6. Estimates, standard errors (SEs), t-values and levels of significance for the age group comparison between adolescents and older adults (N = 105).

Effect	**ß**	**SE**	**t**	**p <**
logExpectedWordDur	0.8258	0.0053	155.5	***
logFrequency	-0.0055	0.0051	-1.1	n.s.
AgeGroup_OlderAdults	0.0564	0.0084	6.7	***
Predictability	-0.0628	0.0020	-30.6	***
logFrequency*AgeGroup	-0.0022	0.0028	-0.8	n.s.
Predictability *AgeGroup	0.0140	0.0036	3.9	**

*** p < 0.0001, ** p < 0.001, n.s. not significant

Discussion

The models we obtained in this age group comparison speak to our research questions in two ways. First, as shown in Sections A and B, probabilistic reduction effects generalize to a complex task such as reading aloud, in adolescents as well as older adults; second, we investigated whether older adults would differentially use predictability cues compared to adolescents. This prediction was confirmed in the direction that older adults had smaller predictability effects than younger adults. We will come back to these findings in more detail in the general discussion of this paper.

D – Content versus function words

Bell and colleagues (2009; see also Jurafsky et al., 2001) report that probabilistic reduction may affect function words and content words differently. As function words are encountered so frequently and learned early during language acquisition, function words may be more prone to age-related ceiling effects in frequency and predictability effects compared to content words. Since differences in probabilistic knowledge among age groups may consequently be larger for content than function words, we separated the two word categories and investigated differences for content versus function words with the data reported in Section C (more experienced older adults versus less experienced adolescents). We used the POS tags provided in the JASMIN annotations to distinguish between content and function words. In line with Bell and colleagues (2009), we assigned nouns, verbs, adjectives, and adverbs to the content word class (N = 19729 observations). The function word category included conjunctions, pronouns, prepositions, articles, quantifiers and demonstratives (N = 22972 observations).

Table 7 and Table 8 provide inter-correlation matrices for word durations and all other continuous variables in our dataset for content versus function words, respectively. The two matrices show a very similar pattern of intercorrelations for content and function words.

Table 7. Correlation matrix for continuous variables in the content word dataset.

	WD	ExpWD	FREQ	JNW	JPW	BTP	FTP	PRED
WD	1							
ExpWD	0.87	1						
FREQ	-0.65	-0.66	1					
JNW	-0.43	-0.41	0.64	1				
JPW	-0.19	-0.19	0.32	0.24	1			
BTP	-0.53	-0.44	0.66	0.56	0.23	1		
FTP	-0.45	-0.44	0.66	0.36	0.50	0.38	1	
PRED	-0.59	-0.53	0.80	0.54	0.45	0.83	0.82	1
AGE	0.14	0.13	0.01	0.01	0.01	0.02	0.01	0.02

WD = log word duration, ExpWD = log expected word duration, FREQ = log relative frequency, JNW = log joint frequency with next word, JPW = log joint frequency with previous word, BTP = log backward transitional probability, FTP = log forward transitional probability, PRED = scaled predictability, AGE = participant age

Table 8. Correlation matrix for continuous variables in the function word dataset.

	WD	ExpWD	FREQ	JNW	JPW	BTP	FTP	PRED
WD	1							
ExpWD	0.76	1						
FREQ	-0.55	-0.42	1					
JNW	0.07	0.03	0.00	1				
JPW	-0.48	-0.29	0.45	-0.07	1			
BTP	-0.46	-0.3	0.59	0.11	0.36	1		
FTP	-0.46	-0.31	0.65	-0.01	0.55	0.38	1	
PRED	-0.56	-0.37	0.76	0.08	0.54	0.82	0.82	1
AGE	0.15	0.20	0.01	-0.01	0.02	0.02	0.01	0.02

WD = log word duration, ExpWD = log expected word duration, FREQ = log relative frequency, JNW = log joint frequency with next word, JPW = log joint frequency with previous word, BTP = log backward transitional probability, FTP = log forward transitional probability, PRED = scaled predictability, AGE = participant age

We fitted separate mixed-effects models for both word classes following the exact same procedure described above (in Section C containing the comparison between the adolescents and the older adults). Summaries of the models, including estimates, standard errors, t-values and levels of significance, are provided in Table 9 (content words) and Table 10 (function words). Both content and function word durations were influenced by speech rate, participant age, as well as predictability from left and right neighboring words. Hence, faster speech rate led to shorter words, older adults had longer word durations than adolescents, and predictable words had shorter durations than unpredictable words. In the models for both content and function words, we found significant predictability by age group interactions, such that older adults had smaller predictability effects on word durations than adolescents (as in Section C). Thus, the results obtained for content and function words showed a very similar pattern. The only difference between the content and function word models was that the simple effect of word frequency was significant (and in the expected direction) in the complete model for function words (Table 10), but not for content words (Table 9). Note that in the respective frequency-only models, significant simple effects of word frequency were observed for both word classes. In light of these results, we can conclude that frequency affects the oral reading of functions words, but we cannot draw firm conclusions about frequency effects in content words, due to covariation of frequency and predictability.

Table 9. Estimates, standard errors (SEs), t-values and levels of significance for the age group comparison between adolescents and older adults (N = 105); model investigating frequency and predictability effects in content words (19729 observations of 252 different words).

Effect	**ß**	**SE**	**t**	**p <**
logExpectedWordDur	0.6990	0.0087	80.8	***
logFrequency	0.0002	0.0056	0	n.s.
AgeGroup_OlderAdults	0.0575	0.0080	7.2	***
Predictability	-0.0656	0.0026	-25.2	***
logFrequency*AgeGroup	-0.0023	0.0035	-0.6	n.s.
Predictability *AgeGroup	0.0128	0.0037	3.5	**

*** $p < 0.0001$, ** $p < 0.001$, n.s. not significant

Table 10. Estimates, standard errors (SEs), t-values and levels of significance for the age group comparison between adolescents and older adults (N = 105); model investigating frequency and predictability effects in function words (22972 observations of 78 different words).

Effect	**ß**	**SE**	**t**	**p <**
logExpectedWordDur	0.8604	0.0070	122.71	***
logFrequency	-0.0347	0.0158	-2.19	*
AgeGroup_OlderAdults	0.0717	0.0165	4.36	***
Predictability	-0.0643	0.0026	-25.01	***
logFrequency*AgeGroup	-0.0053	0.0081	-0.66	n.s.
Predictability *AgeGroup	0.0158	0.0049	3.21	**

*** $p < 0.0001$, ** $p < 0.001$, * $p < 0.05$, n.s. not significant

General discussion

The present study investigated predictability effects on spoken word durations in reading aloud in young readers and older adults. Our goals for this study were to investigate: (1) whether frequency and predictability effects, as found in reading aloud of middle-aged adults in previous studies, generalize to read aloud speech of different age groups; (2) whether frequency and predictability effects vary with age and hence reading experience in children and adolescents;

(3) whether frequency and predictability effects in read aloud speech change with age among older adult readers; and (4) whether older adults show differential benefits of predictability compared to adolescents.

Regarding the first research question, our results indicate that three different groups – younger children, adolescents, and older readers – showed probability-based facilitation effects on spoken word durations in reading aloud. Therefore, TP effects, which had been found in more homogeneous groups of middle-aged adults, generalized to reading aloud for readers with varying reading experience. All regression models, both within-group models and group comparison models, yielded significant effects of TPs (of a target word with its right and left neighboring word), when other variables known to affect word duration (e.g., speech rate) were controlled for. Thus, our data show that the probabilistic reduction hypothesis (Jurafsky et al., 2001) can be confirmed for children and older adults when reading aloud entire sentences. Note, however, that our analyses did not allow us to draw straightforward conclusions about the effect of word frequency on probabilistic reduction across the different age groups. Although frequency-related acoustic reduction may exist in our data (as evidenced by the frequency-only models), frequency and TP were co-variates, and it turned out that frequency effects were not significant in models in which both variables were entered. Thus, TP effects are clearly present in our data, but the evidence for frequency effects is less robust. We will now discuss the different age group results in more detail.

Frequency and predictability effects in children and adolescents

With regard to our second research question, our results provided some evidence that probabilistic effects increased with age in the child reader group, although the results are not clear on what probabilistic variable to assign the interaction to. Note that the direction of this interaction is different from that observed in the comparison between adolescents and older adults, where probabilistic effects decrease with age. This discrepancy can be accounted for in at least two ways. First, the observation in the group of children may be due to text differences, with children reading different texts of various reading levels. This text difference may impact on the word frequency range in the different texts within the group. Second, children in this age range may differ in reading strategy with the youngest children still employing a more technical reading strategy and the older children processing words and word combinations more holistically (e.g., Huestegge, Radach, Corbic, & Huestegge, 2009; Share, 1995; Vorstius et al., 2014; Ziegler & Goswami, 2005). Word and co-occurrence frequency can be expected to play a more important role once the words and combinations are read as wholes, rather than still being decomposed into their constituent parts. Further research with more controlled materials would be required to investigate these alternatives further.

Our results extend existing knowledge on the use of lexical statistics in young school children in several ways. Lexical frequency of single words has been shown to influence children’s fixation durations and word skipping rates in single word recognition or sentence reading, from as early as 7 years of age (e.g., Aghababian & Nazir, 2000; Blythe & Joseph, 2011; Hyönä & Olson, 1995). Our results do not allow us to draw straightforward conclusions about the effect of single-word frequency. However, our study yields two novel findings: first, lexical co-occurrence effects play a role in reading, even for young children with only about two years of reading practice. In other words, young readers seem to exploit inter-word predictability to make their reading more efficient. Second, word co-occurrence effects are not just found in the eye fixation patterns during children’s reading (and thus in their text comprehension), but are also found in young children’s spoken word durations in reading aloud (i.e., in their production).

Exploiting word or inter-word probabilities in children as a result of increasing reading experience represents a form of distributional learning. Words generally do not occur in isolation but rather in phrasal frames. Finding out about distributional regularities in the linguistic environment consequently contributes to the development of reading skills and reading efficiency (e.g., Perea et al., 2013). Knowledge about the frequencies of words and word pairs can then influence both the ease of word identification and their subsequent pronunciation in reading aloud. From the current data we cannot tell whether efficiency is solely developed in the word identification process, or the pronunciation of words, or both. As motor skills improve with age in terms of speed and accuracy, any changes in reading behavior in our young sample may be a combination of changes in motor practice effects, word decoding speed, age-related cognitive development (cf. Blythe & Joseph, 2011), as well as changes in contextual facilitation and probabilistic knowledge (e.g., Nation & Snowling, 1998).

As part of the second research question, we also investigated whether probabilistic effects changed in size in read speech data of adolescents aged 12 to 18. Even though the adolescents also clearly showed evidence of TP effects, the within-group analyses yielded no significant interactions between age and TP, or between age and word frequency. Thus, this group was rather homogeneous in the exploitation of surrounding context and the use of probabilistic knowledge in reading. This suggests that probabilistic reduction effects do not necessarily steadily increase with more language experience. We conclude for our reading study that the use of probabilistic cues and the exploitation of immediate context (that is, neighboring words) is particularly relevant for children learning to read, but once a certain level of reading experience is reached, facilitation from word frequency and surrounding context may level off. Our findings are also in line with research showing that school children reach adult performance in reading around the age of 11 years (cf. Blythe & Joseph, 2011; see also Schroeder, 2011, using self-paced reading). Furthermore, the use of local predictability cues in the entire young group may be driven by overall improvement in reading abilities and reading efficiency (and not solely by growth of probabilistic linguistic knowledge).

Frequency and predictability effects in senior readers

Turning to our third and fourth research questions, we investigated whether TP effects changed across advanced age and whether any differences in probabilistic processing would be found in older adults versus adolescents. The literature on age-related changes in frequency and predictability effect presents conflicting evidence. On the one hand, older adults have been reported to rely more on frequency information than younger adults (e.g., Rayner et al., 2006, in reading; Revill & Spieler, 2012, in listening) by showing more lexical facilitation for highly frequent target words than younger adults. On the other hand, older adults have been reported to make less use of frequency information in written word recognition (Robert, Mathey, & Postal, 2009). In line with the latter suggestion, our comparison of older adults to adolescents (Section C) suggested that frequency effects (in our frequency-only model) and TP effects (in our full model), if anything, were smaller in older adults than adolescents. This result can best be explained by a lexical entrenchment account according to which increased language experience leads to decreased frequency effects because lexical memory representations become more precise or “entrenched” with increasing language experience, resulting in less interference from co-activated representations.

With regard to predictability effects, our results extend the results of older adults’ decreased use of predictive cues in language comprehension (Federmeier et al., 2010). Our comparison of older adults to adolescents (Section C) showed that the older adults had smaller TP effects compared to adolescents. Federmeier and colleagues (e.g., Federmeier & Kutas, 2005) attribute these age differences to age-related decline in cognitive resources. More specifically, predictive processing may require working memory, which is subject to age-related cognitive decline. Alternatively, the smaller probabilistic effects may reflect that low-frequency words and word combinations have caught up with the higher-frequency ones as experienced language users have also encountered the low-frequency ones multiple times. Additionally, any difference between the two age groups may reflect motivational differences and differences in how each age group responds to the expectations of the task (such that older speakers may be more sensitive to what they think is good practice in a reading task, e.g., clarity of diction, compared to adolescents).

Among older adults, aged 62 to 95, we found no evidence for increasing or decreasing predictability effects with advanced age. The general picture from the older adult data is that local probability effects are mostly stable in older age. This is an important finding because it shows that the demands and complexity of oral reading do not cancel local predictability effects. Furthermore, the stability of probabilistic effects may be accounted for by ceiling effects in frequency measures. The degree to which increased use or experience of word (combinations) can lead to increased accessibility is limited. Hence, the benefit of probabilistic knowledge is unlikely to steadily increase across the entire life span (cf. also the adolescent data in Section A). At some point baseline activation may simply have reached its limits, after which additional occurrence no longer exerts substantial changes (cf. Gollan et al., 2008). Furthermore, there may be a limit to the phonological fine-tuning of lexical representations and motor commands with increased usage. Increasing facilitation in ease or speed of lexical access is thereby naturally restricted. These results suggest therefore that age does not impact local probability effects as much as conceptual and semantic prediction (as e.g., in Janse & Jesse, 2014). Effects of TP may thus differ from higher-level semantic prediction or inference as frequently measured by cloze or sentence completion tasks (e.g., Hahn, 2012; McDonald & Shillcock, 2003a, 2003b; Smith & Levy, 2011; but cf. Frisson et al., 2005).

One effect in our data that is clearly age-related is that word durations become longer in the older adults group (compared to the adolescents) and with increasing age (among the older adults). This slowing of reading with advanced adult age has also been found for silent reading as reflected in longer fixation durations, longer sentence reading times and more regressions in older adults (see Kliegl et al., 2004; Rayner et al., 2006). The slowing in our sample may have several reasons, either in the reading comprehension component (e.g., due to age-related declines in visual processing speed; Salthouse, 1996), or in the speech production component of reading aloud where age-related declines in speech motor execution have been observed (Rodriguez-Aranda & Jakobsen, 2011; Weismer & Liss, 1991), or both. Additionally, slower reading rate for older adults may be due to increased interference effects from what has been processed before (e.g., Stoltzfus, Hasher, Zacks, Ulivi, & Goldstein, 1993). Lastly, age-related slowing may be explained by motivation, style or register differences between age groups, such that older adults actively decide to speak more slowly and/or more formally than adolescents.

Mechanisms behind systematic pronunciation variation

We have now linked our research questions to the observed findings within each age group, but the question remains how exactly word frequency or probability effects find their way down to systematic pronunciation variation. Probabilistic effects on how words and phrases are realized can generally be explained within usage-based models. In a usage-based view of the language system, frequency of occurrence indexes the experience a language user has with a given word or phrase (cf. Janssen & Barber, 2012). Probability-conditioned reduction can lead to permanent representational change in the lexicon, such that frequent reductions of certain words bias later productions of these words (and hence these words may be reduced even in unpredictable contexts; cf. Seyfarth, 2014).

More specific accounts have been proposed which may be classified into two types: listener-driven accounts; and speaker-internal accounts. In listener-driven accounts, language users may systematically choose shorter word forms in more predictive contexts, indicating a tendency to distribute information evenly across the speech signal to help the listener (see Aylett & Turk, 2004). Speakers can make such explicit behavioral choices by drawing on their knowledge about the language system (cf. Mahowald, Fedorenko, Piantadosi, & Gibson, 2013). Reduction processes may also be speaker-internal. In speech production, planning processes are not strictly isolated from lower levels such as the encoding of articulatory plans (e.g., Goldrick, Baker, Murphy, & Baese-Berk, 2011; Kahn & Arnold, 2012; Mousikou & Rastle, 2015). Any kind of facilitation on a lexical-representational level may spill over to the pronunciation level, resulting in shorter word durations (but see Baese-Berk & Goldrick, 2009). In addition, articulatory motor routines may be compressed for high-frequency words or likely word transitions as a result of extensive practice (for discussion see Bybee, 2006; Pluymaekers et al., 2005b). Thus, articulatory plans for pronunciation may either be retrieved faster (during lexical access), or the articulators itself may be faster for highly likely word combinations.

Listener-driven and speaker-driven accounts are not mutually exclusive, and we cannot distinguish between them on the basis of our corpus data. In general, accounts that explain probabilistic reduction effects assume that statistical knowledge about the occurrence and co-occurrence of words influences long-term linguistic representations, as well as online language production strategies (cf. Arnon & Snider, 2010; Gahl, Yao, & Johnson, 2012; Janssen & Barber, 2012; Reali & Christiansen, 2007; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Tremblay, Derwing, Libben, & Westbury, 2011; for more general discussions of these accounts see Buz & Jaeger, 2016; Bybee, 2006; Gahl, 2008; Jacobs, Yiu, Watson, & Dell, 2015).

Conclusions and outlook

Knowledge about the frequency patterns of words and word pairs is acquired with language experience, and thus with age. This study is the first to yield insights into how word frequency and word co-occurrence statistics affect spoken word durations in reading aloud in different age groups. Further experimental research is required to determine at which comprehension or production levels probabilistic effects arise and thus to be able to decide between the probabilistic reduction accounts discussed above. Further experimental studies may also provide building blocks for the development of models of reading aloud entire sentences (e.g., Coltheart et al., 2001; for a discussion, see Rayner & Reichle, 2010). Thus far, models of reading aloud have focused on naming single words, rather than reading aloud sentences. To advance our understanding of how reading aloud works, develops in childhood, and changes throughout adulthood, it is important to investigate the factors that lead to successful orchestration of the subcomponents entailed in reading aloud, with probabilistic knowledge being one of them. One other important point that future studies should address is the relationship between objective, corpus-based frequency estimates and participants' subjective frequencies (e.g., Kuperman & Van Dyke, 2013), as subjective frequencies might be better or worse matches to corpus frequencies and TPs at different ages.

To conclude, this study initiates an interesting path for further research on the strength and origin of word co-occurrence and predictability effects in reading aloud and their development across the life span. Our results clearly show that efficiency in reading aloud in skilled readers, as well as in young language learners, may, at least partly, stem from contextual facilitation.

Acknowledgement

We thank Iris Hanique for her support and helpful comments on the usage of the CGN corpus.

Appendix A: Age group comparison with three groups

In Tables 11 and 12 we report the results of two mixed-effects models we fitted for all three age groups in the JASMIN corpus (61 children, 41 adolescents, and 64 older adults, with a total of 60690 observations). For comparing the three different age groups, we mapped the adolescents on the intercept and compared the older adults to the adolescents as well as the children to the adolescents (using a 3-level treatment coded factor for age group). The models yielded simple effects of expected word duration, word frequency (solely in the frequency-only model, cf. Table 11), age group of the older adults (not for the children), and predictability. Thus, faster speech rate led to shorter words, high-frequency words were more reduced than low-frequency words, and older adults had longer word durations than adolescents. Age group did not interact with word frequency, neither in the comparison of the older adults versus adolescents, nor in the comparison of the children versus adolescents. Hence, frequency effects were similar in size in all three groups. However, we did find significant interactions of predictability and age group, such that older adults showed smaller effects of predictability than adolescents, and children had smaller predictability effects than adolescents (but note that due to the missing simple effect of age group for the children, the latter interaction needs to be interpreted with caution).

Table 11. Estimates, standard errors (SEs), t-values and levels of significance for the age group comparison between children, adolescents and older adults (N = 166); model investigating frequency effects (Akaike information criterion = -9755).

Effect	**ß**	**SE**	**t**	**p <**
logExpectedWordDur	0.8501	0.0044	192.2	***
logFrequency	-0.0352	0.0051	-6.9	***
AgeGroup_Children	0.0149	0.0127	1.2	n.s.
AgeGroup_OlderAdults	0.0470	0.0090	5.2	**
logFrequency*Children	0.0098	0.0050	1.9	n.s.
logFrequency*OlderAdults	0.0052	0.0031	1.7	n.s.

*** p < 0.0001, ** p < 0.001, n.s. not significant

Table 12. Estimates, standard errors (SEs), t-values and levels of significance for the age group comparison between children, adolescents, and older adults (N = 166); model investigating additional predictability effects (Akaike information criterion = -11424).

Effect	**ß**	**SE**	**t**	**p <**
logExpectedWordDur	0.8413	0.0044	191.9	***
logFrequency	0.0052	0.0053	1.0	n.s.
AgeGroup_Children	0.0243	0.0127	1.9	n.s.
AgeGroup_OlderAdults	0.0512	0.0092	5.5	**
Predictability	-0.0694	0.0027	-25.5	***
logFrequency*Children	0.0001	0.0055	0.0	n.s.
logFrequency*OlderAdults	-0.0028	0.0036	-0.8	n.s.
Predictability *Children	0.0198	0.0037	5.4	**
Predictability *OlderAdults	0.0132	0.0033	4.1	**

*** p < 0.0001, ** p < 0.001, n.s. not significant

Appendix B: Additional correlation matrices

Tables 13-16 present additional correlation matrices.

Table 13. Correlation matrix for continuous variables in the dataset of children (8-12 years).

	WD	ExpWD	FREQ	JNW	JPW	BTP	FTP	PRED
WD	1							
ExpWD	0.86	1						
FREQ	-0.70	-0.64	1					
JNW	-0.32	-0.32	0.47	1				
JPW	-0.34	-0.31	0.42	0.13	1			
BTP	-0.61	-0.52	0.75	0.45	0.34	1		
FTP	-0.57	-0.53	0.73	0.36	0.51	0.54	1	
PRED	-0.68	-0.60	0.84	0.46	0.48	0.87	0.87	1
AGE	-0.19	-0.22	0.01	0.01	-0.01	0.02	0.01	0.02

WD = log word duration; ExpWD = log expected word duration; FREQ = log relative frequency; JNW = log joint frequency with next word; JPW = log joint frequency with previous word; BTP = log backward transitional probability; FTP = log forward transitional probability; PRED = scaled predictability; AGE = participant age.

Table 14. Correlation matrix for continuous variables in the dataset of adolescents (12-18 years).

	WD	ExpWD	FREQ	JNW	JPW	BTP	FTP	PRED
WD	1							
ExpWD	0.87	1						
FREQ	-0.79	-0.74	1					
JNW	-0.34	-0.32	0.43	1				
JPW	-0.38	-0.31	0.39	0.12	1			
BTP	-0.67	-0.57	0.73	0.44	0.33	1		
FTP	-0.62	-0.56	0.74	0.28	0.52	0.51	1	
PRED	-0.76	-0.66	0.86	0.42	0.49	0.86	0.86	1
AGE	-0.04	-0.08	0.00	0.00	-0.01	0.00	0.00	0.00

WD = log word duration; ExpWD = log expected word duration; FREQ = log relative frequency; JNW = log joint frequency with next word; JPW = log joint frequency with previous word; BTP = log backward transitional probability; FTP = log forward transitional probability; PRED = scaled predictability; AGE = participant age.

Table 15. Correlation matrix for continuous variables in the dataset of older adults (62-95 years)

	WD	ExpWD	FREQ	JNW	JPW	BTP	FTP	PRED
WD	1							
ExpWD	0.88	1						
FREQ	-0.78	-0.74	1					
JNW	-0.32	-0.33	0.42	1				
JPW	-0.41	-0.33	0.41	0.13	1			
BTP	-0.66	-0.58	0.73	0.43	0.34	1		
FTP	-0.62	-0.56	0.75	0.27	0.54	0.51	1	
PRED	-0.75	-0.67	0.86	0.42	0.51	0.86	0.86	1
AGE	0.02	0.02	0.01	0.00	0.01	0.01	0.01	0.01

WD = log word duration, ExpWD = log expected word duration, FREQ = log relative frequency, JNW = log joint frequency with next word, JPW = log joint frequency with previous word, BTP = log backward transitional probability, FTP = log forward transitional probability, PRED = scaled predictability, AGE = participant age

Table 16. Correlation matrix for continuous variables in the combinded dataset of adolescents and older adults.

	WD	ExpWD	FREQ	JNW	JPW	BTP	FTP	PRED
WD	1							
ExpWD	0.88	1						
FREQ	-0.77	-0.73	1					
JNW	-0.33	-0.32	0.43	1				
JPW	-0.40	-0.31	0.41	0.12	1			
BTP	-0.66	-0.56	0.73	0.44	0.33	1		
FTP	-0.62	-0.56	0.75	0.28	0.53	0.51	1	
PRED	-0.74	-0.66	0.86	0.42	0.50	0.86	0.86	1
AGE	0.11	0.13	0.00	0.00	0.01	0.01	0.01	0.01

WD = log word duration, ExpWD = log expected word duration, FREQ = log relative frequency, JNW = log joint frequency with next word, JPW = log joint frequency with previous word, BTP = log backward transitional probability, FTP = log forward transitional probability, PRED = scaled predictability, AGE = participant age

References

Aghababian, V., & Nazir, T. A. (2000). Developing normal reading skills: aspects of the visual processes underlying word recognition. *Journal of Experimental Child Psychology*, 76(2), 123-150. doi: 10.1006/jecp.1999.2540

Allen, P. A., Smith, A. F., Lien, M.-C., Grabbe, J., & Murphy, M. D. (2005). Evidence for an Activation Locus of the Word-Frequency Effect in Lexical Decision. *Journal of Experimental Psychology: Human Perception and Performance*, 31(4), 713-721. doi: 10.1037/0096-1523.31.4.713

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, 42(02), 239-273. doi: 10.1017/S030500091400049X

Anderson, I. H., & Swanson, D. E. (1937). Common factors in eye-movements in silent and oral reading. *Psychological Monographs,* 48(3), 61-77. doi: 10.1037/h0093393

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General*, 139(2), 299-318. doi: 10.1037/a0018366

Arnon, I., & Cohen Priva, U. C. (2013). More than words: the effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, 56(3), 349-371. doi: 10.1177/0023830913484891

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62, 67- 82. doi: 10.1016/j.jml.2009.09.005

Aylett, M., & Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Language and Speech*, 47(1), 31-56. doi: 10.1177/00238309040470010201

Baayen, R.H. (2011). LanguageR: Data sets and functions with “Analyzing Linguistic Data: A practical introduction to statistics”. R package version 1.4. Retrieved from http://cran.r-project.org/web/packages/languageR/

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, 24(4), 527-554. doi: 10.1080/01690960802299378

Baker, R. E., & Bradlow, A. R. (2009). Variability in Word Duration as a Function of Probability, Speech Style, and Prosody. *Language and Speech*, 52(4), 391-413. doi: 10.1177/0023830909336575

Balota, D. A., & Ferraro, F. R. (1996). Lexical, sublexical, and implicit memory processes in healthy young and healthy older adults and in individuals with dementia of the Alzheimer type. *Neuropsychology*, 10(1), 82-95. doi: 10.1037/0894-4105.10.1.82

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278. doi: http://dx.doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., & Bolker, B. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. Retrieved from http://lme4.r-forge.r-project.org/

Beckman, M. E., & Edwards, J. (1990). Lengthening and shortenings and the nature of prosodic constituency. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I* (pp. 152 – 178). Cambridge: Cambridge University Press.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1), 92-111. doi: http://dx.doi.org/10.1016/j.jml.2008.06.003

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, 113(2), 1001-1024. doi: http://dx.doi.org/10.1121/1.1534836

Ben-David, B. M., Erel, H., Goy, H., & Schneider, B. A. (2015). “Older is always better”: Age-related differences in vocabulary scores across 16 years. *Psychology and Aging*, 30(4), 856-862. doi: 10.1037/pag0000051

Blythe, H.I., & Joseph, H.S.S.L. (2011). Children’s Eye Movements during Reading. In S.P. Liversedge, I. Gilchrist & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 643-662). Oxford: Oxford University Press.

Breznitz, Z., & Berman, L. (2003). The Underlying Factors of Word Reading Rate. *Educational Psychology Review*, 15(3), 247-265. doi: 10.1023/a:1024696101081

Buswell, G. T. (1921). The relationship between eye-perception and voice-response in reading. *Journal of Educational Psychology, 12*(4), 217-227. doi: 10.1037/h0070548

Buz, E., & Jaeger, T. F. (2016). The (in)dependence of articulation and lexical planning during isolated word production. *Language, Cognition and Neuroscience,* 31(3), 404-424. doi: 10.1080/23273798.2015.1105984

Bybee, J. L. (2006). *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195301571.001.0001

Calfee, R. & Drum, P. (1986). Research on teaching reading. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 804–849). New York: MacMillan Publishing Company.

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, 10(7), 335-344. doi: http://dx.doi.org/10.1016/j.tics.2006.05.006

Clopper, C. G., & Pierrehumbert, J. B. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *The Journal of the Acoustical Society of America*, *124*(3), 1682–1688. http://doi.org/10.1121/1.2953322

Coltheart, V., Laxon, V. J., & Keating, C. (1988). Effects of word imageability and age of acquisition on children’s reading. *British Journal of Psychology*, 79(1), 1-12. doi: 10.1111/j.2044-8295.1988.tb02270.x

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256. doi:10.1037/0033-295X.108.1.204.

Cucchiarini, C., Driesen, J., Van Hamme, H., & Sanders, E. (2008). Recording Speech of Children, Non-Natives and Elderly People for HLT Applications: the JASMIN-CGN corpus. *Proceedings of the Sixth International Conference on Language Resources and Evaluation* (LREC’08), Marrakech, Morocco. Retrieved from http://www.lrec-conf.org/proceedings/lrec2008/summaries/366.html

Davies, R. A., Arnell, R., Birchenough, J. M., Grimmond, D., & Houlson, S. (2017). Reading through the life span: Individual differences in psycholinguistic effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 43(8):1298-1338. doi: 10.1037/xlm0000366

DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, 121(3), 226-239. doi: 10.1016/j.bandl.2012.02.006

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2012). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *The Quarterly Journal of Experimental Psychology*, 1-21. doi: 10.1080/17470218.2012.720994

Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, 42(2), 133-141. doi: 10.1111/j.1469-8986.2005.00274.x

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115(3), 149-161. doi: http://dx.doi.org/10.1016/j.bandl.2010.07.006

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 101(6), 3728-3740. doi: 10.1121/1.418332

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of Contextual Predictability and Transitional Probability on Eye Movements During Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 31(5), 862-877. doi:10.1037/0278-7393.31.5.862

Gahl, S. (2008). Time and Thyme are not Homophones: The Effect of Lemma Frequency on Word Durations in Spontaneous Speech. *Language*, 84(3), 474–496. doi:10.1353/lan.0.0035

Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, 80(4), 748-775. doi: 10.1353/lan.2004.0185

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66(4), 789-806. doi: http://dx.doi.org/10.1016/j.jml.2011.11.006

Goldrick, M., Ross Baker, H., Murphy, A., & Baese-Berk, M. (2011). Interaction and representational integration: Evidence from speech errors. *Cognition*, 121(1), 58-72. doi: http://dx.doi.org/10.1016/j.cognition.2011.05.006

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, 58(3), 787-814. doi: http://dx.doi.org/10.1016/j.jml.2007.07.001

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. Chicago Linguistics Society (CLS-99), 151-166.

Hahn, L. W. (2012). Measuring local context as context-word probabilities. *Behavior Research Methods*, 44(2), 344-360. doi: 10.3758/s13428-011-0148-y

Hand, C.J., Miellet, S., O'Donnell, P.J. & Sereno, S.C. (2010). The Frequency-Predictability Interaction in Reading: It Depends Where Your're Coming From. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1294-1313. doi: 10.1037/a0020363

Hanique, I., & Ernestus, M. (2011). Final /t/ reduction in Dutch past-participles: The role of word predictability and morphological decomposability. In *Proceedings of the 12th Annual Conference of the International Speech Communication Association* (Interspeech 2011), Florence, Italy (pp. 2849-2852).

Hanique, I., Schuppler, B., & Ernestus, M. (2010). Morphological and predictability effects on schwa reduction: The case of Dutch word-initial syllables. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association* (Interspeech 2010), Makuhari, Japan (pp. 933-936)

Hartshorne, J. K., & Germine, L. T. (2015). When Does Cognitive Functioning Peak? The Asynchronous Rise and Fall of Different Cognitive Abilities Across the Life Span. *Psychological Science,* 26(4), 433 - 443. doi: 10.1177/0956797614567339

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, 115(5), 1090-1103. doi: http://dx.doi.org/10.1016/j.clinph.2003.12.020

Huey, E.B. (1908). *The psychology and pedagogy of reading*. Cambridge, MA: MIT Press.

Huestegge, L., Radach, R., Corbic, D., & Huestegge, S. M. (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision Research*, 49(24), 2948-2959. doi: http://dx.doi.org/10.1016/j.visres.2009.09.012

Hyönä, J., & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 21(6), 1430-1440. http://dx.doi.org/10.1037/0278-7393.21.6.1430

Inhoff, A. W., Solomon, M., Radach, R., & Seymour, B. A. (2011). Temporal dynamics of the eye–voice span and eye movement control during oral reading. *Journal of Cognitive Psychology*, 23(5), 543-558. doi: 10.1080/20445911.2011.546782

Jacobs, C. L., Yiu, L. K., Watson, D. G., & Dell, G. S. (2015). Why are repeated words produced with reduced durations? Evidence from inner speech and homophone production. *Journal of Memory and Language*, 84, 37-48. doi: 10.1016/j.jml.2015.05.004

Janse, E., & Jesse, A. (2014). Working memory affects older adults' use of context in spoken-word recognition. *Quarterly Journal of Experimental Psychology*, 67(9), 1842-1862. doi: 10.1080/17470218.2013.879391

Janssen, N., & Barber, H. A. (2012). Phrase Frequency Effects in Language Production. *PLoS ONE,* 7(3): e33202. doi:10.1371/journal.pone.0033202

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 824-843. doi: 10.1037/0278-7393.20.4.824

Joseph, H., Nation, K., & Liversedge, S. P. (2013). Using eye movements to investigate word frequency effects in children's sentence reading. *School Psychology Review*, 42(2), 207-222.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In J. L. Bybee & P.J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229-254). Amsterdam: John Benjamins.

Kahn, J. M., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*, 67(3), 311-325. doi: http://dx.doi.org/10.1016/j.jml.2012.07.002

Kemper, S., Bontempo, D., Schmalzried, R., McKedy, W., Tagliaferri, B., & Kieweg, D. (2014). Tracking reading: dual task costs of oral reading for young versus older adults. *Journal of Psycholinguist Research*, 43(1), 59-80. doi: 10.1007/s10936-013-9240-z

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology,* 16, 262-284. doi: 10.1080/09541440340000213

Kuperman, V., & Bresnan, J. (2012). The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language*, 66(4), 588-611. doi: http://dx.doi.org/10.1016/j.jml.2012.04.003

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception & Performance,* 39(3), 802-823. doi: 10.1037/a0030859

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6(3), 172–187. doi: 10.1177/002383096300600306

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2), 313-318. doi: http://dx.doi.org/10.1016/j.cognition.2012.09.010

McDonald, S. A., & Shillcock, R. C. (2003a). Eye Movements Reveal the On-Line Computation of Lexical Probabilities During Reading. *Psychological Science*, 14(6), 648-652. doi: 10.1046/j.0956-7976.2003.psci_1480.x

McDonald, S. A., & Shillcock, R. C. (2003b). Low-level predictive inference in reading: the influence of transitional probabilities on eye movements. *Vision Research*, 43(16), 1735-1751. doi: http://dx.doi.org/10.1016/S0042-6989(03)00237-2

Moers, C., Janse, E., & Meyer, A. S. (2015). Probabilistic reduction in reading aloud: A comparison of younger and older adults. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.), *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*. London: International Phonetics Association.

Morrison, C. M., Hirsh, K. W., Chappell, T., & Ellis, A. W. (2002). Age and age of acquisition: An evaluation of the cumulative frequency hypothesis. *European Journal of Cognitive Psychology*, 14(4), 435-459. doi: 10.1080/09541440143000159

Mousikou, B., & Rastle, K. (2015). Lexical frequency effects on articulation: A comparison of picture naming and reading aloud. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.01571

Nation, K., & Snowling, M. J. (1998). Individual Differences in Contextual Facilitation: Evidence from Dyslexia and Poor Reading Comprehension. *Child Development*, 69(4), 996-1011. doi: 10.1111/j.1467-8624.1998.tb06157.x.

Navarrete, E., Basagni, B., Alario, F. X., & Costa, A. (2006). Does word frequency affect lexical selection in speech production? *The Quarterly Journal of Experimental Psychology*, 59(10), 1681-1690. doi: 10.1080/17470210600750558

Oostdijk, N. (2002). The design of the Spoken Dutch Corpus. In P. Peters, P. Collins & A. Smith (Eds.), *New Frontiers of Corpus Research* (pp. 105-112). Amsterdam: Rodopi.

Perea, M., Soares, A. P., & Comesana, M. (2013). Contextual diversity is a main determinant of word identification times in young readers. *Journal of Experimental Child Psychology*, 116(1), 37-44. doi: 10.1016/j.jecp.2012.10.014

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, 11(4), 357-383. doi: 10.1080/10888430701530730

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005a). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, 62(2-4), 146-159. doi: 10.1159/000090095

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005b). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, 118(4), 2561-2569. doi: http://dx.doi.org/10.1121/1.2011150

Praat: Doing Phonetics by Computer, Version 5.3. Boersma, P. & Weenink, D., (2012).

R Development Core Team (2014). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org

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

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422. doi: 10.1037/0033-2909.124.3.372

Rayner, K., & Clifton Jr, C. (2009). Language processing in reading and speech perception is fast and incremental: Implications for event-related potential research. *Biological Psychology*, 80(1), 4-9. doi: http://dx.doi.org/10.1016/j.biopsycho.2008.05.002

Rayner, K., & Reichle, E. D. (2010). Models of the reading process. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 787-799. doi: 10.1002/wcs.68

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, 21(3), 448-465. doi: 10.1037/0882-7974.21.3.448

Reali, F., & Christiansen, M. H. (2007). Word chunk frequencies affect the processing of pronomial object-relative clauses. *The Quarterly Journal of Experimental Psychology*, 60(2), 161 - 170. doi: 10.1080/17470210600971469

Revill, K. P., & Spieler, D. H. (2012). The effect of lexical frequency on spoken word recognition in young and older listeners. *Psychology and Aging, 27*(1), 80-87. doi: 10.1037/a0024113

Robert, C., Mathey, S., & Postal, V. (2009). Différences liées à l'âge dans la reconnaissance visuelle des mots chez l'adulte. *European Review of Applied Psychology*, 59(2), 139-151. doi: http://dx.doi.org/10.1016/j.erap.2008.08.001

Rodriguez-Aranda, C., & Jakobsen, M. (2011). Differential contribution of cognitive and psychomotor functions to the age-related slowing of speech production. *Journal of the International Neuropsychological Society*, 17(5), 807-821. doi: 10.1017/s1355617711000828

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, 274(5294), 1926-1928. doi: 10.1126/science.274.5294.1926

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103(3), 403-428. doi: 10.1037/0033-295x.103.3.403

Schroeder, S. (2011). What readers have and do: Effects of students' verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology*, 103(4), 877-896. doi: 10.1037/a0023731

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523-568. doi: 10.1037/0033-295x.96.4.523

Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *Neuroreport*, 9(10), 2195-2200.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133(1), 140-155. doi: http://dx.doi.org/10.1016/j.cognition.2014.06.013

Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition*, 55(2), 151-218. doi: http://dx.doi.org/10.1016/0010-0277(94)00645-2

Shriberg, E., & Stolcke, A. (1996). Word predictability after hesitations: A corpus-based study. In *Proceedings of the International Conference on Spoken Language Processing* 3: 1868-1871.

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: Learning, Memory, and Cognition*, 37(3), 776-784. doi: 10.1037/a0022531

Smith, N. J., & Levy, R. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pp. 1637-1642.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302-319. doi: http://dx.doi.org/10.1016/j.cognition.2013.02.013

Spieler, D. H., & Balota, D. A. (2000). Factors influencing word naming in younger and older adults. *Psychology and Aging,* 15(2), 225-231. doi:10.1037/0882-7974.15.2.225

Stoltzfus, E. R., Hasher, L., Zacks, R. T., Ulivi, M., & Goldstein, D. (1993). Investigations of inhibition and interference in younger and older adults. *Journal of Gerontology: Psychological Sciences*, 48, 179-188. doi: 10.1093/geronj/48.4.P179

Thompson, S. P., & Newport, E. L. (2007). Statistical Learning of Syntax: The Role of Transitional Probability. *Language Learning and Development*, 3(1), 1-42. doi: 10.1080/15475440709336999

Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*, 1(2), 147 -165. doi: 10.1515/LANGCOG.2009.008, Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing Advantages of Lexical Bundles: Evidence From Self-Paced Reading and Sentence Recall Tasks. *Language Learning*, 61(2), 569-613. doi: 10.1111/j.1467-9922.2010.00622.x

Trouvain, J. (2003). *Tempo Variation in Speech Production. Implications for Speech Synthesis* (Doctoral dissertation, Saarland University). Retrieved from http://www.coli.uni-saarland.de/~trouvain/publications.html

Trouvain, J., & Grice, M. (1999). The effect of tempo on prosodic structure. In *Proceedings of the 14th International Congress of Phonetic Sciences* (ICPhS), San Francisco, USA. Retrieved from http://www.coli.uni-saarland.de/~trouvain/publications.html

Valle, A., Binder, K. S., Walsh, C. B., Nemier, C., & Bangs, K. E. (2013). Eye Movements, Prosody, and Word Frequency Among Average- and High-Skilled Second-Grade Readers. *School Psychology Review,* 42(2), 171-190.

Vorstius, C., Radach, R., & Lonigan, C. J. (2014). Eye movements in developing readers: A comparison of silent and oral sentence reading. *Visual Cognition*, 22(3-4), 458-485. doi: 10.1080/13506285.2014.881445

Wang, H.-C., Pomplun, M., Chen, M., Ko, H., & Rayner, K. (2010). Estimating the effect of word predictability on eye movements in Chinese reading using latent semantic analysis and transitional probability. *The Quarterly Journal of Experimental Psychology*, 63(7), 1374-1386. doi: 10.1080/17470210903380814

Weismer, G., & Liss, J. M. (1991). *Speech motor control and aging*. In D. Ripich (Ed.), *Handbook of geriatric communication disorders* (pp. 205–225). Austin, TX: Pro-Ed.

Whitford, V., & Titone, D. (2013). The effects of reading comprehension and launch site on frequency–predictability interactions during paragraph reading. *The Quarterly Journal of Experimental Psychology,* 67(6), 1151-1165. doi: 10.1080/17470218.2013.848216

Wingfield, A., Alexander, A. H., & Cavigelli, S. (1994). Does memory constrain utilization of top-down information in spoken word recognition? Evidence from normal aging. *Language and Speech*, 37, 221–235. doi:10.1177/002383099403700301

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, 72, 37-48. doi: http://dx.doi.org/10.1016/j.jml.2013.12.003

York, R. (2012). Residualization is not the answer: Rethinking how to address multicollinearity. *Social Science Research*, 41(6), 1379-1386. doi: http://dx.doi.org/10.1016/j.ssresearch.2012.05.014

Zevin, J. D., & Seidenberg, M. S. (2004). Age-of-acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. *Memory & Cognition*, 32(1), 31-38. doi: 10.3758/bf03195818

Ziegler, J. C., & Goswami, U. (2005). Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory. *Psychological Bulletin*, 131(1), 3-29. doi: 10.1037/0033-2909.131.1.3

CHAPTER 3

PROBABILISTIC REDUCTION IN READING ALOUD:

A COMPARISON OF YOUNGER AND OLDER ADULTS

This chapter has been adapted from:
Moers, C., Janse, E., & Meyer, A. S. (2015). Probabilistic reduction in reading aloud: A comparison of younger and older adults. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.), *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*. London: International Phonetics Association.

Abstract

Frequent and predictable words are generally pronounced with less effort and are therefore acoustically more reduced than less frequent or unpredictable words. Local predictability can be operationalized by Transitional Probability (TP), which indicates how likely a word is to occur given its immediate context. We investigated whether and how probabilistic reduction effects on word durations change with adult age when reading aloud content words embedded in sentences. The results showed equally large frequency effects on verb and noun durations for both younger ($M_{age} = 20$ years) and older ($M_{age} = 68$ years) adults. Backward TP also affected word duration for younger and older adults alike. ForwardTP, however, had no significant effect on word duration in either age group. Our results resemble earlier findings of more robust BackwardTP effects compared to ForwardTP effects. Furthermore, unlike often reported decline in predictive processing with aging, probabilistic reduction effects remain stable across adulthood.

Probabilistic Reduction in Reading Aloud: A Comparison of Younger and Older Adults

Introduction

The *Probabilistic Reduction Hypothesis* (Jurafsky, Bell, Gregory, & Raymond, 2001) states that more probable words are acoustically more reduced than less probable words. The probability of a word is influenced by its frequency (a word's prior probability) and its frequency in specific contexts (a word's contextual probability). The latter can be measured by Transitional Probability (TP), which indicates the probability of a word given its right or left neighboring word. Word frequency and transitional probability have been shown to influence acoustic reduction (involving segment deletion or weakening and shortened word duration) in several corpus studies of conversational speech (e.g., Bell et al., 2009; Pluymaekers, Ernestus, & Baayen, 2005; Hanique & Ernestus, 2011).

Apart from their influence on acoustic properties in spontaneous speech production, word frequency and transitional probability influence reading behavior. This has been shown in several studies on frequency and predictability effects on eye movements (Frisson, Rayner & Pickering, 2005; McDonald & Shillcock, 2003). The more likely a word is to occur (given a specific context), the less time the eyes spend on it and the more easily that word can be processed.

Whereas frequency and transitional probability effects have been replicated in several studies investigating silent reading, little is known about these effects on spoken word durations and eye movements in reading aloud. Reading aloud is a complex and demanding task involving the orchestration of several subcomponents: The visual recognition of words, the planning of speech and the actual articulation of the read words need to be synchronized efficiently (Ashby, Yang, Evans, & Rayner, 2012). Therefore, articulation of a word while reading aloud may differ from spontaneous speech production in that articulation can be influenced by orthographic properties of the text such as punctuation, and by simultaneous visual processing of words further ahead in the text (Inhoff, Solomon, Radach, & Seymour, 2011).

Two studies used a reading-aloud task when investigating probabilistic reduction effects: Hanique and Ernestus (2011) investigated predictability effects on final /t/-reduction in different speech registers (conversational speech, semi-formal interview speech and formal speech, i.e., reading aloud). They found probabilistic reduction effects for all registers, but reduction effects were larger for more spontaneous speech. Secondly, Baker and Bradlow (2009) compared effects of second-mention reduction in plain reading aloud to those in clear reading aloud. Reduction effects occurred in both reading styles, but effects were again larger in the less formal speaking style. Note, that neither of these studies investigated effects of TPs between two words on word durations. Thus, there is only limited evidence that probabilistic reduction effects transfer to reading aloud. Knowledge about these effects would give important insights for advancing theories of probabilistic reduction in speech as well as developing models of reading aloud.

Importantly, thus far, all studies on probabilistic reduction effects have investigated student populations or middle-aged adults. However, probabilistic reduction effects may change in size with advanced adult age. Age has been found to modulate word frequency effects in several studies of silent reading or speech production. However, results of these studies did not converge: On the one hand, older adults have been reported to show larger word frequency effects than younger adults in silent reading (Rayner et al., 2006). On the other hand, increased language experience has been shown to lead to decreased frequency effects in picture naming (Gollan, Montoya, Cera, Sandoval, 2008). Studies investigating context effects have also yielded inconsistent age effects: Older adults show either smaller context effects in speech comprehension (Federmeier, Kutas, & Schul, 2010), or make equal use of context to predict upcoming words in silent reading as a younger group (Rayner et al., 2006). Furthermore, it is unclear whether age differences in frequency and predictability effects found in silent reading, speech comprehension or picture naming transfer to probabilistic reduction effects in read-aloud speech.

In the present study, we investigated the following questions: First, do probabilistic reduction effects as found in conversational speech generalize to reading aloud? Second, how do probabilistic reduction effects develop across adult age?

Methods

To study the effects of word frequency and transitional probability (TP) on reading behavior in younger and older adults, we designed an eye-tracking experiment that included two reading tasks: silent reading and reading aloud. The speech recordings from the oral reading condition are analyzed in the current study.

Participants

Thirty older and thirty younger adults participated in the experiment (recruited via the participant pool of the Max Planck Institute for Psycholinguistics) and were paid for their participation. Older participants' mean age was 68.37 years (SD = 4.49; range: 62 to 78 years), younger participants' mean age was 20.83 years (SD = 2.70; range: 18 to 27 years). All participants were native speakers of Dutch, with no self-reported history of language impairments or neurological problems. Furthermore, all participants had normal or corrected-to-normal vision, which was tested prior to the reading experiment to ensure that participants could properly read from the computer screen.

Materials and procedure

Testing and recording took place in a sound-treated booth at the Max Planck Institute for Psycholinguistics. At the start of a session, participants received written information about the upcoming reading task and gave written consent. Participants were instructed to read sentences from a computer screen. They saw two example sentences which they had to read out loud in order to familiarize themselves with the task. Sentences were presented on a single line in an 18 point font (Times New Roman) as black letters on a white background, vertically centered, but horizontally left-aligned. After participants had finished reading a sentence, they had to press a mouse button to proceed to the next sentence. An untimed yes/no comprehension question occurred after one quarter of the sentences. Response accuracy for the comprehension questions was generally high (mean accuracy being 96% for the younger and 93% for the older group).

Participants read 240 Dutch sentences in total, divided over the silent and reading aloud conditions, which alternated in four blocks (block 1: 60 silent trials, block 2: 60 aloud trials, block 3: 60 silent trials, block 4: 60 aloud trials; or reverse order). Each sentence included a target noun-verb combination. Transitional probabilities for these combinations, as well as word frequencies for each noun and verb, were drawn from the 44 million words SUBTLEX-NL corpus (Keuleers, Brysbaert, & New, 2010). Half of the 240 items were infinitives (e.g., “afspraak maken” – *to make an appointment)* and the other half were past participles (e.g., “fouten gemaakt” – *made mistakes*). Forward transitional probability indicates the predictability of the verb from the noun and was computed by taking the total number of occurrences of the noun-verb combination divided by the number of occurrences of the noun. Log-transformed forward transitional probabilities for the 240 noun-verb combinations followed a uniform distribution and thus varied continuously from very high log probability (e.g., logFTP = -0.31 for “afscheid nemen” – *to say farewell*) to very low probability (e.g., logFTP = -9.88 for “tijd kiezen” – *to pick a time*). For each noun-verb combination, we also calculated backward transitional probability, being the predictability of the noun given the verb.

The noun-verb combinations were embedded in neutral, non-predictive sentences that did not contain words that were semantically related to the upcoming target noun or verb. Sentence length ranged from seven to twelve words (including the noun-verb combination). Sentences were constructed such that the target combinations were neither sentence-initial nor sentence-final to avoid initial or final lengthening effects.

Annotations

Participants' speech recordings were annotated by one of eight experienced transcribers in order to derive total sentence reading times and word durations (in milliseconds) for each target noun and verb. Annotations were done using the software Praat (Boersma & Weenink, 2014). Disfluent or misread sentences were excluded from the analysis. A few sample sentences of each transcriber and each participant were evaluated for accuracy by the first author. Annotation accuracy was evaluated as being very high. Speech rate for each sentence was measured by dividing the total sentence reading time by numbers of syllables in the orthographic sentence representation.

Results

Statistical Analysis

To investigate probabilistic reduction effects on word durations, we fitted linear mixed-effects models in R version 3.1.1. (R Development Core Team, 2014), using the lmer-function from package lme4 (Bates, Maechler, & Bolker, 2014). Each model included word length (in number of characters for either the verb or noun), speech rate (in syllables per second, log-transformed), and verb type (infinitive or past participle) as control variables, as well as word frequency (log frequency per million words of either the verb or noun), transitional probability (either log forward or log backward TP) and age group (older versus younger group) as predictor variables. Word durations were log-transformed. Numerical predictors were centered around their mean to reduce non-essential collinearity (Wurm & Fisicaro, 2014). All regression models contained participants and items as random effects, as well as by-subject random slopes for frequency and TP (uncorrelated) and by-item random slopes for age group (Barr, Levy, Scheepers, & Tily, 2013, following a design-driven approach). P-values were obtained by likelihood-ratio tests (comparing nested models). Since all predictor and control variables were theoretically motivated, no model stripping procedure was applied. As our focus was on possible age-related changes in the effect of predictability after controlling for effects of word frequency, we did not remove collinearity between the frequency and TP variables by residualisation (cf. Wurm & Fisicaro, 2014, for a discussion). Instead, we always fitted an initial model that solely investigated the effect of frequency (and its interaction with age group), and only afterwards added either ForwardTP or BackwardTP (to the fixed and to the random part). No changes in frequency effects or frequency by age group interactions occurred in any model due to the addition of the TP variable. The models reported below are based on 6543 trials (out of 7200 trials in total, i.e., 9.1% data loss due to exclusion of disfluent and misread sentences). Older adults took overall more time to read sentences (M = 2797 ms, SD = 548) than younger adults (M = 2429 ms, SD = 442), and thus older adults had slightly slower speech rates (M = 5.14 syllables/second, SD = 0.85, Range = 2.47-11.33) than younger adults (M = 5.86 syllables/second, SD = 0.91, Range = 2.47-11.33). Consequently, older adults had longer verb durations (OA: M = 493 ms, SD = 142, Range = 149-1198) than younger adults (M = 414 ms, SD = 119, Range = 90-1037), as well as longer noun durations (OA: M = 398 ms, SD = 117, Range = 139-954; YA: M = 349 ms, SD = 97, Range = 110-743).

Verb duration

A summary of the best-fitting linear mixed-effects model for the verb duration analysis is shown in Table 1. The model showed significant simple effects of word length, speech rate, verb type, and verb frequency, as well as a significant interaction of ForwardTP by age group. As expected, orthographically longer words had longer spoken durations, higher global speech rate led to shorter word durations, and verbs that occur often were acoustically more reduced than verbs with low frequency of occurrence. Past participle verbs had longer durations than verbs of the infinitive type. Furthermore, the older adults had longer word durations than the younger adults. ForwardTP showed no influence on verb duration for the younger adults (i.e., the group mapped on the intercept). Older adults reduced verbs more when they were more predictable from the noun than younger adults, as indicated by the age group by ForwardTP interaction. However, this interaction is difficult to interpret, as rerunning the same model with the older group mapped on the intercept did not show an effect of ForwardTP either. Hence, the effect of ForwardTP is overall not strong enough to significantly influence word duration.

Table 1. Estimates, standard errors and *t* values for mixed-effects model fitted for the verb duration data; significance ($p<0.05$) indicated by asterix (*).

Variable	**β**	**SE**	***t***
WordLength	0.09223	0.00540	17.1*
LogSpeechRate	-0.55765	0.01640	-34.0*
VerbForm (PP)	0.07588	0.01598	4.7*
LogFreqVerb	-0.03622	0.00955	-3.8*
LogForwardTP	0.00281	0.00404	0.7
Group (Older)	0.09519	0.01846	5.2*
LogFreqVerb by Group	0.00295	0.00418	0.7
LogForwardTP by Group	-0.00393	0.00187	-2.1*

Noun duration

A summary of the best-fitting linear mixed-effects model for the analysis of noun durations is shown in Table 2. The model showed significant simple effects of word length, speech rate, noun frequency, and BackwardTP, but no significant interactions between either noun frequency and age group or BackwardTP and age group. As expected, longer words had longer durations, higher speech rate led to shorter word durations, and frequent nouns were acoustically more reduced than infrequent nouns. Moreover, the predictability of a noun from the verb influenced noun durations such that more predictable nouns were more reduced than less predictable nouns. This effect occurs for younger and older adults alike, as indicated by the non-significant age group by BackwardTP interaction. Figure 1 illustrates age group and TP effects.

Table 2. Estimates, standard errors and *t* values for mixed-effects model fitted for the noun duration data; significance (*p*<0.05) indicated by asterix (*).

Variable	**β**	**SE**	***t***
WordLength	0.10577	0.00523	20.2*
LogSpeechRate	-0.45300	0.01682	-26.9*
VerbForm (PP)	-0.01560	0.01792	-0.9
LogFreqNoun	-0.02515	0.01006	-2.5*
LogBackwardTP	-0.01439	0.00567	-2.5*
Group (Older)	0.06368	0.01426	4.5*
LogFreqNoun by Group	-0.00552	0.00385	-1.4
LogBackwardTP by Group	0.00136	0.00226	0.6

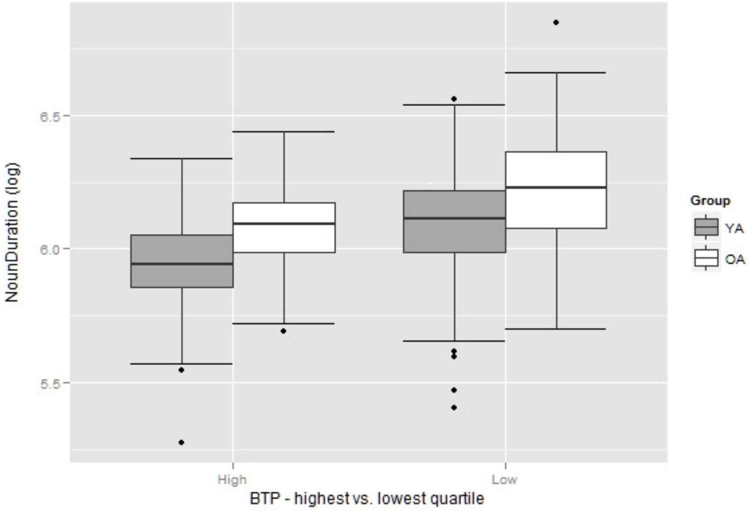

Figure 1. Boxplot illustrating the effect of age group (YA vs. OA) and BackwardTP on noun duration; TP is shown for the first (lowest) and fourth (highest) quartile.

Discussion

We investigated whether effects of probabilistic reduction, as found in studies of conversational speech, generalized to reading aloud. This study showed that the demands and complexity of oral reading did not cancel frequency and local predictability effects.

We further investigated whether frequency and transitional probability effects for noun-verb combinations change across adult age. BackwardTP indicated the predictability of the noun given the following verb, and ForwardTP always was the verb's predictability given the preceding noun. While word frequency influenced probabilistic reduction in all models, effects of transitional probability were mixed. More predictable nouns were more reduced than less predictable nouns, and this effect of BackwardTP was equal in size in younger and older adults. ForwardTP, however, had no influence on acoustic reduction of the verb (note though, that an age-by-TP interaction occurred without a simple ForwardTP effect for either group). Hence, effects of probabilistic reduction were generally similar in size in younger and older adults.

Our finding that BackwardTP effects on duration are more robust than those of ForwardTP are in line with earlier results (Bell et al., 2009). Bell and colleagues found BackwardTP effects for both content and function words in conversational speech, while ForwardTP influenced only the production of high-frequency function words. The authors attributed this finding to differential lexical access for content versus function words in speech production. Since we investigated TP effects on content words only (verbs and nouns), it is not surprising that we only found consistent effects of BackwardTP in both age groups.

The influence of frequency and TP on speech production is often explained by a coordination mechanism that links lexical access and articulation, such that the progress of lexical retrieval is synchronized with speed of articulation (Bell et al., 2009). Our findings of equal probabilistic reduction effects for younger and older adults suggest that the coordination between those levels is well maintained in later adulthood.

Furthermore, the results of equal TP effects for older and younger adults suggest that local TP effects differ from higher-level semantic prediction as frequently measured by cloze or sentence completion tasks (Federmeier et al., 2010; Smith & Levy, 2011), in which older adults usually show smaller predictability effects than younger adults. Clearly, more research is needed to distinguish between TP effects and cloze predictability effects and to explore their possible change across the lifespan.

References

Ashby, J., Yang, J., Evans, K. C., & Rayner, K. (2012). Eye movements and the perceptual span in silent and oral reading. *Attention, Perception, & Psychophysics*, 74(4), 634-640.

Baker, R. E., & Bradlow, A. R. (2009). Variability in Word Duration as a Function of Probability, Speech Style, and Prosody. *Language and Speech,* 52, 391-413.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255-278.

Bates, D., Maechler, M., & Bolker, B. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. http://lme4.r-forge.r-project.org/

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60, 92-111.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115, 149-161.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of Contextual Predictability and Transitional Probability on Eye Movements during Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 31, 862-877.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, 58, 787-814.

Hanique, I., & Ernestus, M. (2011). Final /t/ reduction in Dutch past-participles: The role of word predictability and morphological decomposability. *Proceedings of 12th Interspeech*, Florence, 2849-2852.

Inhoff, A. W., Solomon, M., Radach, R., & Seymour, B. A. (2011). Temporal dynamics of the eye–voice span and eye movement control during oral reading. *Journal of Cognitive Psychology*, 23, 543-558.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In: Bybee, J. L., Hopper, P. J. (Eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins, 229 - 254.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods,* 42, 643-650.

McDonald, S. A., & Shillcock, R. C. (2003). Eye Movements Reveal the On-Line Computation of Lexical Probabilities During Reading. *Psychological Science,* 14, 648-652.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, 118, 2561-2569.

Praat: Doing Phonetics by Computer, Boersma, P., & Weenink, D., (2014). http://www.fon.hum.uva.nl/praat/

R Development Core Team 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging,* 21, 448-465.

Smith, N. J., & Levy, R. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pp. 1637-1642.

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, 72, 37-48.

CHAPTER 4

PREDICTABILITY EFFECTS IN SILENT AND ORAL READING OF YOUNGER AND OLDER ADULTS

This chapter has been adapted from:
Moers, C., Meyer, A.S., & Janse, E. (under revision). Predictability Effects in Silent and Oral Reading of Younger and Older Adults.

Abstract

Efficient silent reading relies on the synchronization of many cognitive processes including visual perception, lexical processing and semantic integration. The synchronization process is even more demanding in oral reading, as it also involves speech production. Word frequency and predictability are known to influence reading efficiency. Frequent and predictable words are processed faster and skipped more often than infrequent or unpredictable words. Older adults have been shown to achieve reading efficiency by using different reading strategies from younger adults, with older adults showing more word skipping and more regressions to previous words than younger adults in silent reading. It is not clear, however, whether those age-related changes in the reading process generalize to reading aloud, and specifically how frequency and predictability effects may be affected. Consequently, this study investigates age-related changes in the use of probabilistic information across reading tasks. Predictability was operationalized by transitional probabilities between neighboring words. The analysis of fixation durations indicated that gazes were generally longer in reading aloud than in silent reading. Predictability effects were similar in size across reading tasks and age groups, but older adults had larger reading task differences and larger word frequency effects on total fixation durations than younger adults. The opposite pattern was seen for skipping rates; here older adults had smaller reading task differences and smaller frequency effects than younger adults. Taken together, our results show that reading task modifies effects of aging on reading patterns, while predictability effects are largely unaffected by reading task.

Predictability Effects in Silent and Oral Reading of Younger and Older Adults

Introduction

For many people reading is a highly practiced skill. Yet, reading is a complex task because efficient reading relies on several components such as visual perception, lexical processing, and semantic integration, which need to be executed in a short amount of time. Oral reading additionally involves speech production, making the synchronization of the components even more demanding. Two variables that contribute to reading efficiency are the frequencies of words and their predictability from context. Highly frequent or predictable words can be processed faster than infrequent or unpredictable words. It is not clear whether and how the ability to make predictions and inferences during reading changes over the adult life span. In this study we investigate whether younger and older adults differ in the use of predictive cues by looking at local word-to-word predictability (transitional probability). We specifically investigate whether frequency and predictability effects differ in reading aloud versus silent reading, due to the additional production level in reading aloud. Possibly, the processing demands of synchronizing visual perception, lexical processing and pronunciation affect older adults more than younger adults.

Frequency and predictability effects in language processing

Word frequency is a robust predictor of processing difficulty across different language tasks, such that words we hear or see often are easier to process than rare words. Specifically, high-frequency words are recognized faster than low-frequency words in language comprehension and produced more efficiently in language production (for comprehension studies see e.g., Allen, Smith, Lien, Grabbe, & Murphy, 2005; Hauk & Pulvermüller, 2004; Revill & Spieler, 2012; Sereno, Rayner, & Posner, 1998; for production studies see e.g., Bell, Brenier, Gregory, Girand,

& Jurafsky, 2009; Jescheniak & Levelt, 1994; Navarrete, Basagni, Alario, & Costa, 2006). Likewise, in reading research, numerous studies have shown that word frequency influences the reader's eye movements. High-frequency words are fixated for shorter periods of time and skipped more often than low-frequency words (for frequency effects in silent reading see e.g., Hand, Millet, O'Donnell, & Sereno, 2010; Kliegl, Grabner, Rolfs, & Engbert, 2004; for reading aloud see e.g., Valle, Binder, Walsh, Nemier, & Bangs, 2013; Vorstius, Radach, & Lonigan, 2014; for general reviews see Rayner, 1998, and Rayner, 2009). Moreover, naming single printed words is faster for high-frequency words than for low-frequency words (e.g., Spieler & Balota, 2000).

Importantly, processing speed and effort are not only influenced by how often single words occur, but also by a word's likelihood in its context. Contextual facilitation has been evidenced in a number of reading studies, such that words that are more predictable from their context are fixated for shorter periods of time, are skipped more often, and need less time or effort for semantic integration compared to less predictable words (cf. Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Fernández, Shalom, Kliegl, & Sigman, 2013; Hand et al., 2010; Schotter, Bicknell, Howard, Levy, & Rayner, 2014; Whitford & Titone, 2014; for discussions see Staub, 2015, and Luke & Christianson, 2016). The likelihood of a word in its immediate context is Transitional Probability (TP), which can be calculated from large language corpora. Transitional probabilities estimate how likely a word is to occur given its right or left neighboring word. Readers are sensitive to this kind of probabilistic information and may use it to predict words from their local context. Hence, readers fixate longer on low-TP words (those words that are less predictable from their local context) than on high-TP words (cf. Frisson, Rayner, & Pickering, 2005; McDonald & Shillcock, 2003a, 2003b; Wang, Pomplun, Chen, Ko, & Rayner, 2010; but see Ong & Kliegl, 2008).

Studies that investigated transitional probability effects in reading have so far solely focused on student or middle-aged adult populations. Thus, while the effects of word frequency and predictability are well established, less is known about the way local probabilistic information in reading changes with age and language experience.

Age-related changes in frequency and predictability effects

Lexical frequency has been found to influence reading times in children from as young as 7 years (e.g., Blythe & Joseph, 2011; Aghababian & Nazir, 2000; Hyönä & Olson, 1995; Davies, Arnell, Birchenough, Grimmond, & Houlson, 2017). Given that probabilistic knowledge grows with language experience, frequency and predictability effects may change with age. A few studies have indeed documented differences in word frequency effects in older adults compared to younger adults (for a general review of age-related changes in frequency and predictability effects in reading see Gordon, Lowder, & Hoedemaker, 2016). However, these studies have led to conflicting conclusions: On the one hand, older adults have been reported to show decreased frequency effects in comparison to younger adults (Robert, Mathey, & Postal, 2009; Davies et al., 2017). One reason for why older age and increased language experience may lead to decreased frequency effects is because one is more likely to have come across infrequent language units with more experience, such that the processing of infrequent units may assimilate to that of frequent ones (cf. Burke & Shafto, 2008). On the other hand, older age has been found to be associated with larger effects of word frequency compared to younger age, which is typically found in reading studies (e.g., Kliegl et al., 2004; Kliegl, Nuthmann, & Engbert, 2006; Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006). Larger frequency effects in older adults can be accounted for by at least two mechanisms. First, lexicon size may be a mediating factor in that vocabulary size determines how long it takes to search through the mental lexicon in order to activate a particular word (e.g., Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014). Older adults often have larger vocabularies than younger adults (Ben-David et al., 2016; Verhaeghen, 2003), which typically entails knowing more low-frequency words. Consequently, older adults may need longer for lexical activation, particularly of low-frequency words (e.g., due to more competition between several low-frequency words) and therefore show increased frequency effects. Second, larger frequency effects for older adults than younger adults in reading may suggest that older adults tend to rely more on lexical information during word recognition to compensate for other age-related effects such as reduced visual acuity or attentional limitations which may impair character recognition (cf. Gordon et al., 2016).

Regarding age-related changes in contextual predictability effects, there is again conflicting evidence. Cheimariou (2016) and Rayner and colleagues (2006) report equal predictability effects for older and younger adults in silent reading. In contrast to these findings, older adults are reported to engage less in anticipation of upcoming words than younger adults in speech comprehension and in text processing (e.g., DeLong, Groppe, Urbach & Kutas, 2012; Federmeier, Kutas, & Schul, 2010; Federmeier, McLennan, De Ochoa, & Kutas, 2002). This difference may be due to different or more variable allocation of processing resources in older compared to younger adults in various language tasks (e.g., Stine-Morrow, Soederberg Miller, & Hertzog, 2006).

Importantly, all studies of age-related differences in predictability effects have operationalized predictability with the cloze procedure, in which participants are asked to complete sentence fragments with the word they think is most likely to follow. Cloze tasks capture higher-level predictions and semantic integration effects over longer stretches of preceding context. However, bigram-based predictability effects, such as TP effects, are more local by definition and may therefore differ from cloze effects (cf. Hahn, 2012; Ong & Kliegl, 2008; Smith & Levy, 2011; but see the discussion in Frisson et al., 2005; Shaoul, Baayen & Westbury, 2014). Working memory ability has been found to relate to the rapid use of preceding semantic sentence context (Janse & Jesse, 2014; Linderholm, 2002). Consequently, it is not clear whether older adults and younger adults only differ in their processing of *global* predictive cues due to age differences in working memory ability, or also of more *local* predictive cues. Any age effects on semantic integration and higher-level prediction may not necessarily generalize to how age groups use local word-to-word predictability. In sum, frequency and predictability effects occur in different language modalities, such as speech production and reading, and these effects may undergo age-related changes. However, the locus and degree of these age-related changes is not clear.

Silent reading versus reading aloud

Age-associated changes in silent reading are well-documented. Besides changes in processing of low-level visual properties of text, such as font size and the spatial frequency of letters (e.g., Jordan, McGowan, & Paterson 2014; Mund, Bell & Buchner, 2010; Rayner et al., 2006), linguistic processes have been found to undergo age-related changes as well. This is evidenced in changes in the effects of word length (e.g., Rayner et al., 2006), and syntactic difficulty (e.g., Kemper & Liu, 2007). Longer words and more difficult words (i.e., words in complex syntactic constructions, also infrequent words) are fixated for longer time periods, even more so the older the reader is. Yet, aging may not only slow the reading process down, but may also affect reading strategy (Rayner et al., 2006). Older readers skip target words more often, and later regress to them more frequently than younger readers (see also Kliegl, Grabner, Rolfs, & Engbert, 2004). This more ‘risky reading strategy’ of older adults may be an attempt to maintain fast reading speed in older age and compensate for slower lexical and parafoveal processing (Rayner et al., 2006).

To the best of our knowledge, no earlier studies have investigated age-related changes in eye-movements in oral sentence reading and compared them to silent reading. In addition to the processes needed for silent reading, reading aloud involves speech production and therefore adds retrieving and executing articulatory plans. As such, oral reading presents a more complex task than silent reading (e.g., Reynolds & Besner, 2006), with increased demands on smooth information transmission between different processes that, in turn, may also be subject to age-associated changes. Furthermore, note that reading aloud does not allow the reader the flexibility that silent reading does. That is, fluent oral reading requires a tight coordination between the input and output levels, such that the eye movements remain synchronized with the speech output (cf. Laubrock & Kliegl, 2015). Articulation is relatively slow, while the identification of words in skilled readers is amazingly fast. Consequently, readers typically look at and process a word in the text that is two to three words to the right of the word they are currently pronouncing (a typical eye-voice gap of 500 ms, Inhoff, Solomon, Radach, & Seymour, 2011). The 'risky reading strategy' of jumping back and forth between words is therefore limited in oral reading, and readers need to adopt a more incremental, word-to-word processing strategy. Consequently, if differences in reading strategies between age groups are reduced, any quantitative differences in lexical processing might show up more strongly. Age differences in frequency and predictability effects are hypothesized to be stronger in oral reading than in silent reading, because the task restrictions of oral reading leave less room for changes in processing strategies.

Note that in reading aloud there are additional levels at which frequency and TP effects could arise, such as during the retrieval or execution of articulatory routines. To the extent that frequency and TP effects on phonetic encoding and articulatory planning of word n affect eye movements, these effects would only be observed 'downstream', that is while the eyes are already inspecting for example word n+3, and hence are not measured in this study.

The current study

The present study had two major goals: First, we aimed to test the hypothesis that younger and older readers differed in the use of local probabilistic cues, that is, word frequency and transitional probability statistics. Second, we tested the hypothesis that age groups particularly differed in the size of frequency and predictability effects in reading aloud compared to silent reading, as the former imposes additional processing demands and leaves less room for changes in processing strategies. We set up an eye-tracking experiment to address the hypotheses above.

Three reading measures will be reported that have all been linked to frequency and predictability effects (i.e., gaze durations, skipping rates, total fixation durations). Generally, reading measures may be grouped into “early” measures (e.g., first-pass reading measures such as first-fixation duration, gaze duration, and first-pass skipping) versus “late” measures (e.g., total fixation duration, regression rate). Early measures reflect perceptual processes during first-pass reading, such as word length effects, frequency, and familiarity effects. Late measures additionally include cognitive top-down influences on reading comprehension, such as semantic wrap-up. We analyzed early and late measures as local predictability and reading condition differences may be reflected in early as well as later processing stages.

Methods

An eye-tracking study was conducted to study effects of word frequency and transitional probability (TP) on reading behavior in different age groups and in different reading modalities. In previous studies both forward TP and backward TP have been found to be predictive of eye fixation behavior in adults (e.g., McDonald & Shillcock, 2003b). While forward TP reflects the predictability of a target word from its preceding context, backward TP measures the dependency of a target word on upcoming context. As such, backward TP effects relate to concurrent processing and are part of parafoveal preview effects in reading. In the present paper we solely report the analysis of forward TP effects, as we were interested in prediction of words on the basis of previously processed words (but see Appendix A on the noun region data). Additional analyses of backward TP effects confirmed the findings concerning predictability by age interactions reported below (cf. Table 7 in Appendix A). Moreover, backward TP has been found to be particularly important in spoken word durations of content words (Bell et al., 2009). We have reported forward and backward TP effects in the speech data from the reading aloud task in a companion paper (Moers, Janse, & Meyer, 2015).

Materials and Design

240 Dutch noun-verb combinations with varying transitional probabilities were drawn from the 44 million words SUBTLEX-NL corpus, which is a collection of movie subtitles (for details see Keuleers, Brysbaert, & New, 2010). Half of the items consisted of combinations of nouns plus infinitives (e.g., “afspraak maken” – *to make an appointment)* and the other half were noun - past participle combinations (e.g., “fouten gemaakt” – *made mistakes*). These two types of combinations were chosen to be able to generalize over different noun-verb combinations. Form frequency of occurrence of each noun-verb combination was used to calculate its transitional probability. Thus the TP of the noun-infinitive combination *afspraak maken* might differ from the TP of the noun-past participle combination *afspraak gemaakt*, due to different frequencies of occurrence of the two constructions. We added verb type (infinitive versus past participle) to the statistical model described below as control variable. In our dataset, forward transitional probability indicates the predictability of the verb from the noun and was computed by taking the total number of occurrences of the noun-verb combination in SUBTLEX divided by the number the noun occurs. The materials were designed such that log-transformed forward transitional probabilities for the 240 noun-verb combinations followed a uniform distribution and thus varied continuously from very high log probability (Max = -0.31) to very low log probability (Min = -10.04). Furthermore, for each word in a noun-verb combination we obtained its relative frequency, which is the number of occurrences of the word per million words, and its word length in number of letters. Word length and word frequencies for both nouns and verbs vary continuously. For all predictor variables, descriptive statistics, including means, standard deviations, minima and maxima are given in Table 1. The selected noun-verb combinations were embedded in neutral sentences that did not contain words that were semantically related to the upcoming target noun or verb. Sentence length differed from seven to twelve words (including the noun-verb combination). The embedded target combinations were always preceded and followed by at least two other words, to avoid potential start-up and

wrap-up lengthening (Kuperman, Dambacher, Nuthman, & Kliegl, 2010).

The sentence materials were distributed over eight lists according to a fully crossed Latin-square design. That is, each participant was presented with each of the 240 sentences once, and the sentences were counterbalanced across silent reading and oral reading. SeAlthough the combination of sentences per block were different on each of the eight lists, mean frequencies and mean transitional probabilities of the noun-verb combinations were kept constant across the four blocks. Participants were randomly assigned to a list.

Table 1. Sentence material characteristics (descriptive statistics with mean, standard deviation (SD), minimum (Min), and maximum (Max)) for the reading task.

	Mean	**SD**	**Min**	**Max**
Verb - noun: total occurrence in SUBTLEX-NL	107.07	328.65	2	4372
Forward TP	0.035	0.078	0.000042	0.730323
Backward TP	0.018	0.062	0.0000992	0.814249
Frequency verb (per million)	190.36	289.12	16.8308	2677.854
Frequency noun (per million)	177.82	234.37	29.1337	1403.815
Word length verb	7.53	1.75	4	13
Word length noun	6.38	1.87	3	12
Total sentence length	9.03	1.04	7	12

Participants

Thirty community-dwelling older adults and thirty younger adults, mainly students from Radboud University Nijmegen, participated in the experiment. All participants were native speakers of Dutch, with no self-reported history of language impairments or neurological problems. They were recruited via the participant pool of the Max Planck Institute for Psycholinguistics and were paid for their participation (€12 for approximately 90 minutes). Older participants’ mean age was 68.37 years (SD = 4.49; range: 62 to 78 years), younger participants’ mean age was 20.83 years (SD = 2.70; range: 18 to 27 years). Older participants had larger vocabulary size compared to younger adults, as assessed by the Peabody Picture Vocabulary Test in Dutch (PPVT-III-NL; Dunn & Dunn, 2005; raw score M = 188.06, SD =

6.02 for older adults versus M = 171.87, SD = 11.36 for younger adults).

Participants wore their appropriate glasses or soft contact lenses if needed. Their vision was tested in two ways: far vision was tested over a three-meter distance with a Landolt C chart, and near-vision was tested using a Snellen chart in which print size was adjusted to the eye-screen distance of approximately 70 centimeters used in the experiment. According to the vision tests, all participants had normal or corrected-to-normal vision for both far and near vision and could therefore be expected to be able to easily read the sentences in the chosen font size from the computer screen.

Procedure

The experimental session lasted approximately 90 minutes, of which about 45 minutes were needed for completing the eye-tracking experiment. The experimental procedure was the same for both age groups. Testing took place at the Max Planck Institute for Psycholinguistics. At the start of a session, participants received written information about the nature and timing of the upcoming tasks after which they had time to ask questions. Then participants gave written consent. Participants were seated in a sound-treated booth. Participants were instructed to read the upcoming sentences for comprehension. They were told that yes/no-comprehension questions would be presented on the screen from time to time and that they should respond with "yes" or "no" by pressing either the left or right mouse button. After participants confirmed that they understood the instructions, they saw two example sentences which they had to read out loud in order to familiarize themselves with the task. One of these sentences was followed by an example comprehension question, and participants received feedback about the correct response.

Eye movements were recorded using an SR Research EyeLink1000 remote desktop eye-tracker with a 500 Hertz sampling rate. No chin rest or bite bar was used. Instead a target sticker that was placed on each participant's forehead helped in tracking head movements and changes in eye-screen distance. Calibration and validation of the eye tracker were performed using a five-point grid. Sentence viewing was binocular, but eye movements were recorded from the left or right eye depending on calibration (if calibration was poor for the left eye, the right eye was tracked). After successful calibration, the first test began with an instruction screen informing participants about the task (silent reading or reading aloud). On each of the following trials, a fixation cross was presented vertically centered, close to the left edge of the computer screen, for 500 milliseconds. The fixation cross then had to be fixated for at least 100 milliseconds in order to trigger the presentation of the upcoming target sentence. Sentences were presented vertically centered, horizontally left-aligned, with the first letter of the first word occurring exactly where the fixation cross had been placed before. Sentences were presented on a single line in an 18 point font (Times New Roman) as black letters on a white background. One degree of visual angle was equivalent to approximately 4.3 characters. After participants had finished reading a sentence, they clicked the mouse button to proceed to the next item. An untimed yes/no comprehension question about the content of the sentence came up after a quarter of the sentences randomly distributed across blocks. Response accuracy for the comprehension questions was generally high, and equaled 96% for the younger group and 93% for the older group.

The 240 sentences were presented to each participant in four blocks, which were separated by pauses. Half of the participants started with a reading-aloud block, which was followed by a silent-reading block, another reading-aloud block and a final silent-reading block. The other half of the participants went through the reading blocks in the reverse order. Participants were encouraged to take short breaks between blocks. Calibration and validation for the eye-tracker were repeated before each new block. Two practice sentences followed to ensure that participants switched to the right reading mode for each block.

Speech recordings were made during the reading aloud condition, and word durations for nouns and verbs were annotated manually. The analysis of the speech data is reported in Moers, Janse, and Meyer (2015).

Data selection

For inspection of the eye-tracking data we used the EyeLink Data Viewer software by SR Research. The software automatically defines separate interest areas for all words in a sentence. Following the conventions in the field, single fixations shorter than 80 milliseconds were merged with the next fixation if they occurred within one degree of visual angle to the next fixation, otherwise they were deleted. This was done as participants cannot extract much information in fixations as short as this (e.g., Frisson et al., 2005). Fixations longer than 800 milliseconds were excluded from the analysis (cf., Rayner et al., 2006). Additionally, we excluded target items from our analysis for which total sentence reading time exceeded 10 seconds.

Out of a total of 14400 trials in the experiment (60 participants times 240 trials) 13369 were included in the eye-movement dataset for analysis (i.e., about 7% data were lost due to tracking losses or unreliable/inaccurate tracking, e.g., in case of several blinks in a trial). Older adults had on average more fixations per sentence (M = 13, SD = 4 in oral reading, M = 10, SD = 3 in silent reading) than younger adults (M = 9, SD = 3 in oral reading, M = 8, SD = 3 in silent reading). Average fixation duration was similar for older (M = 272, SD = 45 ms for reading aloud, M = 232, SD = 43 ms for silent reading) and younger adults (M = 279, SD = 46 ms for reading aloud, M = 227, SD = 46 ms for silent reading).

The statistical analysis of the participants' fixation behavior included several dependent variables, which are all standard measures in reading research. In the following section, we report detailed results for two "early" reading measures and one "late" measure. We chose gaze duration (GD) and skipping rate (SR) as early measures, and total fixation duration (TD; includes regressive fixations on a critical region) as a late measure. Gaze duration and skipping rate are complementary in that the former captures those trials in which initial fixations were made on target verbs or nouns, while the latter includes trials that had no initial fixation. Gaze duration is the summed duration of all fixations on an interest area made during first-pass reading (i.e., before any movements are made to a later interest area, or before moving back to earlier text). Skipping rate is the probability that an item was not fixated during first-pass reading (a binary variable: 1 for fixated, 0 for skipped). Total fixation duration was defined as the sum of all fixation times made on an interest area.

Analyses and results

Regression techniques were used to investigate differential effects of word frequency and transitional probability on various reading measures in the two age groups. We analyzed the data by fitting linear mixed-effects models in R version 3.1.1. (R Development Core Team, 2014), using the lmer-function from the lme4 package (Bates, Maechler, Bolker and colleagues, 2014). Numerical predictors were centered around their mean to reduce collinearity due to scaling of variables (i.e., non-essential collinearity, cf. Wurm & Fisicaro, 2014). Categorical variables (age group, reading condition, verb type) were contrast-coded. All regression models contained participant number and noun-verb combination as random effects, that is, we included random intercepts and random slopes for subjects and items. We adopted a design-driven approach of maximal random effects for hypothesis testing and generalizability advocated by Barr and colleagues (2013). Hence, each regression model included by-subject random slopes for frequency and transitional probability and by-item random slopes for age group and reading condition, unless convergence was not reached. If the fully specified random structure led to convergence errors due to perfect correlations between predictor slopes (1.00 or -1.00 indicate overfitting of a model), we removed the respective slope(s). Any simplifications of the random structure are reported along with the models' estimates. P-values for the predictors in each model were obtained by conducting likelihood-ratio tests, comparing nested models that differed only in the absence or presence of the fixed effect under investigation while keeping all other fixed and random effects constant. Frequency of the verb and forward transitional probability were correlated (i.e., $r = 0.59$ for logFrequency of the verb and logForwardTP), which required some caution when adding both to the same statistical model. Since we were interested in age-related changes of the effect of predictability on reading behavior after controlling for differential effects of word frequency, we did not remove collinearity between frequency and TP by using techniques of residualisation (for a discussion see York, 2012; Wurm & Fisicaro, 2014). Instead, we always fitted an initial model that solely investigated the effect of frequency (and its interaction with age group), and only afterwards added either ForwardTP or BackwardTP and their interactions with age group to the same model (with TP added both to the fixed and random parts). If the effect of any predictor changed due to the addition of the TP variable, we report it below. Improvement of model fit was compared by means of likelihood-ratio tests between the models including and excluding TP and by evaluations of the respective AICs (Akaike Information Criterion).

Results

Summaries of average gaze durations and total fixation durations for younger versus older adults in silent reading versus reading aloud are provided in Table 2. As the analyses reported below will confirm, older adults had larger reading condition differences than younger adult in their initial gaze behavior and in their total gaze behavior. The difference in gaze duration between silent reading and reading aloud is 83 milliseconds for older adults, and 63 milliseconds for younger adults. The reading duration difference for oral versus silent reading amounts to 135 milliseconds for older adults, and 71 milliseconds for younger adults, when looking at total fixation durations. As a “late” reading measure, total fixation duration includes regressive fixations on a target area, which may be important to capture changes in the reading process in older adulthood. Collapsed over reading conditions, older adults have a slightly higher skipping rate (7%) during first-pass reading of the verbs than younger adults (6%), but older adults regress more often to the initially-skipped verbs during later reading (60% for older adults versus 36% for younger adults), and generally have a higher verb refixation rate (43%

for older adults versus 25% for the younger adults).

In our statistical models, we tested for interactions of frequency, age group, and reading condition on gaze durations, skipping rates, and total fixation durations, as well as for interactions of ForwardTP, age group and reading condition. None of these three-way interactions indicated that frequency or TP effects differed in size for silent reading versus reading aloud in younger or older adults, and these interactions were therefore deleted from the models. Word length of the verb was a significant predictor in all models, such that that longer verbs had longer fixation durations and were less likely to be skipped. Verb type (infinitive or past participle) had no overall effect.

A summary of the lmer-model fitted for GD in the verb data is provided in Table 3. The model yielded significant simple effects of reading condition and ForwardTP, such that participants' gaze durations were longer in oral compared to silent reading and longer when the verb was less predictable, as compared to more predictable verbs. Furthermore, the model showed a significant interaction between age group and reading condition, indicating that reading condition differences were larger for older than younger adults. Note, though, that the age group variable had no significant simple effect on GDs (see also Table 2).

When ForwardTP was excluded from this model (yielding a frequency-only model), the frequency effect became significant, $\beta = -7.3555$, SE = 2.3074, $t = -3.19$, $p < 0.05$), such that high-frequency verbs had shorter gaze durations than low-frequency verbs. The frequency effect was of similar size across groups (i.e., no frequency by group interaction). Model comparisons by means of likelihood-ratio tests and comparisons of AICs between the two models spoke in favor of the more complex model in which TP was included, suggesting that TP explains more variance than verb frequency alone. We also set up a model in which TP was the only probabilistic variable (leaving verb frequency out), which confirmed the simple effect of predictability, and yielded no predictability by age group interaction. Thus, frequency and TP effects were similar in size in both groups.

To complement the first-pass fixation duration analysis, we considered whether or not verbs were skipped during first-pass reading. 850 verbs were skipped during first-pass reading (N = 447 for older adults, N = 403 for younger adults) while 12519 trials had one or more fixations on the verb during first-pass reading. We used the generalized linear mixed-effects function *glmer*() for binomial data (1 = fixated, 0 = skipped during first-pass reading) to fit skipping rate models. A summary of the model is provided in Table 4. Skipping was influenced by verb frequency, verb predictability, and reading condition. Specifically, frequent and predictable verbs were skipped more often than infrequent or unpredictable verbs, and verbs were more likely to be skipped in silent reading than in reading aloud. There was also a marginally significant effect of age group, that is, older adults were more likely to skip a target verb than younger adults. Older adults showed smaller frequency effects on skipping than younger adults, and older adults had smaller reading condition differences than younger adults (cf. Table 2). These two interaction effects were confirmed in a frequency-only model (excluding TP). Again, we also set up a model in which TP was the only probabilistic variable (leaving verb frequency out), which confirmed the simple effect of predictability, and yielded no predictability by age group interaction. In sum, TP effects were similar in size in both group, but frequency effects on skipping rates differed for younger versus older adults.

The last analysis reported for the verb region is the analysis of total fixation duration (TD). Before fitting the TD model, we excluded eight outliers that had total fixation durations above 2000 ms. A summary of the best model for TD is provided in Table 5. Total fixation durations were influenced by verb frequency, ForwardTP, age group, and reading condition. Thus, frequent words had shorter fixations than infrequent verbs, predictable verbs had shorter total fixation durations than less predictable verbs, older adults spent more time reading the verbs than younger adults and total fixation durations were longer in reading aloud, compared to silent reading. The model also showed a significant interaction between verb frequency and age group, in that frequency effects were larger in older than in younger adults. Note that this interaction is in the opposite direction compared to the skipping rate analysis. The interaction was confirmed in a frequency-only model in which the TP variable was left out. Predictability effects were equal in both groups, which was confirmed in the respective TP-only model (leaving verb frequency out). Finally, older adults had larger reading condition differences than younger adults such that older adults' fixation durations were slowed more by oral reading than those of younger adults. Again, the TD data is in contrast to our findings from the skipping rate analysis (in which younger adults had larger condition differences). The interaction of age group and condition is also apparent in the descriptive statistics in Table 2: the difference in average TD between reading aloud and silent reading is 71 milliseconds for younger adults, but 135 milliseconds for older adults.

In sum, our results show that older adults had overall larger reading condition differences in their fixation durations than younger adults. Moreover, predictability effects were equal in size in younger and older adults, while frequency effects differed between the two groups, specifically in total fixation durations, see Figure 1 (top right panel). If we apply a median-split to the continuous verb frequency variable, the difference in total fixation duration between high-frequency and low-frequency items was 66 milliseconds for older adults, versus 49 milliseconds for younger adults. Note that frequency and predictability effects were not modulated by reading task, neither in the entire sample nor in any of the groups individually.

Table 2. Average gaze durations and average total fixation durations for the target verbs split by age group and reading condition (in milliseconds; standard deviations provided in brackets); skipping rates are provided in percentages.

	Older adults		**Younger adults**	
	Reading aloud	Silent reading	Reading aloud	Silent reading
Gaze duration (ms)	355 (174)	272 (114)	340 (148)	275 (125)
Total duration (ms)	521 (266)	386 (230)	414 (199)	343 (192)
Skipping rate (%)	6.51	7.39	2.93	8.54

Table 3. Summary of linear mixed-effects model fitted for gaze durations (GD) of the verb region (N=12519 trials), including β-coefficients, standard errors, *t*-values and levels of significance; the model converged with random intercepts for subjects and items, as well as by-subject random slopes for verb frequency and ForwardTP.

Variable	**β**	**SE**	***t***	***p***
WordLengthVerb	11.130	1.404	7.93	***
VerbType	0.004	4.157	0.00	
Condition – Reading aloud	77.427	2.347	32.99	***
logFreqVerb	-3.598	2.482	-1.45	
logForwardTP	-3.918	1.053	-3.72	**
Group – Older adults	5.231	11.877	0.44	
logFreqVerb by Group	-5.096	2.830	-1.80	
logForwardTP by Group	1.350	1.301	1.04	
Condition by Group	18.106	4.738	3.82	**

*** $p < 0.001$, ** $p < 0.01$

Table 4. Summary of generalized linear mixed-effects model fitted for skipping rate (SR) of the verb region (N = 13369 trials), including β-coefficients, standard errors, z-scores and levels of significance; the model converged with random intercepts for items and subjects.

Variable	**β**	**SE**	***z***	***p***
WordLengthVerb	0.368	0.035	10.57	***
VerbType	0.051	0.087	0.59	
Condition – Reading aloud	0.653	0.079	8.24	***
logFreqVerb	-0.143	0.046	-3.12	*
logForwardTP	-0.070	0.021	-3.39	*
Group – Older Adults	-0.433	0.198	-2.19	m.s.
logFreqVerb by Group	0.197	0.074	2.65	*
logForwardTP by Group	-0.072	0.038	-1.89	
Condition by Group	-1.156	0.159	-7.26	***

*** $p < 0.001$, * $p < 0.05$, m.s. $p < 0.1$

Table 5. Summary of linear mixed-effects model fitted for total fixation duration (TD) of the verb region (N=12922 trials), including β-coefficients, standard errors, t-values and levels of significance; the model included random intercepts for subjects and items, by-subject random slopes for frequency and ForwardTP, and by-item random slopes for reading condition.

Variable	**β**	**SE**	***t***	***p***
WordLengthVerb	18.073	2.323	7.78	***
VerbType	2.211	6.871	0.32	
Condition – Reading aloud	104.353	4.024	25.93	***
logFreqVerb	-10.465	4.051	-2.58	*
logForwardTP	-6.522	1.711	-3.81	**
Group – Older adults	79.013	22.729	3.48	**
logFreqVerb by Group	-14.323	4.106	-3.49	**
logForwardTP by Group	2.785	1.883	1.48	
Condition by Group	61.581	6.989	8.81	***

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

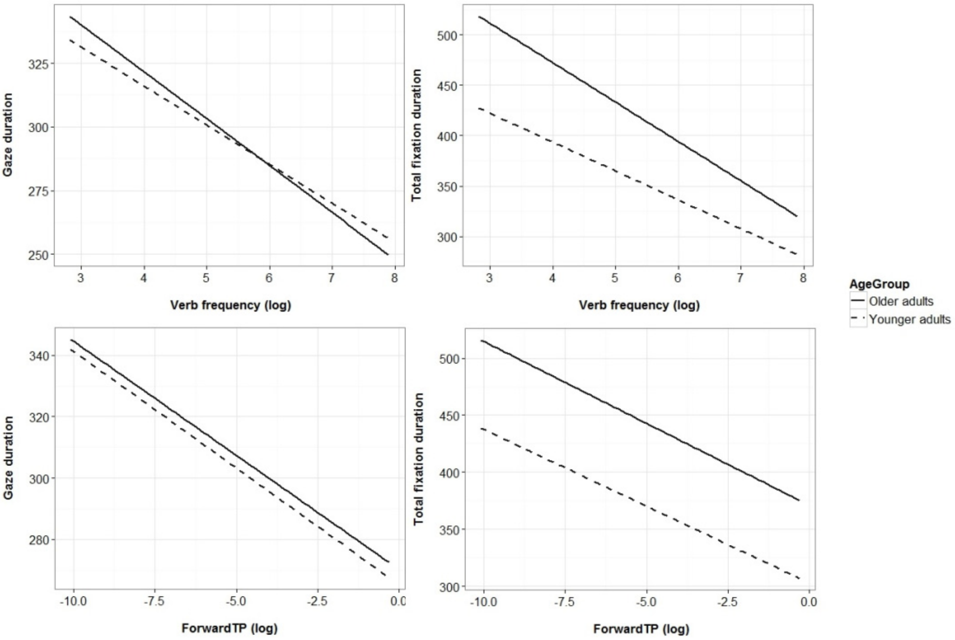

Figure 1. Linear regression plots for the critical verb region aggregated across reading aloud and silent reading; graphs are split by younger adults (dashed lines) and older adults (solid lines) for the effect of verb frequency on gaze durations (top left) and total fixation durations (top right), and the effect of forward TP on gaze durations (bottom left) and total fixation durations (bottom right). Differences between older and younger adults in frequency or predictability effects were not significant for the gaze durations (cf. top left and bottom left graph), but frequency effects did differ significantly for total fixation durations (top right graph).

Discussion

We investigated, first, whether and how younger and older readers differ in the use of local probabilistic cues, that is, word frequency and transitional probability statistics, and, second, whether any age-related differences in the size of frequency and predictability effects would be more pronounced in reading aloud than in silent reading. Reading aloud imposes additional processing demands (e.g., Reynolds & Besner, 2006) and may therefore leave less room for specific reading strategies that have been associated with older adults' silent reading (Rayner et al., 2006; for a discussion see Gordon et al., 2016). Regarding our first research question, both older and younger adults showed word frequency and transitional probability effects on fixation durations and skipping rates. While predictability effects were similar in size in younger and older adults, frequency effects differed across age groups. Regarding our second research question, frequency and predictability effects turned out to be similar in reading aloud and silent reading. This similarity was observed across age groups, but also for each age group individually. The most salient differences between silent reading and reading aloud were observed in increased total fixation durations in reading aloud, particularly for the older adults, and in a differential skipping pattern across age groups and reading tasks.

Age-related differences in frequency effects

Frequency effects in total fixation durations (but not in initial gaze durations) were larger in older adults than in younger adults. It is not surprising that group differences in frequency effects are most pronounced in total fixation duration given that older adults regress to and refixate critical verbs more often. Older adults may particularly regress to low-frequency verbs, which increases the total fixation duration on those words. Consequently, the difference in fixation duration for low- versus high-frequency verbs increases more in older adults, compared to younger adults (who regress less often). Increased frequency effects with older age are in line with findings of studies by Kliegl and colleagues (2004, 2006), as well as Rayner and colleagues (2006), which investigated silent reading. Our study extends these findings by showing that frequency effects increase with age, regardless of whether the task is silent reading, which solely involves comprehension, or reading aloud, which also involves speech production and is therefore more demanding. Increased frequency effects with older age and more language experience may be explained by larger vocabulary size as a mediating factor. Older adults typically have larger vocabularies than younger adults (e.g., Ben-David et al., 2016), which was also the case in our sample. Larger vocabulary entails knowing more low-frequency words. Consequently, there is more competition in the lexical activation of low-frequency words, which enhances the difference in processing time between low- and high-frequency words (e.g., Revill & Spieler, 2012; Spieler & Balota, 2000). Another account of increased frequency effects with age relates to the reliance on lexical information during word recognition, which increases with age to compensate for age-related declines in visual acuity and character recognition (cf. Gordon et al., 2016).

Note that our results differ from those obtained by Kliegl and colleagues (2004) and by Rayner and colleagues (2006), in that we found smaller frequency effects in skipping rates with older age, rather than larger frequency effects. The opposite effects in total fixation durations versus first-pass skipping rates may be related to age group differences in reading behavior. Although both younger and older readers skip words during first-pass reading, the older adults later regress to these words more often than the younger adults. Particularly infrequent words are likely to receive regressive fixations, such that any age-related increase in frequency effects may be better captured by late reading measures (like total fixation duration) rather than early measures.

Predictability effects

In previous studies that investigated age by predictability interactions, predictability was generally operationalized by cloze norms. On the basis of earlier results (cf. Cheimariou, 2016;

Kliegl et al., 2004, and Rayner et al., 2006), it was not clear whether older and younger adults would differ in their use of TP, as it was unclear whether TPs would pattern with age patterns for more global predictive cues (cloze effects) or with age patterns for frequency effects. In line with the silent reading studies by Kliegl and colleagues (2004), as well as Rayner and colleagues (2006), who found similar predictability effects in older and younger adults, we found equal predictability effects across age groups. Hence, our results do not only replicate findings from silent reading, but extend these to reading aloud and to local TP effects.

However, our results differ from those obtained by DeLong and colleagues (2012), and Federmeier and colleagues (e.g., Federmeier, McLennan, De Ochoa, & Kutas, 2002), who found smaller predictability effects in older adults compared to younger adults in a number of language comprehension studies (for a review see Federmeier, 2007). Importantly, decreased contextual effects for older, compared to younger, adults were not only observed when the contextual cues were global and had to be integrated over longer distances (e.g., Federmeier & Kutas, 2005), but also in the use of a local contextual cue (cf. DeLong and colleagues (2012) on article-noun combinations such as “a bicycle” versus “an elephant”). Besides the basic differences in experimental paradigms and measures (i.e., direct, neural measures in EEG studies versus indirect, behavioral measure in the current eye-tracking study), there are two possible ways to account for this. First, our study used eye-tracking during normal reading, whereas DeLong and colleagues used EEG and presented the words at a fixed word-by-word presentation rate, which put readers under considerable time pressure. By contrast, the natural reading task in our study did not involve any time pressure. Timing, alongside with cognitive capacity, has been shown to be a crucial factor for making predictions during language processing (especially for prediction of phonological form), such that even high-functioning younger adults may not engage in predictive processing if time pressure is high (Ito et al., 2016; Linderholm 2002; Wlotko & Federmeier, 2015). In our study, older adults may have had enough time for involving prediction in their processing, leading to equal predictability effects in both age groups. Second, our local context cue differed from the contextual information in DeLong's study in that our noun-verb combinations, varying in transitional probability, were surrounded by otherwise neutral contexts, while DeLong and colleagues embedded their ”a”/ ”an” combinations in lead-in sentences, which rendered the determiner-noun combinations more or less predictable. Thus, age may interact with the type of information that readers preferentially process and integrate. The longer the distance over which information needs to be integrated, the more older adults will be disadvantaged in their rapid use of contextual information because of age-related declines in processing speed or working memory. This may be why frequency effects dissociate from TP effects and cloze effects in that older adults rely more on single-word information than younger adults, but the benefit from probabilistic information between words is offset by integration cost with older age. Further research is needed to systematically differentiate age-related changes in longer distance versus local integration types of predictability.

Reading task differences

Considering general differences in silent reading versus reading aloud across the two age groups, eye gaze data showed age-related slowing in both reading tasks. Older adults needed longer times to read the verbs as reflected by longer total fixation durations compared to younger adults. Again, we replicated results from Kliegl and colleagues (2004), and Rayner and colleagues (2006), who reported slowed sentence reading in older adults compared to younger adults in silent reading. Importantly, our study extends these results and shows that age-related slowing is more pronounced in reading aloud, as reflected by differentially longer total fixation durations in older adults' oral reading compared to younger adults' oral reading (cf. Table 2). The slowing effect in oral reading, particularly in older adults, may be explained by assuming a tight synchronization between production and comprehension processes. As reported in our companion paper (Moers et al., 2015), older adults spoke more slowly than younger adults in reading aloud. This was reflected in slower speech rates (syllables per second) when reading entire sentences (older adults were 12% slower in their speech rate than younger adults). Simultaneously, older adults were 16% slower than younger adults in total sentence reading times in the eye-gaze data, which points to roughly equal amounts of slowing in the comprehension and production data. We would like to acknowledge that these findings of age-related slowing in reading and articulation do not necessarily mean that older adults are generally slower in their language processing than younger adults. Another reason may be that older adults process linguistic information more thoroughly and/or have more complex lexical and semantic processing relative to younger adults, due to more diverse and complex linguistic systems with language experience and age. Moreover, older adults may simply choose to articulate more carefully when reading out loud.

Apart from slowing in the eye gaze duration data, a differential pattern of skipping behavior occurred across age groups and reading tasks. On the basis of earlier studies that associated older adults' silent reading with a more 'risky' reading strategy (cf. Rayner et al., 2006), we hypothesized that the additional production level in oral reading would leave less room for risky eye-movement patterns due to a tight synchronization between eye gazes and the rather linear process of articulation. Consequently, less skipping was expected in oral compared to silent reading, particularly for the older adults. In silent reading both younger and older adults had similar verb skipping rates (cf. Table 2). The predicted pattern of reduced skipping in oral reading was observed for the younger adults, but not for the older adults (i.e., older adults' skipping rate remained relatively high in reading aloud). This finding of similar skipping rates in oral and silent reading of older adults may relate to the differential slowing of oral reading in the older adults. Possibly, due to longer reading times, older adults fixated words long enough to benefit relatively more from parafoveal preview of the verbs than younger adults (but see e.g., Payne & Stine-Morrow, 2012, for a discussion on parafoveal preview effects in silent reading). Consequently, older adults may have lexically processed the verbs sufficiently to not need to make another fixation right on the verb. According to this explanation, the relatively high skipping rates in oral reading of older adults would not indicate a risky reading strategy, but would be a consequence of age-related slowing in reading rate. More research is needed to confirm reading task-specific accounts.

In sum, our study confirms that older adults show different reading patterns compared to younger adults in that there was age-related slowing in fixation durations, there were changes in skipping and regression rates, as well as changes in frequency effects. Older adults' silent reading has been linked to a 'risky' reading strategy in previous studies because of relatively high skipping rates during first-pass reading that were accompanied by high regression rates during later reading stages. Our data confirmed the latter result of higher regression and refixation rates with age, and found that older adults skipped more words than younger adults in oral reading. Word skipping is generally a "good" feature in eye gaze behavior in that it enables readers to read at fast rates (hence proficient readers may often skip words), but skipping may become inefficient if accompanied by many regressions during later reading. Gordon and colleagues (2016) reviewed several studies on age-related changes in the reading process, and concluded that higher rates of skipping may result from a compensation mechanism for slower visual and motor processes. By adopting a risky reading strategy, older adults rely more heavily on their intact semantic and conceptual representations and less heavily on perceptual processing of text than younger adults (cf. Gordon et al., 2016). This reading pattern may be an automatic adaptation or compensatory mechanism in response to cognitive and/or physiological limitations in older age, rather than being a consciously applied strategy (cf. Gordon et al., 2016).

Conclusion and Outlook

This study investigated probabilistic language processing in silent reading and reading aloud in younger and older adults. We expected age groups to differ particularly in the size of frequency and predictability effects in reading aloud (more so than in silent reading), because reading aloud leaves less room for a risky reading strategy due to the required synchronization of visual decoding, word recognition, and articulation. Older adults make differentially longer fixations in reading aloud, which may be attributed mainly to constraints on reading speed imposed by speech production. However, frequency and predictability effects in reading aloud did not differ from those in silent reading. Hence, even the more constraining task of reading aloud does not change the use of basic language features such as probabilistic lexical information. This is in line with the conclusion of Gordon and colleagues (2016), who argue that although older adults read more slowly than young adults, their patterns of eye movements in response to lexical characteristics (e.g., frequency) and sentence characteristics (e.g., word predictability) strongly resemble those of younger adults, demonstrating the influence of preserved crystalized intelligence in the form of language knowledge. We used transitional probabilities to estimate predictability, instead of sentence cloze values, because TPs represent local probabilistic knowledge that may reflect an easy-to-use predictability clue and does not require long-distance integration of information.

Future research should further investigate how age, cognitive resources and different measures of predictability (e.g., TPs, multi-word chunks, cloze values) influence the cost of integration into the sentential context during reading. Moreover, more research is needed to differentiate subjective predictability measures that reflect individual language experience (such as subjective ratings of frequency or of co-occurrence) from corpus-based measures like TPs, and evaluate which type of measure is a better predictor for predictive processing in younger versus older adults.

Appendix A – Analysis of the noun region

As before, we tested for the effects of frequency, TP, age group and reading condition (and their interactions) on fixation durations and skipping rates. The noun-region analysis confirmed findings from the verb data. First, older adults skipped nouns more often (23.21%) than younger adults (18.56%) during first-pass reading, but older adults also regressed to those nouns more often (41.97%) than younger adults (22.09%), and had a higher noun refixation rate in general (42.17% versus 27.14%). Second, predictability effects did not differ in younger versus older adults, nor across reading modalities, which again confirms the findings from the verb data.

Table 6. Average gaze durations and average total fixation durations for the target nouns split by age group and reading condition (in milliseconds; standard deviations provided in brackets); skipping rates are provided in percentages.

	Older adults		**Younger adults**	
	Reading aloud	Silent reading	Reading aloud	Silent reading
Gaze duration (ms)	291 (152)	237 (102)	309 (148)	238 (103)
Total duration (ms)	456 (266)	336 (205)	389 (206)	304 (187)
Skipping rate (%)	23.39	23.05	16.37	20.67

Table 7. Summaries of mixed-effect models (including β-coefficients, standard errors, *t*-values or *z*-scores, and levels of significance) for skipping rates (SR; N = 13363 trials) and total fixation durations (TD; N = 11491 trials) in the noun region.

	SR				**TD**			
Variable	**β**	**SE**	***z***	***p***	**β**	**SE**	***t***	***p***
WordLengthNoun	0.382	0.020	18.97	***	21.296	1.700	12.52	***
Condition – Reading aloud	0.157	0.046	3.43	**	105.295	4.438	23.73	***
logFreqNoun	-0.054	0.034	-1.61		-2.344	3.350	-0.70	
logBackwardTP	-0.012	0.020	-0.58		-4.907	1.873	-2.62	**
Group – Older adults	-0.318	0.120	-2.65	**	51.479	20.786	2.48	*
logFreqNoun by Group	-0.024	0.044	-0.56		-15.618	4.021	-3.88	**
logBackwardTP by Group	0.025	0.029	0.85		2.240	2.369	0.95	
Condition by Group	-0.336	0.092	-3.65	***	41.706	7.448	5.60	**

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

References

Aghababian, V., & Nazir, T. A. (2000). Developing normal reading skills: aspects of the visual processes underlying word recognition. *Journal of Experimental Child Psychology*, 76(2), 123-150. doi: 10.1006/jecp.1999.2540

Allen, P. A., Smith, A. F., Lien, M.-C., Grabbe, J., & Murphy, M. D. (2005). Evidence for an Activation Locus of the Word-Frequency Effect in Lexical Decision. *Journal of Experimental Psychology: Human Perception and Performance*, 31(4), 713-721. doi: 10.1037/0096-1523.31.4.713

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language,* 68(3), 255-278. doi: http://dx.doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., & Bolker, B. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. Retrieved from http://lme4.r-forge.r-project.org/

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1), 92-111. doi: http://dx.doi.org/10.1016/j.jml.2008.06.003

Ben-David, B. M., Erel, H., Goy, H., & Schneider, B. A. (2015). “Older is always better”: Age-related differences in vocabulary scores across 16 years. *Psychology and Aging*, 30(4), 856-862. doi: 10.1037/pag0000051

Blythe, H.I., & Joseph, H.S.S.L. (2011). Children’s Eye Movements during Reading. In S.P. Liversedge, I. Gilchrist & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 643-662). Oxford: Oxford University Press.

Burke, D.M. & Shafto, M.A. (2008). Language and aging. In F.I.M. Craik & T.A. Salthouse (Eds.), *The handbook of aging and cognition* (pp.373-443). New York: Psychology Press.

Cheimariou, S. (2016). *Prediction in aging language processing*. Doctoral dissertation, University of Iowa. Retrieved from http://ir.uiowa.edu/etd/3056/

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, 1084(1), 89-103. doi: 10.1016/j.brainres.2006.02.010

Davies, R. A., Arnell, R., Birchenough, J. M., Grimmond, D., & Houlson, S. (2017). Reading through the life span: Individual differences in psycholinguistic effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 43(8):1298-1338. doi: 10.1037/xlm0000366

DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, 121(3), 226-239. doi: 10.1016/j.bandl.2012.02.006

Dunn M. & Dunn, LM. (2005). Peabody Picture Vocabulary Test-III-NL (Nederlandse versie Liesbeth Schlichting).

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491-505. doi: 10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, 42(2), 133-141. doi: 10.1111/j.1469-8986.2005.00274.x

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115(3), 149-161. doi: http://dx.doi.org/10.1016/j.bandl.2010.07.006

Federmeier, K.D., McLennan, D.B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, 39, 133–146. doi: 10.1017.S004857720139203X

Fernández, G., Shalom, D. E., Kliegl, R., & Sigman, M. (2014). Eye movements during reading proverbs and regular sentences: the incoming word predictability effect. *Language, Cognition and Neuroscience*, 29(3), 260-273. doi: 10.1080/01690965.2012.760745

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of Contextual Predictability and Transitional Probability on Eye Movements During Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 31(5), 862-877. doi: 10.1037/0278-7393.31.5.862

Gordon, P. C., Lowder, M. W., & Hoedemaker, R. S. (2016). Reading in normally aging adults. In H. Wright (Ed*.), Cognition, Language and Aging* (pp. 165-192). Amsterdam: Benjamins.

Hahn, L. W. (2012). Measuring local context as context-word probabilities. *Behavior Research Methods*, 44(2), 344-360. doi: 10.3758/s13428-011-0148-y

Hand, C.J., Miellet, S., O'Donnell, P.J. & Sereno, S.C. (2010). The Frequency-Predictability Interaction in Reading: It Depends Where Your're Coming From. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1294-1313. doi: 10.1037/a0020363

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, 115(5), 1090-1103. doi: http://dx.doi.org/10.1016/j.clinph.2003.12.020

Hyönä, J., & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6), 1430-1440. doi: http://dx.doi.org/10.1037/0278-7393.21.6.1430

Inhoff, A. W., Solomon, M., Radach, R., & Seymour, B. A. (2011). Temporal dynamics of the eye–voice span and eye movement control during oral reading. *Journal of Cognitive Psychology*, 23(5), 543-558. doi: 10.1080/20445911.2011.546782

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, 86, 157-171. doi: http://dx.doi.org/10.1016/j.jml.2015.10.007

Janse, E., & Jesse, A. (2014). Working memory affects older adults' use of context in spoken-word recognition. *Quarterly Journal of Experimental Psychology*, 67(9), 1842-1862. doi: 10.1080/17470218.2013.879391

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 824-843. doi: 10.1037/0278-7393.20.4.824

Jordan, T. R., McGowan, V. A., & Paterson, K. B. (2014). Reading with filtered fixations: adult age differences in the effectiveness of low-level properties of text within central vision. *Psychology and Aging*, 29(2), 229-235. doi: 10.1037/a0035948

Kemper, S., & Liu, C. J. (2007). Eye movements of young and older adults during reading. *Psychology and Aging*, 22(1), 84-93. doi: 10.1037/0882-7974.22.1.84

Keuleers, E., Brysbaert, M., & New, B. SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, 42(3), 643-650. doi: 10.3758/brm.42.3.643

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16, 262-284. doi: 10.1080/09541440340000213

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, 135(1), 12-35. doi: 10.1037/0096-3445.135.1.12

Kuperman, V., Dambacher, M., Nuthmann, A., & Kliegl, R. (2010). The effect of word position on eye-movements in sentence and paragraph reading. *The Quarterly Journal of Experimental Psychology*, 63(9), 1838-1857. doi: 10.1080/17470211003602412

Laubrock, J., & Kliegl, R. (2015). The eye-voice span during reading aloud. *Frontiers in Psychology*, 6. doi: 10.3389/fpsyg.2015.01432

Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and causal text constraints. *Discourse Processes*, 34, 259–280. doi: 10.1006/jmla.1994.1008

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology,* 88, 22-60. doi: http://dx.doi.org/10.1016/j.cogpsych.2016.06.002

McDonald, S. A., & Shillcock, R. C. (2003a). Eye Movements Reveal the On-Line Computation of Lexical Probabilities During Reading. *Psychological Science*, 14(6), 648-652. doi: 10.1046/j.0956-7976.2003.psci_1480.x

McDonald, S. A., & Shillcock, R. C. (2003b). Low-level predictive inference in reading: the influence of transitional probabilities on eye movements. *Vision Research*, 43(16), 1735-1751. doi: http://dx.doi.org/10.1016/S0042-6989(03)00237-2

Moers, C., Janse, E., & Meyer, A. S. (2015). Probabilistic reduction in reading aloud: A comparison of younger and older adults. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.), *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*. London: International Phonetics Association.

Mund, I., Bell, R., & Buchner, A. (2010). Age differences in reading with distraction: sensory or inhibitory deficits? *Psychology and Aging*, 25(4), 886-897. doi: 10.1037/a0019508

Navarrete, E., Basagni, B., Alario, F. X., & Costa, A. (2006). Does word frequency affect lexical selection in speech production? *The Quarterly Journal of Experimental Psychology*, 59(10), 1681-1690. doi: 10.1080/17470210600750558

Ong, J. K. Y. & Kliegl, R. (2008). Conditional co-occurrence probability acts like frequency in predicting fixation durations. *Journal of Eye Movement Research*, 2(1):3, 1-7.

Payne, B. R., & Stine-Morrow, E. A. (2012). Aging, parafoveal preview, and semantic integration in sentence processing: testing the cognitive workload of wrap-up. *Psychology and Aging*, 27(3), 638-649. doi: 10.1037/a0026540

R Development Core Team (2014). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org

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

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422. doi: 10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye Movements in Reading: Models and Data. *Journal of eye movement research*, 2(5), 1-10.

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, 21(3), 448-465. doi: 10.1037/0882-7974.21.3.448

Revill, K. P., & Spieler, D. H. (2012). The effect of lexical frequency on spoken word recognition in young and older listeners. *Psychology and Aging, 27*(1), 80-87. doi:10.1037/a0024113

Reynolds, M., & Besner, D. (2006). Reading aloud is not automatic: processing capacity is required to generate a phonological code from print. *Journal of Experimental Psychology: Human Perception and Performance,* 32(6), 1303-1323. doi: 10.1037/0096-1523.32.6.1303

Robert, C., Mathey, S., & Postal, V. (2009). Différences liées à l'âge dans la reconnaissance visuelle des mots chez l'adulte. *European Review of Applied Psychology*, 59(2), 139-151. doi: http://dx.doi.org/10.1016/j.erap.2008.08.001

Schotter, E. R., Bicknell, K., Howard, I., Levy, R., & Rayner, K. (2014). Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading. *Cognition*, 131(1), 1-27. doi: http://dx.doi.org/10.1016/j.cognition.2013.11.018

Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *Neuroreport*, 9(10), 2195-2200.

Shaoul, C., Harald Baayen, R., & Westbury, C. F. (2014). N-gram probability effects in a cloze task. *The mental lexicon,* 9(3), 437–472. doi: 10.1075/ml.9.3.04sha

Smith, N. J., & Levy, R. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pp. 1637–1642.

Spieler, D. H., & Balota, D. A. (2000). Factors influencing word naming in younger and older adults. *Psychology and Aging, 15*(2), 225-231. doi:10.1037/0882-7974.15.2.225

Staub, A. (2015). The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation. *Language and Linguistics Compass,* 9(8), 311-327. doi: 10.1111/lnc3.12151

Stine-Morrow, E. A. L., Soederberg Miller, L. M., & Hertzog, C. (2006). Aging and Self-Regulated Language Processing. *Psychological Bulletin*, 132(4), 582-606. doi: 10.1037/0033-2909.132.4.582

Valle, A., Binder, K. S., Walsh, C. B., Nemier, C., & Bangs, K. E. (2013). Eye Movements, Prosody, and Word Frequency Among Average- and High-Skilled Second-Grade Readers. *School Psychology Review*, 42(2), 171-190.

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging,* 18(2), 332-339. doi:10.1037/0882-7974.18.2.332

Vorstius, C., Radach, R., & Lonigan, C. J. (2014). Eye movements in developing readers: A comparison of silent and oral sentence reading. *Visual Cognition*, 22(3-4), 458-485. doi: 10.1080/13506285.2014.881445

Wang, H.-C., Pomplun, M., Chen, M., Ko, H., & Rayner, K. (2010). Estimating the effect of word predictability on eye movements in Chinese reading using latent semantic analysis and transitional probability. *The Quarterly Journal of Experimental Psychology*, 63(7), 1374-1386. doi: 10.1080/17470210903380814

Whitford, V., & Titone, D. (2013). The effects of reading comprehension and launch site on frequency–predictability interactions during paragraph reading. *The Quarterly Journal of Experimental Psychology,* 67(6), 1151-1165. doi: 10.1080/17470218.2013.848216

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, 68, 20-32. doi: http://dx.doi.org/10.1016/j.cortex.2015.03.014

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, 72(0), 37-48. doi: http://dx.doi.org/10.1016/j.jml.2013.12.003

York, R. (2012). Residualization is not the answer: Rethinking how to address multicollinearity. *Social Science Research*, 41(6), 1379-1386. doi: http://dx.doi.org/10.1016/j.ssresearch.2012.05.014

CHAPTER 5

CONTEXTUALLY-CONSTRAINED SPEECH PRODUCTION: PREDICTABILITY EFFECTS IN YOUNGER AND OLDER ADULTS

This study was conducted at the Department of Linguistics of Northwestern University, in collaboration with Prof. dr. Matthew Goldrick.

Abstract

Language processing often entails making predictions about words that are likely to occur. Predictability effects may undergo age-related changes as they depend on language experience and on rapid processing of context information to generate predictions. We investigated age-related differences in sentence-level and word-pair prediction. Participants read sentences in a self-paced fashion that were either constraining or neutral towards a final noun, which was presented as a picture to be named aloud. In the constraining condition, pictures were preceded by adjectives that formed a high or low transitional probability combination with the noun. Both cloze predictability and transitional probability influenced picture naming latencies, but not word durations. Cloze and TP effects were similar in younger and older adults. In sum, under conditions that do not impose strict time constraints, as was the case in our picture naming task, older adults do not differ from students in their processing of predictability information.

Contextually-Constrained Speech Production: Predictability Effects in Younger and Older Adults

Introduction

Readers, listeners, and speakers use prediction to facilitate language processing. Evidence for the role of predictability in language processing comes from studies in which predictable words are processed faster and more efficiently than unpredictable words. Language users build up their predictions on different types of information, such as world knowledge, semantic and grammatical constraints, and word (combination) frequencies. This information is by definition dependent on language experience. Therefore, predictive language processing may change with age. A number of studies have demonstrated that predictability effects may decrease in older adults compared to students, but note that these studies did not distinguish different types of predictive information, such as global semantic and local probabilistic information. In this study, we contrasted two types of predictability cues: sentence-level cloze predictability and transitional probabilities between word pairs.

Predictability effects in language processing

During language processing, people can anticipate likely upcoming materials. Adequate predictions make language processing fast and efficient as they allow the pre-processing of words which facilitates their integration once they are actually encountered. Using a predictive processing strategy can thereby facilitate language tasks, for example in conversations where listeners may anticipate the interlocutor’s words in order to efficiently process and integrate them into their own utterance planning (for a discussion see e.g., Gambi & Pickering, 2011). A large number of experimental studies have shown that predictable words receive processing benefits compared to unpredictable words. For example, predictable words have shorter fixation durations and are skipped more often during reading, they are processed faster during speech comprehension, and they can be named faster in speech production relative to unpredictable words (e.g., Frisson et al., 2005; Griffin & Bock, 1998; McDonald & Shillcock, 2003a; van Berkum et al., 2005). Word predictability is also associated with changes in pronunciation, such that more predictable words are acoustically reduced relative to less predictable ones, which has been termed *Probabilistic Reduction* (Bell et al., 2003).

Types of predictability information

Several types of linguistic information have been shown to reflect predictability from context, such as general discourse information (van Berkum et al., 2005), functional and general word associations (Hintz, 2015), syntactic structures (Arai & Keller, 2012; Staub & Clifton, 2006), linguistic givenness and second mention (Kahn & Arnold, 2012), case marking (Kamide, Scheepers, & Altmann, 2003), prosody (Weber, Grice, & Crocker, 2006), hesitations (Shriberg & Stolcke, 1996), and even fine-grained phonological and orthographic information (DeLong, Urbach, & Kutas, 2005; Laszlo & Federmeier, 2009). Of course, different types of information can be processed and used in parallel to derive predictions.

Two measures of predictability that are commonly used in studies on predictability effects are Cloze Probability and Transitional Probability (TP). In the cloze procedure, participants have to complete sentence fragments on the basis of what they think is likely to follow. As such, cloze probabilities are subjective measures of how likely words are to occur given some sentence context and may depend on the participant group. Cloze predictability influences, for instance, eye fixations in reading (e.g., Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006), response latencies in speech production (e.g., Griffin & Bock, 1998), and the effort of semantic integration in speech comprehension (as reflected by the N400 component in EEG studies, e.g., Federmeier, 2007). The second measure, TP, is derived from word and word pair frequencies of large language corpora. The transitional probability reflects how likely a word is to occur given the right or left neighboring word. People are sensitive to these local dependencies between words, as evidenced by the findings that readers fixate on high-TP words (those words that are more predictable from their local context) for shorter periods of time than on low-TP words (cf. McDonald & Shillcock, 2003a, 2003b; Frisson, Rayner, & Pickering, 2005; Wang, Pomplun, Chen, Ko, & Rayner, 2010; but see Ong & Kliegl, 2008) and speakers acoustically reduce high-TP words more than low-TP words (e.g., Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Jurafsky, Bell, Gregory, & Raymond, 2001; Moers, Janse, & Meyer, 2015).

Age-related changes in predictions

Over the course of their lives people acquire knowledge about the contexts in which certain words occur and this knowledge supports making appropriate expectations (e.g., Federmeier, 2007). Linking predictive processing to language experience, one might assume that older people are better at making predictions than younger people, simply because older adults have more experience with the contexts and combinations in which certain words occur. Indeed, older adults have been shown to benefit more from context relative to younger adults (e.g., Pichora-Fuller, Schneider, & Daneman, 1995; Sheldon, Pichora-Fuller, & Schneider, 2008), particularly in speech-in-noise tasks, in which the reliance on top-down lexical knowledge is encouraged for (post-perceptual) guessing. However, Cheimariou (2016) and Rayner and colleagues (2006) reported equal predictability effects for older and younger adults in silent reading. In EEG studies by DeLong and colleagues, and Federmeier and colleagues, older adults engaged less in anticipation of upcoming words than younger adults during speech comprehension and text processing (e.g., DeLong, Groppe, Urbach & Kutas, 2012; Federmeier, Kutas, & Schul, 2010; Federmeier, McLennan, De Ochoa, & Kutas, 2002). In sum, it is not clear whether and how the ability to use predictive processing changes across the adult life span.

Importantly, almost all studies investigating age group differences have operationalized predictability using sentence completion tasks. The cloze measure, if measured on entire sentences with a final-word gap, captures higher-level predictions and semantic integration effects over longer stretches of preceding context (for a discussion see Staub, Grant, Astheimer, & Cohen, 2015). Experimental research suggests that people need more time for semantic integration as they age (e.g., Huang, Meyer, & Federmeier, 2012; Payne & Stine-Morrow, 2012). Moreover, the rapid use and integration of preceding semantic sentence context has been linked to working memory capacity (Huettig & Janse, 2016; Janse & Jesse, 2014; Linderholm, 2002), which declines with age (e.g., van den Noort, Haverkort, Bosch, & Hugdahl, 2006). Any age effects on higher processing levels (such as conceptual and semantic levels), may not necessarily generalize to how age groups process more local predictability cues. The longer the distance over which information needs to be integrated, the more older adults may be disadvantaged in their use of contextual information. TP effects are more local in nature, as they capture the dependency of only two words. Age-related changes may be less pronounced in TP effects, as TP predictability requires fewer processing demands and cognitive resources relative to cloze predictability. Linking predictability effects to working memory and age-related limits in cognitive resources, we hypothesized that older adults and younger adults differ in their processing of global predictive cues, but not so much in their processing of local word-to-word information. Thus, age may interact with the type of information that readers readily process and integrate.

Another factor that may affect the size of predictability effects in younger versus older adults is vocabulary knowledge. Vocabulary knowledge is influenced by literacy and the amount of exposure to language in diverse contexts. Recent studies suggest that larger vocabulary size is associated with larger predictability effects, compared to smaller vocabulary knowledge. For instance, Rommers and colleagues (2015), as well as Hintz (2015), observed that vocabulary size influenced the anticipatory eye gaze towards target objects in a visual world paradigm in student populations, such that participants with larger vocabulary had larger predictability effects, compared to lower-vocabulary participants. Furthermore, Borovsky and colleagues (2012) reported that children (3 to 10 years old) with higher vocabulary scores were faster to anticipate target words relative to children with lower vocabulary scores. Moreover, Huettig and colleagues (2014; Mishra, Singh, Pandey, & Huettig, 2012) linked literacy and the amount of exposure to texts to the size of predictability effects by arguing that enhanced literacy increases the speed of lexical access (particularly the rapid availability of orthographic and phonemic representations) and thus enables the rapid availability of lexical representations during predictive speech processing. Increasing language experience is associated with an increase in vocabulary knowledge, such that older adults typically have larger vocabularies than younger adults (Ben-David, Erel, Goy, & Schneider, 2015; Hartshorne & Germine, 2015; Verhaeghen, 2003). Therefore, we hypothesized that the group of older adults in our study would have higher vocabulary scores than the student group. Moreover, we assumed that vocabulary knowledge would be a mediating factor in predictability effects, and we assessed its contribution to age group differences in predictability effects. Federmeier and colleagues (2002) showed that older adults with larger receptive vocabulary scores were more likely to show the young adult's pattern of responses (ERPs) in predictive processing than older adults with smaller vocabularies. They concluded that the use of predictive processing strategies in older adults is related to their ability to quickly access lexical items and assign meaning to a wide range of words.

The current study

In order to investigate age-related changes in the use of predictability information, we asked younger and older adults to participate in a picture naming task, in which we manipulated the lexical information that was available from previous sentence context. In this paradigm, the convergence of written context information and pictorial information is in the speech production, hence we are going to report the analyses of picture naming latencies and word durations. We had three research questions: First, we investigated whether cloze predictability and transitional probabilities both affect experimentally controlled speech production in the paradigm we used. Second, we aimed to test our hypothesis that younger and older adults would process transitional probabilities similarly, but that the two groups would differ in their benefit from cloze predictability as sentence-level predictability imposes more processing demands and requires longer-distance integration relative to the use of local, co-occurrence information. Third, we investigated whether cloze and TP predictability effects on speech production are modulated by vocabulary size, as vocabulary size reflects the rapid availability of lexical representations which may influence the speed with which predictions can be derived.

Methods

We designed a picture-naming experiment in order to investigate how cloze predictability and transitional probability affect speech production in younger versus older adults. On each trial a participant read a sentence of which the last word (a noun) was provided as a picture that had to be named aloud. Sentence presentation was word-by-word and self-paced. The sentence context was either neutral, or highly constraining such that the last word could be predicted (as cloze manipulation). Moreover, sentences included specific adjectives that formed either a high or low forward transitional probability combination together with the depicted noun (as TP manipulation). Similar paradigms that involved both sentence reading and picture naming have been successfully used by other researchers in order to investigate speech production in semantically constraining contexts (e.g., Ferreira & Griffin, 2003; Gollan, Slattery, Goldenberg,Van Assche, Duyck, & Rayner, 2011; Kleinmann, Runnqvist, & Ferreira, 2015; Piai, Roelofs, Rommers, & Maris, 2015; Severens, Ratinckx, Ferreira, & Hartsuiker, 2008; Staub, Grant, Astheimer, & Cohen, 2015).

Participants

Twenty-four younger adults (aged 18 to 33 years, Mean = 20.5, SD = 3.39) and twenty-four older adults (aged 63 to 78 years, Mean = 71.9, SD = 4.45) participated in the experiment in exchange for course credits or money ($10 per hour). Participants were recruited via Northwestern University's Linguistic Department participant pool or by advertising our experiment on flyers and in email newsletters. All participants were tested in an experimentation lab of the Linguistics Department at Northwestern University. The study was approved by the Institutional Review Board of Northwestern University (as part of the project STU00077121) in accordance with ethical principles underlying research with human subjects. All participants were native speakers of American English and reported no history of language impairments or neurological problems. All participants reported normal or corrected-to-normal vision.

Materials

We selected 48 mono- and disyllabic nouns (36 monosyllabic, e.g., "house"), each of which was paired with two adjectives: a high transitional-probability adjective (e.g., "haunted house") and a low transitional-probability adjective (e.g., "shabby house"). Forward transitional probabilities (FTP) for noun-verb combinations (high FTP: min = 0.007, max 0.6; low FTP: min = 0.00005, max 0.01) were calculated from the 440-million-words *Corpus of Contemporary American English* (COCA; Davies, 2008). For each noun a full-color photograph (picture on white background) was selected from either the BOSS picture database (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010) or from the internet using Google Images. Pictures matched both the high and low TP adjective, which was assessed by a separate norming study (see Appendix A for a description of the norming procedure). For each noun two sentences were constructed: one from which the noun could be predicted as sentence-final word (e.g., "The suburban real estate agent could not sell the…" > haunted / shabby house), and a neutral sentence from which the noun could not be predicted, but would still be plausible (e.g., "It was hard for Liz to recognize the…" > house). Neutral sentences did not include an adjective (and hence no high/low TP manipulation). Picture naming agreement (above 70% for each picture) and predictability of target sentences (above 0.7 for high cloze predictability sentences, below 0.3 for the neutral sentences) were normed in separate surveys (cf. Appendix A).

The target materials were distributed over twelve different lists. Each list included a total of 196 trials: 48 target sentences with high cloze predictability and a high TP adjective, 48 target sentences with high cloze predictability and a low TP adjective, 48 neutral target sentences (no adjective), and additionally 48 neutral filler sentences. The filler sentences were added in order to match the rate of neutral versus highly predictive sentences across all trials in the experiment. The fillers were created by pairing a new set of 16 filler pictures with the neutral target sentences (i.e., each of the 16 pictures was added to 3 out of 48 neutral sentences). Each experiment list consisted of three blocks. That means that during the experiment, each participant saw each target picture three times, once in the high cloze probability – high TP condition, once in the high cloze probability – low TP condition and once in the neutral condition (plus three times a filler picture). Each block contained the same number of items of each condition. The order of the blocks in which the items occurred was randomized across lists according to a fully crossed Latin-square design. The twelve lists were randomly assigned to participants (two younger and two older participants per list).

Procedure

At the start of the testing session, participants first received information about the nature of the study and gave written consent. Then, they filled in a standard questionnaire about their language and education background. Participants were then asked to take a seat in a sound-attenuated booth, in which the picture-naming experiment was presented on a 15 inch computer screen. Speech responses were recorded using a head-mounted microphone. Participants saw five practice trials to familiarize themselves with the task. They were asked to read each sentence silently, and to name the sentence-final picture aloud. Sentences were presented word-by-word in the center of the screen with black letters on a white background. Presentation rate was self-paced. That is, every trial started with a fixation cross and participants had to tap on a trackpad each time they wanted to see the next word. After the sentence had been read up to and including the penultimate word, the picture occurred in a centered, eight by six inch frame for 600 milliseconds. Participants received the instructions to say the name of the picture as fast as they could using a single word. The trackpad was blocked during picture presentation for a total time of 1500 milliseconds, such that participants could not accidentally skip the picture naming period. An illustration of the timing and setup within trials is provided in Figure 1.

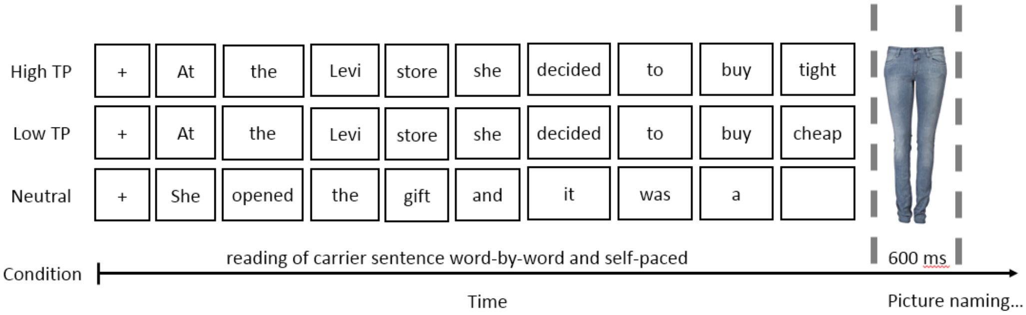

Figure 1. Illustration of the time course within an experimental trial, split by condition (top = sentence with a high-TP adjective-noun combination; middle = sentence with a low-TP adjective-noun combination; bottom = neutral sentence).

As noted above, the 196 trials of a list were presented in three blocks. Participants were encouraged to take breaks between blocks. After finishing the picture-naming experiment, participants were asked to perform two additional tasks: the MINT task (Multilingual Naming Test; Gollan, Weissberger, Runnqvist, Montoya, & Cera, 2012), measuring productive vocabulary, and the Shipley questionnaire (*Shipley-2*, 2009), measuring receptive vocabulary. The MINT consisted of a set of 68 black and white line drawings (e.g., of a seesaw) that were presented one by one on a computer screen and had to be named aloud in English. If a participant did not recognize the picture, a semantic cue was provided by the experimenter (e.g., "found on a playground"). The participant's score was the number of correctly named items. Age group differences in productive vocabulary were small but significant (M = 67.00, SD =

1.74 for older adults; M = 65.79, SD = 1.67 for younger adults; t(45.91) = 2.45, p = .02). The Shipley vocabulary task consisted of 40 multiple-choice questions in which synonyms of varying difficulty have to be identified (e.g., “mollify” > mitigate, direct, pertain, abuse). The survey was presented on a computer screen and participants clicked on one of four response options. Each participant’s score was the number of correctly identified synonyms (Shipley raw score). There were no significant age group differences in receptive vocabulary size (M = 35.67, SD = 1.74 for older adults; M = 34.42, SD = 2.59 for younger adults; t(40.23) = 1.97, p = .06), but note the same trend as in productive vocabulary scores. The entire testing session took about 90 minutes for younger adults and up to two hours for older adults.

Data selection

Picture onsets were automatically marked in a second auditory channel co-registered with the speech recordings. Speech onsets and offsets in the recordings from the picture-naming experiment were automatically measured by an acoustic trigger system (using intensity thresholds). In a second step, the automatic annotations were hand-corrected in Praat (Boersma & Weenink, 2012) by two trained transcribers. About half of the participants were processed by myself, and the other half by a student assistant from the Max Planck Institute for Psycholinguistics. Both transcribers processed data of younger and older adults, such that any age group differences were not confounded with transcriber differences. Naming latencies were measured from onset of picture presentation to speech onset. Word durations were calculated from speech onset to speech offset. During annotation the transcribers could not see in which predictability condition the pictures had been named and thus were not influenced in their choices regarding the word onsets and offsets by the experiment design. Incorrect picture names, disfluent and self-corrected responses, as well as responses with more than one word were excluded from the analysis. To keep the number of observations equal across all conditions, those responses were removed from the data in all three conditions (i.e., removed from all three blocks an item occurred in). Out of a total of 6912 trials in the experiment (48 participants times 144 target trials) 1686 were removed from the analysis, with an effective drop-out rate of 24% (21% for the younger adults; 28% for the older adults). Most errors were made in picture responses that occurred after neutral sentences (i.e., non-predictable responses; 340 trials), followed by those in the low TP condition (246 trials), and the fewest errors were made in the high TP condition (197 trials). In all conditions, older adults made more errors than younger adults (error rates for older adults: 18% neutral, 13% low TP, 9% high TP; error rates for younger adults: 12% neutral, 8% low TP, 8% high TP).

Analysis and Results

Statistical techniques

We analyzed reaction times (RTs) and word durations (WDs) from our experiment by fitting linear mixed-effects models in R version 3.1.1. (R Development Core Team, 2014), using the lmer-function from the lme4 package (Bates, Maechler, Bolker and colleagues, 2014). Age group was entered as a contrast-coded factor (binary contrast coded as -0.5 for younger adults and 0.5 for older adults). The three-level factor "predictability" (highTP, lowTP, neutral condition) was also contrasted-coded, such that the contrasts allowed for two comparisons: First, we compared the effect of high sentence predictability to the neutral sentence condition (contrast named *Cloze vs. Neutral* coded as 0.25 for highCloze-highTP items, 0.25 for highCloze-lowTP items, -0.5 for neutral items). Second, the effect of highTP predictability could be compared to lowTP predictability (contrast named *HighTP vs. LowTP* coded as 0.5 for highCloze-highTP items, -0.5 for highCloze-lowTP items, 0 for neutral items).

Block, age group, and predictability were entered as simple effects in each model. Moreover, the two-way interaction of predictability and age group was entered to test whether predictability affected the age groups differently. We also added an interaction of predictability and block as predictability effects may become weaker across blocks. All regression models contained subjects and items as random effects. We further added by-subject random slopes for predictability and by-item random slopes for age group and predictability in order to adhere to a design-driven approach of maximal random effects for hypothesis testing and generalizability advocated by Barr and colleagues (2013). Outliers of more than three standard deviations above and below the mean RT or mean WD were removed before analyzing the data. RTs and WDs were log-transformed in all mixed models reported below.

Results

Descriptive statistics for RTs and WDs split by age group and experimental condition are provided in Table 1. Reaction times and word durations were longer in older adults compared to younger adults across all conditions. RTs decreased across predictability conditions in both younger and older adults, such that picture naming latencies were fastest in the high TP condition and slowest in the neutral sentence reading condition. The differences in WDs across experimental conditions were very small in both age groups. The observation of condition differences in RTs, but not WDs, was also confirmed in the mixed-models reported below.

Table 1. Descriptive statistics (with mean, standard deviation (SD), minimum (Min) and maximum (Max)) for reaction times and word durations in milliseconds split by age group and predictability condition.

	Younger adults				**Older adults**			
	Reaction times							
Condition	Mean	SD	Min	Max	Mean	SD	Min	Max
High TP	477	150	19	1161	589	149	257	1289
Low TP	530	166	239	1182	649	173	320	1293
Neutral	669	138	414	1283	796	167	441	1296
	Word durations							
Condition	Mean	SD	Min	Max	Mean	SD	Min	Max
High TP	442	123	106	882	515	132	173	966
Low TP	436	121	126	928	509	126	149	963
Neutral	439	121	161	947	513	127	207	947

A summary of the full lmer-model fitted for RTs is provided in Table 2 (left panel). The model yielded significant simple effects of block, age group, cloze predictability and TP predictability. Hence, reaction times decreased across blocks (i.e., the second or third time a picture is named), older adults were slower than younger adults in their response times, pictures in predictable contexts were named faster than pictures in neutral contexts, and pictures following high TP adjectives were named faster than pictures following low TP adjectives. Both age groups benefitted from both cloze and TP predictability to a similar extent, as indexed by the non-significant interactions of age groups and predictability contrasts. The only significant interaction in the model was the two-way interaction of block and the TP manipulation, which indicated that the effect of high versus low TP became smaller across blocks.

As the strongest predictability effects were observed in the first block (cf. the TP-by-Block interaction in Table 2), we fitted an additional model in which we solely entered data from the first block (leaving block and block-interactions with other variables out; summarized in Table 2, right panel). The first-block-only model confirmed our most important findings: older adults responded more slowly than younger adults, predictable sentences generated faster responses than neutral sentences, and highTP items were named faster than lowTP items. Importantly, predictability benefits were similar in size for younger and older adults.

Table 2. Summary of mixed-effects models (including β-coefficients, standard errors, *t*-values, and levels of significance) fitted for reaction times; data from all three blocks (N = 5153 trials) on the left side; data from the first block only (N = 1700 trials) on the right side; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

	All three blocks				**First-block-only**			
Predictor	**β**	**SE**	***t***	***p***	**β**	**SE**	***t***	***p***
Block	-0.039	0.004	-10.90	***				
AgeGroup	0.221	0.036	6.11	***	0.199	0.038	5.18	***
Cloze vs. Neutral	-0.429	0.032	-13.23	***	-0.414	0.027	-15.15	***
HighTP vs. LowTP	-0.167	0.021	-7.83	***	-0.134	0.019	-7.18	***
Cloze vs. Neutral x Block	0.013	0.010	1.32					
HighTP vs. LowTP x Block	0.032	0.009	3.72	*				
Cloze vs. Neutral x AgeGroup	0.054	0.039	1.39		0.019	0.045	0.42	
HighTP vs. LowTP x AgeGroup	0.012	0.019	0.65		0.006	0.031	0.19	

Turning to the analysis of WD data, we provide a summary of the full model in Table 3 (left panel). The model yielded a significant simple effect of age group, showing that older adults had longer word durations than younger adults. None of the other variables or interactions between them contributed significantly to word durations. Note that even very simple models (not reported here), in which interactions of predictability with other variables were removed, did not yield significant predictability effects on word durations. As with the RT data, we also fitted a first-block-only model, which confirmed the full model in that it yielded a significant simple effect of age group and no other significant effects.

Table 3. Summary of mixed-effects models (including β-coefficients, standard errors, *t*-values, and levels of significance) fitted for word durations; data from all three blocks (N = 5198 trials) on the left side; data from the first block only (N = 1737 trials) on the right side; ** p < 0.05.

	All three blocks				**First-block-only**			
Predictor	**β**	**SE**	***t***	***p***	**β**	**SE**	***t***	***p***
Block	-0.003	0.003	-1.23					
AgeGroup	0.175	0.045	3.90	**	0.169	0.046	3.66	**
Cloze vs. Neutral	-0.007	0.016	-0.45		-0.008	0.011	-0.68	
HighTP vs. LowTP	-0.008	0.014	-0.62		-0.011	0.011	-1.02	
Cloze vs. Neutral x Block	0.002	0.007	0.25					
HighTP vs. LowTP x Block	0.010	0.006	1.55					
Cloze vs. Neutral x AgeGroup	-0.003	0.013	-0.24		-0.004	0.022	-0.16	
HighTP vs. LowTP x AgeGroup	-0.002	0.011	-0.16		-0.006	0.021	-0.31	

Given that we found effects of almost all variables in the RT models, but not in the word duration models, we tested whether there was a relationship between RT and WD per trial, such that trials with relatively long response latencies also had relatively long word durations. We entered RT as a predictor to the word duration model, which yielded no significant simple effect of RT on WDs. Hence, in this analysis response speed seemed to not predict articulation duration. We further investigated whether there was a by-participant trade-off between response speed (RT) and speech articulation (WD). Some participants may start giving their response before they are done with planning the entire response. Consequently, effects of ease of word planning due to contextual constraint may spill over to articulation (and hence to word durations) for participants who started speaking quickly (cf. Buz & Jaeger, 2016; Kello, 2004). To test for this relationship, we extracted the random by-participant intercepts from the RT model (reported in Table 2) as an index of individual response speed, and included these as simple effect as well as in interaction with the predictability effect in the word duration model. A negative correlation between the predictability effect and the by-participant RT intercept would reflect that speakers who start responding relatively quickly will show larger predictability effects on word duration. The model yielded a significant effect of RT intercept on word duration ($\beta = 0.4088$, $SE = 0.178902$, $t = 2.29$), but no interaction of the intercept with either TP predictability or cloze predictability was observed. Hence, those participants who had longer response latencies also had longer word durations, but across participants the facilitation from sentence or TP predictability is solely found in the planning of words (reflected by naming latency), but not in the subsequent articulation.

In an additional analysis, we added the word-by-word reading rate of the carrier sentences to the mixed-effects models of the RTs and WDs as we hypothesized that the time participants spend on reading a sentence may interact with the predictability effect on word naming. That is, those participants who take more time to read the sentences may show larger predictability effects due to more attentive processing of context. Reading rate in words per second was calculated using the time stamps from the trackpad tapping during the self-paced reading of the sentence up to (but not including) the presentation of the final noun. Older adults had slower reading rates than younger adults ($M = 1.97$ words/sec, $SD = 1.00$ for older adults; $M = 2.81$ words/sec, $SD = 1.07$ for younger adults). We observed that the confound between age group and reading rate increased the estimates of interactions with age group, which is why we report the additional analysis without including age group in the modelling. Reading rate was log-transformed and centered before entering the variable to the RT and WD models, for which summaries are provided in Table 4. Reading rate was entered as simple effect and in interaction with the predictability contrasts. We also added by-participant and by-item random slopes for reading rate. As before, the WD model yielded no significant effects. However, in the RT model, presented in Table 4, reading rate interacted with cloze predictability such that faster sentence reading led to smaller predictability effects on picture naming latency (without a significant simple effect of reading rate). This was confirmed in analyses on subsample datasets of younger and older adults, such that across age groups participants who took more time to read the context sentences had larger predictability effects. The RT model on data of both age groups (Table 4) also showed interactions between TP and cloze predictability and block, with the two types of predictability effects decreasing in later blocks.

Table 4. Summary of mixed-effects models (including β-coefficients, standard errors, *t*-values, and levels of significance) including sentence reading rate fitted for reaction times (left panel, N = 5153 trials) and word durations (right panel, N = 5198 trials); *** p < 0.001, ** p < 0.01, * p < 0.05.

	RTs				**WDs**			
Predictor	**β**	**SE**	***t***	***p***	**β**	**SE**	***t***	***p***
Block	-0.050	0.005	-10.69	***	-0.006	0.003	-1.76	
LogReadingRate	0.016	0.025	0.63		0.008	0.018	0.46	
Cloze vs. Neutral	-0.352	0.035	-10.13	***	-0.006	0.017	-0.36	
HighTP vs. LowTP	-0.161	0.024	-6.81	***	-0.005	0.015	-0.36	
Cloze vs. Neutral x Block	-0.024	0.012	-2.00	*	0.002	0.008	0.22	
HighTP vs. LowTP x Block	0.029	0.010	2.92	*	0.008	0.007	1.14	
Cloze vs. Neutral x LogReadingRate	0.144	0.026	5.53	***	-0.002	0.015	-0.16	
HighTP vs. LowTP x LogReadingRate	0.017	0.020	0.86		0.010	0.013	0.72	

Lastly, we tested whether predictability effects on picture naming related to vocabulary size differences. We fitted models on the RT and WD data in which we entered vocabulary scores (either Shipley or Mint) as simple effects as well as in interaction with the predictability contrasts. Neither Mint nor Shipley scores contributed to RTs or WD, nor yielded any significant interaction effects. Thus, predictability effects (particularly observed for RT) were not modified by an individual's vocabulary knowledge. Note though, that vocabulary scores reached ceiling-levels in both tasks. Hence, there were only subtle between-participant differences in the scores, which makes it hard to assess the contribution of vocabulary to predictability effects per se.

In sum, we found evidence for block, age group, cloze sentence and TP effects on picture naming latencies. TP and cloze predictability effects were similar in size in younger and older adults. Moreover, reading rate interacted with cloze predictability, such that faster reading times led to smaller predictability effects on naming latencies. The word duration data reflected some age-related slowing in speech production, but there were no signs of predictability-related acoustic reduction. Vocabulary size made no significant contribution to any model.

Discussion

In the present study we investigated age-related differences in predictability effects in a picture-naming task. Participants read context sentences and provided the sentence-final noun by naming a picture. Specific age-related changes, particularly regarding the types of information used in predictive processing, have not been studied in detail so far. We manipulated cloze predictability at sentence level by providing contexts that were either constraining or neutral towards a final noun. Importantly, we also investigated local word-pair predictability such that high or low transitional probability adjectives preceded the picture (noun) presentation. We had three research questions: 1) whether cloze predictability and transitional probabilities both affect experimentally controlled speech production in the paradigm we used; 2) whether younger and older adults differ in their benefit from cloze predictability versus TP predictability, as the two types of predictability may differ in their processing demands; 3) whether cloze and TP predictability effects on speech production are modulated by vocabulary size, as vocabulary size reflects the rapid availability of lexical representations which may influence the speed with which predictions can be set up.

Regarding the first research question, we found evidence that both cloze predictability and TP predictability affected speech production in the task we used. Note, though, that predictability effects were found only in reaction times, but not in the word durations. Therefore, the facilitation from sentence or TP predictability is solely found in the planning of words, not in the subsequent realization. Generally, all variables that were entered into the statistical modes (block, age group, cloze predictability, TP predictability, and reading rate) affected picture naming latencies, while the only effect seen in the analyses of the word duration being that of age-related slowing. Considering our second research question, contrary to our hypothesis we saw that younger and older adults have similar predictability effects, both for local TP contexts and for cloze sentence context. Longer-distance integration of context in this task does not affect older adults more negatively relative to younger adults. With respect to the third research question, we found no contribution of vocabulary size to predictability effects in younger versus older adults. The scores from both vocabulary tasks (productive and the receptive vocabulary) showed very little variation among individuals and across age groups, which made it difficult to detect any potential contribution of vocabulary knowledge to predictability effects in the first place. We will now discuss our findings in more detail.

Effects in reaction times but not word durations

Given that predictability effects, and in particular transitional probabilities, have been linked to probabilistic reduction (i.e., the acoustic reduction of words that are likely to occur in their context; Jurafsky et al., 2001), it may be surprising that such effects were absent in our word duration data, despite clear effects of cloze probability and transitional probability on reaction times. Several studies reported that high-TP words are acoustically more reduced than low-TP words (e.g., Bell et al., 2003; Bell et al., 2009; Gregory et al., 1999). Note, that most of these studies were corpus-studies that investigated spontaneous, conversational speech and, importantly, that excluded phrase-final or sentence-final words which are likely to undergo phrase-final lengthening (e.g., Beckman & Edwards, 1990; Fougeron & Keating, 1997). In our study, all target words were sentence-final and thus may have undergone lengthening, such that predictability-related acoustic reduction effects were minimized (or not statistically detectable). Nevertheless, the null-effect in word durations may be surprising given that previous studies used similar paradigms, in which constraining sentence context was provided before a sentence-final picture had to be named aloud, and which found predictability effects in articulation measures (e.g., Drake & Corley, 2015). However, in those studies the experimental paradigm typically involved a semantically incongruent condition, in which the picture continuation was implausible given the sentence context information. That is, lexical access may have been disrupted for incongruent items, which may have increased the difference in articulation for plausible versus implausible items. In our study, we contrasted highly-constraining versus neutral sentence context, as well as high-TP versus low-TP items, which were plausible and semantically congruent in all cases. Thus, any effects on the acoustic realization of the target words may have been very small in our study, as compared to studies that accentuated the contrast in experimental conditions. Also note that Drake and Corley (2015) presented their context-constraining sentences auditorily, such that presentation rate was not controlled by the participant, which may have increased spill-over effects into articulation.

Age-related changes in predictability effects

In previous studies that investigated age-related changes in cloze predictability effects groups of older adults typically had smaller predictability effects compared to groups of younger adults (Federmeier & Kutas, 2005; for a review see e.g., Federmeier, 2007; see also Huang, Meyer, & Federmeier, 2012, for age-related differences in the processing of adjective-noun combinations). The results from our study differ from these earlier findings since our statistical analysis yielded equal predictability effects for younger and older adults. This was the case for both predictability measures we used. Note that it was not clear whether TP effects would pattern with age patterns for more global predictive cues (cloze effects) as age-related changes in TP effects had not been studied. Importantly, decreased contextual effects for older, compared to younger, adults were not only observed in cloze effects but also in a study that investigated age-group differences in local contextual predictability. DeLong and colleagues (2012) used article-noun combinations, such as “a bicycle” versus “an elephant”, in which the noun was differentially predictable from the article by the English article rule (“a” for words with initial consonant, “an” for words with initial vowel). Older adults differed in the processing of this local cue from younger adults such that older adults had smaller predictability effects. We can think of two explanations to account for these differences across the studies.

The first explanation relates to the task-set used in our paradigm. The involvement of speech production - additional to the reading of the carrier sentence - may have particularly encouraged older adults to engage in predictive processing. Gollan and colleagues (2011) contrasted strongly and weakly constraining context in a picture naming and a reading task. In both tasks predictable items were processed faster than less predictable items, but the predictability effect was much larger in the naming task than in the reading task. More evidence for enhanced predictability effects in tasks involving speech production comes from Hintz and colleagues (2016), who found larger facilitation effects in mixed-task blocks, where participants also had to name aloud the outcome of a predictive stimulus, relative to comprehension-only trials. One explanation for these findings is that participants are more engaged (aroused) and/or motivated if the task actively requires language production as compared to comprehension-only tasks. Relatedly, the mixed task set involved in our paradigm, in which participants first had to read a carrier sentence and then overtly produce the sentence-final word, may have encouraged particularly older adults to engage in predictive processing more so than in standard comprehension studies (e.g., Federmeier & Kutas, 2005). This task engagement account relates to recent prediction-by-production models, stressing the role of the speech production system during predictive processing (e.g., Dell & Chang, 2014; Pickering & Garrod, 2013; for a discussion see Huettig, 2015). The idea behind these models is that predicting a word may be closely related to producing the word internally (i.e., covert production is involved during language comprehension). Consequently, the question arises whether predictability effects differ in language tasks that involve solely comprehension from tasks that involve both comprehension and overt production, as was the case in our experiment.

The second explanation is connected to stimulus presentation timing and the timing of semantic integration. DeLong and colleagues (and in fact, many other studies investigating age-related changes in predictability effects) used an EEG paradigm in which the context sentences were presented at a fixed word-by-word presentation rate, which put readers under considerable time pressure. By contrast, the sentences in our study were presented word-by-word in a self-paced manner such that participants could take as much time as they needed to process the context sentences. Stimulus timing has been shown to be a crucial factor when drawing on predictive processing strategies during language processing (especially for prediction of fine-grained phonological form, cf. Ito, Corley, Pickering, Martin-Nieuwland, & Nieuwland, 2016). If time-pressure in the experimental task is high, even high-functioning younger adults may not engage in predictive processing (Ito et al., 2016; Linderholm 2002; Wlotko & Federmeier, 2015). In our study, however, older adults may have had enough time for involving prediction in their processing due to the self-paced task, leading to equal predictability effects in both age groups. Support for the "time-pressure" explanation comes from studies that used eye-tracking during sentence reading. Rayner and colleagues (2006), as well as Kliegl and colleagues (2004), investigated cloze predictability effects in reading and found no difference between younger and older adults when whole sentences were presented (rather than individual words) and participants could view them as long as they needed to process them.

Additional support for this explanation comes from the statistical analysis that involved sentence reading rate: reading rate of the context sentence interacted with cloze predictability, such that slower reading times led to larger predictability effects on naming latencies in both younger and older adults. One interpretation of this finding would be that those participants who had slower reading rates, read the sentences more attentively and thus processed the predictability information more intensely. Given that older adults generally read sentences more slowly than younger adults, any age-related decline in predictability effects may have been washed out by the self-paced reading. In sum, our study did not (artificially) impose/increase age-related differences in predictive processing, as we allowed enough time for the processing of predictability cues, which may lead to similar predictability effects between younger and older adults.

Conclusions and Outlook

This study investigated age-related changes in predictability effects in a contextually-constrained picture naming paradigm. The study confirmed that both local word-pair predictability and sentence-global cloze predictability affected picture naming latencies. Younger and older adults, however, did not differ in their processing of these predictability cues, although we expected older adults to benefit less from the availability of cloze predictability information than younger adults due to increased processing demands (i.e., the need for longer-distance semantic integration in cloze predictability, relative to TP predictability). We attribute the similarity of predictive processing across the age groups to the self-paced nature of the task. In other words, if the experimental paradigm allows older adults enough time to process predictive information, age group differences in prediction during language processing may be minimal. Future research should further investigate how age, cognitive resources and different measures of predictability (e.g., general world knowledge, lexical associations, phrasal predictability in multi-word chunks) influence the cost of integration into the sentential context during language processing. Moreover, more research is needed to assess the dependency of age-related changes in predictability effects on the language task (comprehension only versus comprehension and production, dialog) and to evaluate how stimulus-presentation timing influences younger versus older adults.

Appendix A – Description of the norming procedure for pictures and sentences

Pictures were normed for their naming agreement and adjective-acceptability. Sentences were normed for their cloze predictability, as assessed by a sentence completion task. We designed the norming surveys with the help of the Qualtrics software (Qualtrics, Provo, UT, 2014) and utilized Amazon Mechanical Turk (MTurk, 2014) to run them online. Each survey was taken by 11 participants, most of which were middle-aged adults. The data of 4 participants was discarded later because these participants did not follow the instructions or had high error rates. None of the participants participated in more than one survey. Potential participants could preview their task on MTurk. Upon accepting the assignment, they gave written consent. After completing a survey (which took about 15 minutes) participants received \$3. Participants saw three practice trials to familiarize themselves with the task before continuing to the target trials. In the first and second survey (the picture-norming surveys) participants saw pictures one by one (e.g., a baby) and were asked to type in a single word to name the picture. Subsequently, while still seeing the picture, they were asked to rate on a 5-point scale whether they would use a target adjective-noun combination (high TP or low TP) to describe the picture ("Would you use the words 'newborn baby' to describe the following picture" > "Definitely (1), probably (2), maybe (3), probably not (4), never (5)"). The first and second survey only differed in the adjective-noun part, in terms of the condition in which each item was presented (as high versus low TP adjective-noun combination), which was counterbalanced (i.e., both surveys had the same amount of high TP and low TP adjective-noun combinations, but highTP items from the first survey were included as lowTP items in the second survey, and vice versa). We also interspersed one quarter of nonsense adjective-noun fillers into the surveys to probe for "never"-responses (e.g., "Would you use the words 'silver fries' to describe the following picture?", with fries depicted as normal, yellow fries). All target adjective-noun combinations had a mean score of at least 3.4 (i.e., tended towards the "maybe"-rating). Picture naming agreement was above 70 percent for all pictures that were selected in the experiment materials. The third survey (the sentence-norming survey) tested for the cloze predictability of highly constraining and neutral target sentences. On each trial participants were randomly presented with either a neutral or a constraining sentence of which the last word was missing and had to be provided by the participant (e.g., "She does not want to spend much money on the…"). Participants were instructed to type in a single word. Sentence completion agreement (cloze predictability) was above 70 percent for highly constraining sentences, and below 30 percent for neutral sentences.

In a post-experiment norming phase, we modified the Qualtrics surveys to obtain scores of picture naming agreement from 9 older adults, as well as sentence cloze predictability norms from 9 older adults (aged above 60). For older adults the surveys were part of an in-lab experiment session (rather than online). None of the older participants had participated in the actual experiment. Picture naming agreement was above 70% per item, except for five out of 48 pictures (these had a naming agreement above 50%). Cloze predictability was above 0.7 for highly constraining sentences, except for eight out of 48 sentences (in which cloze predictability ranged from 0.22 to 0.66 in the older adults). Re-analyses of our data, in which the five items with lower naming agreement or the eight items with lower cloze scores were excluded, yielded the exact same results as reported in the Results section.

Acknowledgements

We would like to thank Chun Liang Chang for technical support, Erin Gustafson for her help with participant recruitment, as well as Hanna de Groot and Jessica Aguilar Diaz for their help with data coding.

References

Amazon Mechanical Turk (MTurk), 2014, Amazon Web Services. Available from: www.mturk.com

Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes,* 28(4), 525-560. doi: 10.1080/01690965.2012.658072

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278. doi: http://dx.doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., & Bolker, B. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. Retrieved from http://lme4.r-forge.r-project.org/

Beckman, M. E., & Edwards, J. (1990). Lengthening and shortenings and the nature of prosodic constituency. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I* (pp. 152-178). Cambridge: Cambridge University Press.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1), 92-111. doi: http://dx.doi.org/10.1016/j.jml.2008.06.003

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, 113(2), 1001-1024. doi: http://dx.doi.org/10.1121/1.1534836

Ben-David, B. M., Erel, H., Goy, H., & Schneider, B. A. (2015). "Older is always better": Age-related differences in vocabulary scores across 16 years. *Psychology and Aging*, 30(4), 856-862. doi: 10.1037/pag0000051

Borovsky, A., Elman, J., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, 112(4), 417–436. doi: http://doi.org/10.1016/j.jecp.2012.01.005

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS ONE*, *5*(5), 1-13.

Buz, E., & Jaeger, T. F. (2016). The (in)dependence of articulation and lexical planning during isolated word production. *Language, Cognition and Neuroscience,* 31(3), 404-424. doi: 10.1080/23273798.2015.1105984

Cheimariou, S. (2016). *Prediction in aging language processing*. Doctoral dissertation, University of Iowa. Retrieved from http://ir.uiowa.edu/etd/3056/

Davies, M. (2008). The Corpus of Contemporary American English (COCA): 520 million words, 1990-present. Available from: http://corpus.byu.edu/coca/

Dell, G. S., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society London,* 369(1634), 20120394. doi: 10.1098/rstb.2012.0394

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience,* 8(8), 1117-1121. doi: 10.1038/nn1504

DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, 121(3), 226-239. doi: 10.1016/j.bandl.2012.02.006

Drake, E., & Corley, M. (2015). Articulatory imaging implicates prediction during spoken language comprehension. *Memory & Cognition,* 43(8), 1136-1147. doi: 10.3758/s13421-015-0530-6

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115(3), 149-161. doi: http://dx.doi.org/10.1016/j.bandl.2010.07.006

Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, 42(2), 133-141. doi: 10.1111/j.1469-8986.2005.00274.x

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, 39(2), 133-146. doi: 10.1017/S0048577202001373

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491-505. doi: 10.1111/j.1469-8986.2007.00531.x

Ferreira, V. S., & Griffin, Z. M. (2003). Phonological influences on lexical (mis)selection. *Psychological Science*, 14(1), 86-90. doi: 10.1111/1467-9280.01424

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 101(6), 3728-3740. doi: 10.1121/1.418332

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of Contextual Predictability and Transitional Probability on Eye Movements During Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 31*(*5), 862-877. doi:10.1037/0278-7393.31.5.862

Gambi, C., & Pickering, M. J. (2011). A cognitive architecture for the coordination of utterances. *Frontiers in Psychology,* 2 (275). doi: 10.3389/fpsyg.2011.00275

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, 140(2), 186-209. doi: 10.1037/a0022256

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. Chicago Linguistics Society (CLS-99), 151-166.

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and levels of processing in spoken word production. *Journal of Memory and Language*, 38(3), 313-338. doi: http://dx.doi.org/10.1006/jmla.1997.2547

Hartshorne, J. K., & Germine, L. T., (2015). When Does Cognitive Functioning Peak? The Asynchronous Rise and Fall of Different Cognitive Abilities Across the Life Span. *Psychological Science,* 26(4), 433-443. doi: 10.1177/0956797614567339

Hintz, F. (2015). *Predicting language in different contexts: The nature and limits of mechanisms in anticipatory language processing*. PhD Thesis, Radboud University Nijmegen.

Hintz, F., Meyer, A. S., & Huettig, F. (2016). Encouraging prediction during production facilitates subsequent comprehension: Evidence from interleaved object naming in sentence context and sentence reading. *Quarterly Journal of Experimental Psychology*, 69(6), 1056-1063. doi:10.1080/17470218.2015.1131309.

Huang, H.-W., Meyer, A. M., & Federmeier, K. D. (2012). A "concrete view" of aging: Event related potentials reveal age-related changes in basic integrative processes in language. *Neuropsychologia*, 50(1), 26–35. doi: http://doi.org/10.1016/j.neuropsychologia.2011.10.018

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, 1626, 118-135. doi:10.1016/j.brainres.2015.02.014.

Huettig, F., & Mishra, R. K. (2014). How literacy acquisition affects the illiterate mind - A critical examination of theories and evidence. *Language and Linguistics Compass*, 8(10), 401-427. doi:10.1111/lnc3.12092

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.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, 86, 157-171. doi: http://dx.doi.org/10.1016/j.jml.2015.10.007

Janse, E., & Jesse, A. (2014). Working memory affects older adults' use of context in spoken-word recognition. *Quarterly Journal of Experimental Psychology*, 67(9), 1842-1862. doi: 10.1080/17470218.2013.879391

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In J. L. Bybee & P.J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229-254). Amsterdam: John Benjamins.

Kahn, J. M., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*, 67(3), 311-325. doi: http://dx.doi.org/10.1016/j.jml.2012.07.002

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of Syntactic and Semantic Information in Predictive Processing: Cross-Linguistic Evidence from German and English. *Journal of Psycholinguistic Research,* 32(1), 37-55. doi: 10.1023/a:1021933015362

Kello, C. T. (2004). Control over the time course of cognition in the tempo-naming task. *Journal of Experimental Psychology: Human Perception and Performance,* 30(5), 942-955. doi: 10.1037/0096-1523.30.5.942

Kleinman, D., Runnqvist, E., & Ferreira, V. S. (2015). Single-word predictions of upcoming language during comprehension: Evidence from the cumulative semantic interference task. *Cognitive Psychology,* 79, 68-101. doi: 10.1016/j.cogpsych.2015.04.001

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology,* 16*,* 262-284. doi: 10.1080/09541440340000213

Laszlo, S., & Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology,* 48(2), 176-186. doi: 10.1111/j.1469-8986.2010.01058.x

Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and causal text constraints. *Discourse Processes*, 34, 259–280. doi: 10.1006/jmla.1994.1008

McDonald, S. A., & Shillcock, R. C. (2003a). Eye Movements Reveal the On-Line Computation of Lexical Probabilities During Reading. *Psychological Science*, 14(6), 648-652. doi: 10.1046/j.0956-7976.2003.psci_1480.x

McDonald, S. A., & Shillcock, R. C. (2003b). Low-level predictive inference in reading: the influence of transitional probabilities on eye movements. *Vision Research*, 43(16), 1735-1751. doi: http://dx.doi.org/10.1016/S0042-6989(03)00237-2

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye movements are modulated by reading ability: Evidence from Indian low and high literates. *Journal of Eye Movement Research*, 5(1), 1-10. doi:10.16910/jemr.5.1.3.

Moers, C., Janse, E., & Meyer, A. S. (2015). Probabilistic reduction in reading aloud: A comparison of younger and older adults. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.), *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*. London: International Phonetics Association.

Ong, J. K. Y. & Kliegl, R. (2008). Conditional co-occurrence probability acts like frequency in predicting fixation durations. *Journal of Eye Movement Research*, 2(1):3, 1-7. doi: http://dx.doi.org/10.16910/jemr.2.1.3

Payne, B. R., & Stine-Morrow, E. A. (2012). Aging, parafoveal preview, and semantic integration in sentence processing: testing the cognitive workload of wrap-up. *Psychology and Aging*, 27(3), 638-649. doi: 10.1037/a0026540

Piai, V., Roelofs, A., Rommers, J., & Maris, E. (2015). Beta oscillations reflect memory and motor aspects of spoken word production. *Human Brain Mapping,* 36(7), 2767-2780. doi: 10.1002/hbm.22806

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, 97(1), 593–608. doi: 10.1121/1.412282

Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169-190. doi: 10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329-347. doi: 10.1017/s0140525x12001495

Praat: Doing Phonetics by Computer, Version 5.3. Boersma, P. & Weenink, D., (2012).

Qaltrics, Provo, Utah, USA, 2014. Available from: http://www.qualtrics.com

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, 21(3), 448-465. doi: 10.1037/0882-7974.21.3.448

Rommers, J., Meyer, A. S., & Huettig, F. (2015). Verbal and nonverbal predictors of language-mediated anticipatory eye movements. *Attention Perception, & Psychophysics,* 77(3), 720-730. doi: 10.3758/s13414-015-0873-x

Severens, E., Ratinckx, E., Ferreira, V. S., & Hartsuiker, R. J. (2008). Are phonological influences on lexical (mis)selection the result of a monitoring bias? *Quarterly Journal of Experimental Psychology (2006)*, 61(11), 1687–1709. http://doi.org/10.1080/17470210701647422

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *Journal of the Acoustical Society of America*, 123, 489–499. doi: 10.1121%2F1.2783762

Shipley W. C., Gruber C. P, Martin T. A., Klein A. M. (2009). Shipley-2. Los Angeles, CA: Western Psychological Services.

Shriberg, E., & Stolcke, A. (1996). Word predictability after hesitations: A corpus-based study. In *Proceedings of the International Conference on Spoken Language Processing* 3: 1868-1871.

Staub, A., & Clifton, C. (2006). Syntactic Prediction in Language Comprehension: Evidence From Either…or. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 32(2), 425-436. http://doi.org/10.1037/0278-7393.32.2.425

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language,* 82, 1-17. doi: http://dx.doi.org/10.1016/j.jml.2015.02.004

Van Berkum, J. J. A., Brown, M. C., Zwitserlood, P., Kooijman, V. & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443–467. doi: 10.1037/0278-7393.31.3.443

Van den Noort, M., Haverkort, M., Bosch, P., & Hugdahl, K. (2006). Is there a decline in verbal working memory over age?. *Europe's Journal of Psychology,* 2(2). doi:10.5964/ejop.v2i2.321

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging,* 18(2), 332-339. doi:10.1037/0882-7974.18.2.332

Wang, H.-C., Pomplun, M., Chen, M., Ko, H., & Rayner, K. (2010). Estimating the effect of word predictability on eye movements in Chinese reading using latent semantic analysis and transitional probability. *The Quarterly Journal of Experimental Psychology*, 63(7), 1374-1386. doi: 10.1080/17470210903380814

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition,* 99(2), B63-B72. doi: http://dx.doi.org/10.1016/j.cognition.2005.07.001

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, 68, 20-32. doi: http://dx.doi.org/10.1016/j.cortex.2015.03.01

CHAPTER 6

SUMMARY AND CONCLUSIONS

Summary and Conclusions

The key question that all four empirical studies in this thesis addressed was how aging and language experience modify probabilistic effects in predictive language processing. As such, the findings reported in the present dissertation contribute to psycholinguistic research in at least three ways: First, the findings allow for a more elaborate view on the ways in which probabilistic knowledge can influence lexical processing and acoustic realization. Second, the findings are informative with respect to the benefits and limitations of linguistic predictions in different language tasks. Third, the research specifies in more detail under which circumstances cognitive aging in later adulthood may or may not alter the use of predictability information throughout the life span.

In this chapter, I will first summarize the findings from my research studies. Subsequently, I will discuss the results in a broader, psycholinguistic perspective and suggest some directions for future research.

Summary of results

In the corpus study reported in Chapter 2, probabilistic effects on reading durations were analyzed in speech samples (taken from the Dutch JASMIN corpus) of three different age groups: children, adolescents, and older adults. Predictability was operationalized as a combined predictor of forward transitional probability (which is the likelihood of a target word given the preceding word) and backward transitional probability (i.e., the likelihood of a target word given the following word). The analysis yielded stable effects of Transitional Probabilities (TPs) on spoken word durations: all groups of readers showed probability-based facilitation effects, such that more likely words were acoustically more reduced relative to less likely words. Note, though, that the analyses did not allow me to draw straightforward conclusions about the effect of single-word frequency on probabilistic reduction across the different age group due to collinearity between frequency and TP. With regard to age-related differences in predictability effects, the following patterns were observed: Within the young child reader group, probabilistic effects slightly increased with age, indicating that probabilistic knowledge about how often words co-occur is building up during childhood. Within the adolescent group, there were no such significant interactions between age and TP. Within the group of older adults, there were again no significant interactions between age and TP. Note, however, that in an age-group comparison between older adults and adolescents, older adults had slightly smaller predictability effects than adolescents. Hence, I concluded that facilitatory TP effects level off after adolescence.

For the research in Chapters 3 and 4, I designed a follow-up experiment on changes in probabilistic effects in adulthood, in which I investigated age-related changes in frequency and co-occurrence effects in silent reading versus reading aloud using an eye-tracking paradigm. Younger and older adults read sentences that contained Dutch noun-verb combinations varying in frequency and co-occurrence predictability both silently and aloud. In Chapter 3, the analysis of the spoken word duration data from the reading aloud task confirmed the main findings from Chapter 2: word probabilities influenced speech articulation, such that more likely words were acoustically more reduced. This effect only occurred for backward TP, while forward TP had no significant influence on spoken word durations. Moreover, backward TP effects were similar in size in younger and older adults, which shows that changes in the use of probabilistic information are very subtle, if any, across adulthood.

Subsequently, in Chapter 4 of this thesis I analyzed the eye-tracking data from the experiment on silent reading versus reading aloud. Once again, TPs influenced reading durations, such that likely words needed less processing time, relative to words that were less likely in their context. Both older and younger adults showed these facilitatory probabilistic effects. While predictability effects were similar in size, frequency effects differed across age groups: In fixation durations older adults showed larger frequency effects relative to younger adults, although the opposite pattern (smaller frequency effects in older adults than younger adults) occurred in the analysis of skipping rates. In this study, I also investigated task-related differences (reading aloud versus silent) on the use of probabilistic information, but it turned out that frequency and predictability effects on gaze behavior during reading were equal in both tasks. This task similarity was observed across age groups, but also for each age group individually. The most salient differences between silent reading and reading aloud were observed in increased total fixation durations in reading aloud, particularly for the older adults, and in a differential skipping pattern across age groups. Because I was also interested in individual differences in frequency and predictability effects during reading, I had added measures of vocabulary size and attentional ability (Stroop and Flanker performance) to the study. However, the statistical analysis showed that there were no systematic effects of these background measures on the size of probabilistic effects and therefore the results were not reported in detail in Chapter 4.

Lastly, the study described in Chapter 5 investigated age-related changes in predictability effects during contextually constrained picture naming. In the previous studies I had used data from tasks that solely involved reading. The picture naming task in this last study included an additional layer of lexical processing, as the word form necessarily has to be retrieved from the semantic concept, whereas the word form is provided in reading. Additionally, the experiment was designed to contrast two different types of predictability information: global sentence predictability (cloze predictability) versus local word co-occurrence predictability. In my previous studies, age-related differences in TP predictability were minimal or in fact absent. For the final research study, I hypothesized that age may interact with the type of information that people readily process, such that younger and older adults differ in their use of cloze predictability, but not in TP predictability. Yet, the data analysis showed that older adults did not significantly differ from younger adults in their use of prediction: cloze and TP predictability affected picture naming times (i.e., response times) equally in both age groups. I also analyzed the spoken word duration data from the picture responses, in order to link this study to the preceding studies in investigating whether ease of lexical retrieval affected articulation. However, neither cloze nor TP predictability affected word durations in this study. The only effect that persisted throughout the analysis of picture naming times and word durations was the effect of age-related slowing. Older adults needed more time than younger adults to start and to articulate the response, independently from their use of prediction. Additionally, as for the reading study reported in Chapter 4, I investigated whether productive or receptive vocabulary predicted the size of predictability effects (cf. Federmeier, McLennan, De Ochoa, & Kutas, 2002) in older versus younger adults in picture naming, but found no such contribution of vocabulary size.

Taken together, the following key findings and conclusions can be derived from my studies: First, probabilistic knowledge, as for instance quantified by transitional probability, generally facilitates language processing. Consequently, words that are more frequent and likely given their context require less processing time and are produced with less articulatory effort, relative to less likely words. The facilitatory effect of contextual predictability on lexical processing and pronunciation is, of course, not a novel finding in itself. However, the research in this thesis offers more systematic insights into the levels at which local word predictability (in the form of forward and backward TP) may exert its influence. In this thesis, facilitatory effects were found in eye fixation behavior, picture naming latencies, as well as in spoken word durations across different language tasks that combined comprehension and production processes. As such, TPs seem to particularly influence lexical processing, and to spill over to articulation under certain circumstances (e.g., during the articulation of longer stretches of speech). Note that one interesting finding was the difference in backward TP effects versus forward TP effects: forward TP influenced eye-movements in reading (Chapter 4) and response latencies in contextually-constrained picture naming (Chapter 5), but *NOT* word durations (neither in Chapter 3 in reading aloud, nor in Chapter 5 in picture naming); in contrast, backward TP influenced both eye fixation behavior (see Chapter 4) and speech production in the form of word durations (Chapter 3). As such, backward TP seems to be more important in speech articulation than forward TP, while forward TP affects language processing up to the start of motor movements. These differences may be linked to the coordination processes involved in the preparation of speech. By definition, forward TP reflects predictability of a current word on the basis of previously encountered words, which may be important for the pre-activation of the lexical representations of an upcoming word. If the current word is unlikely given previous context, it will be less pre-activated and consequently it will take a longer time to initiate the lexical selection of the current word (which leads to a longer response latency in contextually-constrained picture naming, for instance). Backward TP measures the dependency of a current word on upcoming context. If the current word is unlikely given the context ahead, it means that the upcoming words are more difficult to process and more time or resources need to be spent on them. Hence the articulation of the current word may be lengthened in order to buy time for processing the next word. This serial coordination mechanism (word-by-word) and the constant look ahead may explain the relative importance of backward TP effects in speech: upcoming context determines how much time will be spent to articulate the current word, while previous context solely determines how long it took to preactivate and initiate the word selection (and therefore forward TP may influence the actual articulation less compared to backward TP; for a discussion of FTP vs. BTP effects on articulation see also Bell, Brenier, Gregory, Girand, & Jurafsky, 2009).

Second, children, adolescents, younger adults, and older adults (aged above 60) make use of probabilistic information during language processing in similar ways. In this thesis, all groups benefitted from local context information, and showed probabilistic processing reduction. While there may be small age- or experience-related differences in the size of predictability effects, the studies confirm that probabilistic processing is a fundamental principle in human language use across diverse groups. Previous studies had demonstrated the strong influence of word frequency and other lexical statistics on language processing within childen or adults (e.g., Cohen-Shikora & Balota, 2016). The current research (and particularly Chapter 2) contributes to the field by way of a more in-depth investigation of local word-to-word predictability effects (like TP effects) in sentence reading of heterogeneous participant groups.

Third, older adults barely differ from younger adults in their processing of predictability information during relatively simple language tasks, such as normal reading or picture naming. Note, that manifold studies have reported such age-related differences in prediction during reading, speech comprehension or speaking (e.g., DeLong, Groppe, Urbach & Kutas, 2012; Federmeier, Kutas, & Schul, 2010; Federmeier & Kutas, 2005; Federmeier et al., 2002; Roe et al., 2000; but see e.g., Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006). However, in this thesis none of the studies involved any processing pressure (such as time pressure, secondary tasks) or misleading context information (i.e., that would disconfirm or hinder predictions), which may have been the case in earlier research studies. Rather, in the studies in this thesis the timing of the input information was largely controlled by participants: participants could take as much time as they needed to read or speak the words. Under these circumstances, older adults seem to hardly differ from younger adults in their language processing strategies.

The only age group difference that persisted across all studies in this thesis was that the older adults processed the materials more slowly than the younger ones. Age-related slowing was found in word durations in Chapter 2, Chapter 3, and Chapter 5, in eye gaze data in Chapter 4, and in response latencies as well as self-paced reading rate in Chapter 5. Although slowing is a very common finding in research studies with older participants, the causes often remain unclear, as many factors play a role. On the one hand, slowing may be caused by decreased processing speed, which would have consequences for all processing, transmission and articulatory execution stages (e.g., Salthouse, 1996). On the other hand, slowing may in fact be a "positive" sign of more complex and deeper processing of linguistic materials with language experience (for a discussion see Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014), or may reflect an active choice by older adults to speak carefully and to pay close attention to the task at hand (for discussions see e.g., Abrams & Farrell, 2011, and Mortensen, Meyer, & Humphreys, 2006). It is likely that age-related slowing was caused by a combination of these factors in the production and comprehension studies reported in this thesis.

Aging and predictability

As previously outlined in the summary of the results, a striking difference between earlier studies (e.g., Federmeier et al., 2010) and the research in this thesis lies in the finding of age-by-predictability interactions: some researchers have reported a decline in prediction effects in older adults relative to younger adults, which is generally not replicated in the studies in this thesis. The only exception to this general absence of decreased contextual facilitation for older adults was observed in Chapter 2, with older adults showing smaller TP effects than adolescents on their spoken word durations. However, in the same study, the within-senior group analysis – which included participants from a large age range from 62 to 95 years – showed stable predictability effects, rather than any age-by-predictability interaction. The discrepancy between the findings of this thesis and the literature on age-related decline in predictive processing can be explained in multiple ways.

First, EEG paradigms, which directly measure online processes, may simply better capture subtle differences between younger and older adults, relative to the indirect behavioral measures presented in this thesis in which group differences may be masked due to the measuring delays. Second, the use of predictive processing strategies, particularly in older adults, may be connected to stimulus presentation timing and the timing of semantic integration. DeLong and colleagues (2012), for instance, used an EEG paradigm in which the context sentences were presented at a fixed word-by-word presentation rate, which put readers, older readers in particular, under considerable time pressure. Here, a decrease of prediction effects was reported for older adults relative to younger adults. By contrast, the reading tasks in this thesis allowed participants to self-control the word input rate, such that participants could take as much time as they needed to process the context in which words occurred. Stimulus timing has been shown to be a crucial factor for predictive processing during language processing (cf. Ito, Corley, Pickering, Martin-Nieuwland, & Nieuwland, 2016; Linderholm 2002; Wlotko & Federmeier, 2015). Support for the "time-pressure" explanation also comes from other eye-tracking studies (e.g., Rayner et al., 2006; Kliegl, Grabner, Rolfs, & Engbert, 2004), which report no difference between younger and older adults in predictability effects if whole sentences are presented and participants can view sentences as long as they liked. Given that older adults generally processed stimuli more slowly than younger adults (cf. age-related slowing observed in all chapters), any age-related decline in predictability effects may have been washed out by the self-paced timing. Note, however, that in an unpublished dataset we found no age group-by-predictability interaction, even in a dual-task situation in which participants listened to auditory lead-in sentences and named sentence-final pictures while walking on a treadmill. Hence, more demanding task conditions and the lack of self-controlled stimulus presentation rate may not necessarily hamper predictive processing in older adults per se.

The third explanation for the general absence of age-by-predictability interactions relates to the involvement of speech production in almost all tasks used in this thesis (except silent reading in Chapter 4). Participants may be more engaged (aroused) and/or motivated if the task actively requires language production as compared to comprehension-only tasks. Gollan and colleagues (2011), for instance, compared predictability effects in a picture naming and a reading task. They report that in both tasks predictable items were processed faster than less predictable items, but the predictability effect was much larger in the naming task than in the reading task. Furthermore, Hintz and colleagues (2016) found larger facilitation effects in mixed-task blocks, in which participants also had to name aloud the outcome of a predictive stimulus, relative to comprehension-only trials. The involvement of speech production in the studies throughout this thesis may have particularly encouraged older adults to engage in predictive processing. This is in line with recent theoretical models that stress the role of the speech production system during predictive processing (e.g., Dell & Chang, 2014; Pickering and Garrod 2013; for a discussion see Huettig, 2015). These models build on a tight coupling of language production and predictive processing, as they hypothesize that predictive processing involves the use of the internal production system. Under these accounts, generating a prediction means to create an internal representation about how oneself would produce likely upcoming input. After building an internal representation of expected words, the actual overt production of these words may be easier and prone to similar-sized facilitation effects in both younger and older adults.

The third factor that may have played into the observation of similar benefits of predictive processing for younger and older adults is that the design and materials of my research studies did not involve the contrast between predictable versus misleading stimuli. Presenting misleading sentence context (or contrasting logical versus illogical sentence continuation) may discourage predictive processing because participants are "punished" for predicting. Rather, I used a range of differentially predictable – yet always plausible – contexts for target words. Previous research studies had largely used different experimental conditions, in which predictable trials were compared to semantically incongruent trials such that conflicts or mismatches occurred between context and target words. Aging has been linked to a decline in the ability to ignore irrelevant context (cf. Mortensen, Meyer, & Humphreys, 2006). Thus, the effect of "interfering" context information, or violated predictions, may particularly hinder (or confuse) older adults in their processing strategies, while any "facilitation" from congruent context information (more or less predictable) may be similar in younger and older adults (cf. Roe et al., 2000). The enlarged interference effects in older adults are predicted by accounts of decreased inhibitory control in later adulthood (e.g., the inhibition deficit model; Lustig, Hasher

& Zacks, 2007; cf. also the study by Lash, Rogers, Zoller, & Wingfield, 2013), as for instance in situations in which a logical, predicted sentence continuation strongly competes with an incongruent continuation. However, there are no such expected age differences for "facilitation" effects when the sentence context may support multiple, relatively predictable words, as in my studies.

Conclusion and Outlook

This thesis focused on investigating interactions between three lines of psycholinguistic research: research on how language users may generate predictions, how probabilistic knowledge benefits lexical access, and how cognitive aging – as well as the age-associated increase in language experience – affects language processing. In our daily language use, words usually occur in context and as such in particular recurrent combinations with other words. In this thesis, these co-occurrences were quantified by drawing on large-scale corpora and computational methods. Experience with a language allows users to pick up on fine statistical patterns and fosters the generation of associations between words in the mental lexicon. As such, probabilistic knowledge supports predicting upcoming language input. The empirical studies in this thesis showed that there is little change to the use of this probabilistic information in later adulthood, as older adults largely maintain using probabilistic knowledge in order to facilitate their lexical processing. Evidence of stable facilitation effects across adulthood was found for probabilistic reduction on spoken word durations (in Chapter 2 within a group of older adults with a large age-range, and in Chapter 3 for younger vs. older adults), for TP effects on eye gaze patterns in reading (Chapter 4), and for TP as well as cloze predictability effects on picture naming speed (Chapter 5).

While the research in this thesis contributed to the study of language processing over the life span, further research is needed to specify, among others, the following issues. First, it is still unclear whether predictive processing may be hampered (particularly in older adults) if language processing is made more demanding by either changing the language task or the circumstances. Hence, future studies on prediction in the context of aging may vary the time pressure (i.e., increase stimulus presentation rate, impede the processing of lead-in sentences), or use challenging dual-task situations, in order to modulate the benefit of setting up predictions. This would clarify whether the time course during language processing and available cognitive resources indeed determine whether predictions are beneficial in later adulthood, or not. Second, despite the large body of literature demonstrating the importance of prediction for language processing, it is not clear which types of predictive processes are mandatory components of language processing and which are optional strategies. Additionally, for those predictive processes that are choice parameters, do language users either become more or less flexible in that choice with increasing age and language experience? Third, earlier research had shown that individual differences, such as vocabulary knowledge, inhibitory skills, and working memory (e.g., Cheimariou, 2016), are mediating factors in the use of probabilistic knowledge and prediction. Yet, individual differences seemed to hardly play any role in using simple, word co-occurrence patterns to benefit language processing. Possibly, this may be due to the fact that there was little variation in the individual differences measures and participant samples. Further research is needed to determine how individual differences contribute to acquiring and storing probabilistic knowledge and on drawing on that knowledge for language processing across the life span. Relatedly, future research should investigate whether working memory capacity, and age-related changes in working memory, influence the use of local versus more global predictability information during language processing.

References

Abrams, L., & Farrell, M.T. (2011). Language Processing in Normal Aging. In Guendouzi, J., Loncke, F., & Williams, M.J. (Eds.), *The Handbook of Psycholinguistic and Cognitive Processes: Perspectives in Communication Disorders* (pp. 49-74). New York: Psychology Press.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1), 92-111. doi: http://dx.doi.org/10.1016/j.jml.2008.06.003

Cheimariou, S. (2016). *Prediction in aging language processing*. Doctoral dissertation, University of Iowa. Retrieved from http://ir.uiowa.edu/etd/3056/

Cohen-Shikora, E. R., & Balota, D. A. (2016). Visual Word Recognition Across the Adult Lifespan. *Psychology and Aging*, 31(5), 488-502. doi: 10.1037/pag0000100

Dell, G. S., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society London*, 369(1634), 20120394. doi: 10.1098/rstb.2012.0394

DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, 121(3), 226-239. doi: 10.1016/j.bandl.2012.02.006

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115(3), 149-161. doi: http://dx.doi.org/10.1016/j.bandl.2010.07.006

Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, 42(2), 133-141. doi: 10.1111/j.1469-8986.2005.00274.x

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, 39(2), 133-146. doi: 10.1017/S0048577202001373

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, 140(2), 186-209. doi: 10.1037/a0022256

Hintz, F., Meyer, A. S., & Huettig, F. (2016). Encouraging prediction during production facilitates subsequent comprehension: Evidence from interleaved object naming in sentence context and sentence reading. *Quarterly Journal of Experimental Psychology*, 69(6), 1056-1063. doi:10.1080/17470218.2015.1131309.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, 1626, 118-135. doi:10.1016/j.brainres.2015.02.014.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, 86, 157-171. doi: http://dx.doi.org/10.1016/j.jml.2015.10.007

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16, 262-284. doi: 10.1080/09541440340000213

Lash, A., Rogers, C. S., Zoller, A., & Wingfield, A. (2013). Expectation and Entropy in Spoken Word Recognition: Effects of Age and Hearing Acuity. *Experimental Aging Research*, 39(3), 235–253. doi: http://doi.org/10.1080/0361073X.2013.779175

Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and causal text constraints. *Discourse Processes*, 34, 259–280. doi: 10.1006/jmla.1994.1008

Lustig, C., Hasher, L., & Zacks, R. T. (2007). Inhibitory deficit theory: Recent developments in a "new view". In Gorfein, D. S., & MacLeod, C. M. (Eds.), *Inhibition in cognition* (pp. 145-162). Washington, DC, US: American Psychological Association.

Mortensen, L., Meyer, A. S., & Humphreys, G. W. (2006). Age-related effects on speech production: A review. *Language and Cognitive Processes*, 21(1-3), 238-290. doi: 10.1080/01690960444000278

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329-347. doi: 10.1017/s0140525x12001495

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

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, 21(3), 448-465. doi: 10.1037/0882-7974.21.3.448

Roe, K., Jahn-Samilo, J., Juarez, L., Mickel, N., Royer, I., & Bates, E. (2000). Contextual effects on word production: A lifespan study. *Memory & Cognition*, 28(5), 756-765. doi: 10.3758/bf03198410

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103(3), 403-428. doi: 10.1037/0033-295x.103.3.403

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, 68, 20-32. doi: http://dx.doi.org/10.1016/j.cortex.2015.03.014

Nederlandse samenvatting

Wanneer we taal verwerken, gebruiken we kennis over hoe waarschijnlijk het is dat bepaalde woorden in een bepaalde context voorkomen. Effecten van deze voorspelbaarheid op de snelheid waarmee we taal kunnen gebruiken, noemen we *probabilistische effecten*. Hoe vaker we een woord lezen, horen, schrijven, zeggen, des te beter kunnen we inschatten hoe voorspelbaar dit woord in een bepaalde context is. Het zou dus kunnen dat naarmate we ouder worden en meer *taalervaring* opdoen, we de informatie over de waarschijnlijkheid van woorden anders gaan gebruiken. De vier studies beschreven in hoofdstukken twee tot en met vijf deden hier onderzoek naar.

Hoofdstuk 2 rapporteert een corpusstudie. Deze studie onderzocht het effect van de voorspelbaarheid van woorden op de duur van die woorden wanneer ze hardop gelezen worden. Met behulp van gesproken taalsamples uit het Nederlandse JASMIN corpus werd een vergelijking gemaakt tussen drie leeftijdsgroepen: kinderen, adolescenten en ouderen. De voorspelbaarheid van een woord werd uitgedrukt als de *overgangswaarschijnlijkheid* van dat woord. In deze overgangswaarschijnlijkheidsmaat (*transitional probability*, ofwel TP) werden voortwaartse en achterwaartse overgangswaarschijnlijkheid gecombineerd, waarbij voorwaartse overgangswaarschijnlijkheid de waarschijnlijkheid aangeeft van een woord gegeven het woord ervoor en achterwaartse waarschijnlijkheid de waarschijnlijkheid van een woord gegeven het erop volgende woord. Alle groepen proefpersonen toonden *probabilistische facilitatie-effecten*, wat inhoudt dat meer waarschijnlijke woorden akoestisch meer gereduceerd werden dan minder waarschijnlijke woorden. De duur van een gesproken woord werd dus beïnvloed door de voorspelbaarheid van dat woord. Omdat dit effect van TP sterk correleerde met effecten van woordfrequentie, konden er geen eenduidige conclusies getrokken worden over het effect van de frequentie van losse woorden op akoestische reductie in de verschillende leeftijdsgroepen. Wat betreft leeftijdgerelateerde verschillen kunnen we wel zeggen dat het TP-effect in de groep kinderen iets toenam met leeftijd. Dit betekent dat kennis over hoe waarschijnlijk het is dat woorden samen voorkomen, opgebouwd wordt tijdens onze kindertijd. Zulke interacties tussen leeftijd en TP werden niet gevonden in de groep adolescenten en ook niet voor de oudere volwassenen. Toch is het belangrijk om op te merken dat ouderen minder grote TP-effecten vertoonden dan adolescenten. Het lijkt erop dat informatie over de voorspelbaarheid van woorden minder gebruikt wordt bij de taalverwerking na de adolescentie.

Hoofdstuk drie en vier beschrijven een vervolgexperiment over de verandering in effecten van voorspelbaarheid in volwassenen. Daarin onderzocht ik of effecten van woordfrequentie en effecten door kennis over het samen voorkomen van woorden veranderen met de leeftijd. Jongere en oudere volwassenen lazen Nederlandse zinnen waarin combinaties van een zelfstandig naamwoord en een werkwoord (zoals bijvoorbeeld “afspraak maken”) voorkwamen die verschilden in frequentie en in hoe vaak de woorden in de combinatie samen gebruikt worden. Proefpersonen lazen deze zinnen in zichzelf en hardop (spraakresultaten van het hardop lezen worden besproken in hoofdstuk drie). Hierbij werden oogbewegingen gemeten door middel van een eye-tracking opstelling (oogbewegingsresultaten in hoofdstuk vier).

De spraakresultaten van het hardop lezen bevestigden de bevindingen uit de corpusstudie in hoofdstuk twee. Opnieuw bleek dat de voorspelbaarheid van een woord de uitspraak ervan beïnvloedde. Woorden met een hogere achterwaartse overgangswaarschijnlijkheid werden akoestisch meer gereduceerd dan minder waarschijnlijke woorden. Voorwaartse overgangswaarschijnlijkheid daarentegen had geen significante invloed op de duur van gesproken woorden. De grootte van het achterwaartse TP-effect was gelijk voor beide leeftijdsgroepen. Als het gebruik van informatie over de waarschijnlijkheid van woorden al verandert met leeftijd in de volwassenheid, zijn de veranderingen in elk geval erg subtiel.

In hoofdstuk vier worden de eye-trackingdata geanalyseerd. Daarbij worden fixaties bij stillezen en hardop lezen met elkaar vergeleken. Ook hier bleek dat TPs de leesduur beïnvloedden. Het kostte minder tijd om waarschijnlijke woorden in hun context te verwerken dan minder waarschijnlijke woorden. Dit was in gelijke mate het geval voor jongere en oudere volwassenen. Effecten van woordfrequentie verschilden wel voor beide groepen: bij ouderen was er een groter frequentie-effect te zien op de duur van fixaties dan bij jongere volwassenen. Het verschil in fixatieduur voor vaak voorkomende woorden (makkelijker, kortere fixatietijd) versus minder frequente woorden (moeilijker, langere fixatietijden) was dus groter bij oudere volwassenen. Het omgekeerde effect (kleinere frequentie-effecten voor oudere versus jongere volwassenen) werd gevonden in de *skipping rates.* In deze studie heb ik ook getoetst of het gebruik van informatie over de voorspelbaarheid van woorden verschilt in verschillende leestaken (hardop lezen versus stillezen). De effecten van frequentie en overgangswaarschijnlijkheid waren echter even groot in beide soorten taken voor beide leeftijdsgroepen. De grootste verschillen tussen stillezen en hardop lezen waren (1) de langere totale duur van fixaties tijdens het hardop lezen, vooral bij oudere volwassenen, en (2) een verschillend patroon in skipping rates voor de twee manieren van lezen voor beide leeftijdsgroepen. Omdat ik ook geïnteresseerd was in mogelijke individuele verschillen in effecten van woordfrequentie en voorspelbaarheid tijdens het lezen, heb ik twee variabelen aan de studie toegevoegd, namelijk de grootte van de woordenschat van de proefpersonen en hoe goed proefpersonen hun aandacht ergens op kunnen richten (gemeten met de Strooptaak en de Flankertaak). Statistische analyse toonde aan dat deze twee extra variabelen geen systematisch effect hadden op de grootte van effecten van voorspelbaarheid.

Tenslotte onderzocht ik in de studie beschreven in hoofstuk vijf leeftijdgerelateerde veranderingen in effecten van voorspelbaarheid tijdens een plaatjesbenoemtaak waarbij het benoemen voorafgegaan werd door het lezen van een contextzin. In de vorige experimenten lag de focus op taken waarin proefpersonen enkel hoefden te lezen. Deze plaatjesbenoemtaak voegt een extra laag in de verwerking van woorden toe, aangezien hier de woordvorm eerst nog opgehaald moet worden aan de hand van het concept (het plaatje). Bij lezen daarentegen is de woordvorm al gegeven. Bovendien kon ik met dit experiment twee verschillende soorten voorspelbaarheid met elkaar vergelijken, namelijk *globale voorspelbaarheid* in een zin (*cloze predictability*) versus *locale voorspelbaarheid* (*transitional probability,* TP) van combinaties van opeenvolgende woorden. In de vorig studies waren er geen of erg kleine leeftijdgerelateerde verschillen in het gebruik van overgangswaarschijnlijkheid. In dit laatste experiment was mijn hypothese dat er een interactie zou kunnen zijn tussen leeftijd en het soort waarschijnlijkheidsinformatie dat mensen snel verwerken. Jongere en oudere volwassenen zouden kunnen verschillen in het gebruik van globale voorspelbaarheid, maar wellicht niet in het gebruik van locale voorspelbaarheid. Data-analyse heeft echter aangetoond dat beide leeftijdsgroepen hier niet in verschillen: globale en locale voorspelbaarheid hadden in beide groepen een even groot effect op de reactietijden van proefpersonen, dat wil zeggen op de tijd die ze nodig hadden om een plaatje te benoemen. Verder heb ik ook de duur van de gesproken woorden geanalyseerd om deze data in verband te kunnen brengen met de vorige studies en te kunnen kijken of voorspelbaarheid een effect had op de uitspraak van de afgebeelde objecten. Noch globale, noch locale voorspelbaarheid had echter een effect op de duur van het woord. Wat wel duidelijk werd uit de analyse van reactietijden en woordduur, was een leeftijdgerelateerd effect van snelheid. Oudere volwassenen hadden over het algemeen meer tijd nodig dan jongere volwassenen om hun uitspraak te beginnen en te voltooien, los van hun gebruik van informatie over de voorspelbaarheid van woorden. Net als in hoofdstuk vier, had de grootte van woordenschat van proefpersonen geen effect op de resultaten voor beide leeftijdsgroepen.

De resultaten die gepresenteerd worden in deze thesis vormen op minstens drie manieren een belangrijke toevoeging aan psycholinguïstisch onderzoek. Ten eerste geven ze een breder beeld van hoe kennis over de voorspelbaarheid van woorden de lexicale verwerking en akoestische realisatie (de uitspraak) van woorden kan beïnvloeden. Ten tweede geven de resultaten duidelijk weer dat kinderen, adolescenten, jongvolwassenen en ouderen op vergelijkbare manieren gebruik maken van probabilistische informatie met betrekking tot overgangswaarschijnlijkheden tijdens het verwerken van taal. Tenslotte heeft dit onderzoek laten zien dat jongvolwassenen en ouderen nauwelijks verschillen in het gebruik van contextinformatie tijdens relatief eenvoudige taaltaken. Dit betekent dat cognitieve veroudering geen belemmering vormt voor het gebruikmaken van contextinformatie zolang de taaltaak de taalgebruiker maar niet onder druk zet of met vreemde zinswendingen confronteert.

Verder onderzoek is nodig om uit te zoeken of en hoe er mogelijk individuele verschillen bestaan in het verwerven, opslaan of gebruiken van probabilististische taalinformatie. Vervolgonderzoek zou zich ook kunnen richten op de vraag of werkgeheugen en leeftijdsgebonden veranderingen daarin samenhangen met het gebruik van locale dan wel globale waarschijnlijkheidsinformatie voor het taalverwerken.

Acknowledgements

This PhD brought me to three different research institutes, which had two things in common: beautiful minds that offered inspiration and friendship, and first-class infrastructure for doing research. Unfortunately, I cannot acknowledge each and everyone here, but I am most thankful for all the support I received.

First and foremost, my gratitude goes to you - Esther. You have been the best daily supervisor and PhD company I could ask for. You helped me out when I was struggling and gave me freedom when I followed my own path. Your positive feedback was often the one missing piece in my PhD puzzles. Not only did you round off my papers, polished my presentations and scattered my doubts, but your kind way of thinking and encouraging chats made me leave your office with a happy smile. Thank you also for the very personal, warm space you created outside the MPI, for instance when we (accidentally) had boerenkool-soup, you hid behind Kranenburg curtains for a perfect Conny-surprise, and when our group shared little Sinterklaas poems. Most importantly, you were always optimistic that things would work out in the end. And they did because of you.

Dear Antje, you offered me a research home in the PoL department and created an excellent work environment. I am still amazed by the speed with which you can give feedback on papers, posters and talks. You helped me overcome disappointments and your generosity enabled me to finally carry out the research visit to Northwestern University. Group events with dinners at your place fostered a positive & social atmosphere in your group. I am so very thankful for all your support.

Matt, those eight months in the US could have not been better spent than at your lab and Northwestern University's beautiful environment. You are a mastermind when it comes to constructive feedback on research ideas, data analysis and linking different research lines. Thank you for our wonderful collaboration, for excellent discussions, and for fun times in Evanston, which became my favorite place in the US (right after Alaska, of course) because of this experience.

I would also like to acknowledge the members of the reading committee, Prof. Mirjam Ernestus, Prof. Eva Belke, and Dr. Stefan Frank. Thank you for devoting your time and efforts to evaluating this thesis.

Looking back at my time as PhD student, I think that I personally benefitted the most from the company of creative, warm-hearted people in the research networks I joined. I learned so many things from my fellow PhDs and colleagues. Your open-mindedness is something that I take with me and will always cherish.

My dear paranymphs, I am very happy to call you my friends and to have you by my side during the defense. Kranenburg is such a tiny village, but you made it a great spot full of laughs, tasty food and comfort. Thordis, I will always have a big smile on my face when I think about our US road trip, your dedicated baking of Donauwellen and all the little chats we shared in and out of the MPI world. Caitlin, thank you for the occasional sharing of your cutie-cat plus your place with me, for keeping the ties even now I am far in the south, and for all the work you invested in helping me finalize this thesis.

Dear fellow Maxies, the six months in Rostock could have been an isolated time. Instead, this period was shaped by the privilege to be in our cohort and the luck of getting to know you as my friends. Of course, we shared fruitful discussions about aging research and science in general. But drinks at Warnow shore, burgers at Helgas Kitchen and trips to the beach were the very best that Rostock offered.

Dear PoL-members and everyone linked to the department, past or present – it was my pleasure to be part of this amazing group! Though you might remember me as the one who had office-breakfast while you were having lunch :-), I always enjoyed your company so much. Our lab meetings were linked to professional research discussions and exchanging ideas. The rest of the time was filled with your helping hands, funny jokes, and occasional Friday drinks. Some of you were also my office mates, and I would like to particularly thank Agnieszka, Joost, Joe, and Jeroen for providing happy vibes throughout the work day (and sometimes in the nights and weekends). Thanks also to you, Evelyn, for managing all department-related and personal things and for your always-open door.

Then there are so many other PhDs I need to acknowledge because they influenced my thinking and work-life balance so much. Louise, Jule, Gabriella, Carmen, Svetlana, Franziska, Xaver, Antje, I am speaking about you. Not only did we share the struggles of being PhD students, but what sticks with my memory is the in-depth conversations and private times we shared. I can only hope that we will see each other often, rather sooner than later.

A huge thank you also goes to the members of the SoundLab at Northwestern University, who welcomed me to their group and provided so much support. Sharing my time with you was such a privilege and Chicago winters (partly) lost their scariness when filled with your good vibes. A special thanks to Nayoun for your friendship and for introducing me to the

best Evanston food, even late at night. Another special thank you to Heidi (and of course to the cat who runs the world for kindly accepting my presence). You offered a warm and welcoming home full of good spirits, and I learned so much from your inspirational take on whatever happens in life.

Dear administration team, secretaries, technical group, library staff, canteen and reception people - for sure no daily business would be possible without running computers, functional experiment labs, paper-work support, open doors for participants, background literature and food. You do such an important job, thank you for being there.

I would also like to thank the coordinators of the IMPRS (Kevin, Dirkje, and Els), my 2012 IMPRS cohort, and the Max Planck International Research Network on Aging for organizing interesting events, good times and funding opportunities. Moreover, I would like to thank the CLS group for invitations to their lab meetings and feedback on my research. I am indebted to the younger and older participants, some of whom made it to the lab even under difficult conditions and who showed a keen interest in my work. Thank you to the student assistants for your help with annotations and data processing, and Maria – my intern – for your research contributions.

Zum Schluss ein Dank & Gruß an die Menschen, die die wertvolle Zeit außerhalb der Arbeit mit Leben füllen.

Meine Freunde aus der Bonner Zeit, liebe Island-Mädels, lieber Anton & Co, liebe Hofmans (bei euch gibt es noch immer die beste Tonatensauce dieser Welt): Danke an Euch für fröhliche Wiedersehen, Stunden voller Reden, Lachen & Feiern, und für das Gefühl, mit den richtigen Begleitern an allen Orten ein Zuhause finden zu können.

Anna, Kathrin, ihr beiden begleitet mich schon seit Jahrzehnten. Das macht uns zwar irgendwie alt, aber alte-Zeiten-Geschichten, Kochabende am "anderen Berg" und Lehreranekdoten sind eben auch der allerbeste Stressausgleich. Danke für eure Freundschaft und dass ihr immer dabei seid, selbst an anderen Enden der Welt.

Mama, Papa, Roland, Matthias und Vanessa - ihr seid was auch immer ich brauche: Umzugshelfer, Logistikzentrum, Katzensitter, IT-Support, Grillmeister, Reisebegleitung, Hausbank, Mentaltrainer, Seelsorger, Spaßvögel, Glücksbringer und meine allerliebste Donauwellen-Kaffeegesellschaft. Es ist so viel mehr als ich auflisten kann. Unsere Familie ist mein Ankerplatz und ohne diesen würden meine vielen, kleinen Abenteuer nichts bedeuten. Danke für Wurzeln und Flügel.

Curriculum Vitae

Cornelia Moers was born in 1984 in Olpe, Germany. In 2011, she obtained a Master’s degree (Magistra Artium) from the University of Bonn, Germany, where she studied Communication Research with a major in Phonetics, as well as English, and Psychology. As part of the ERASMUS program, she also spent one semester at the University of Iceland in Reykjavik, Iceland.

Cornelia was awarded a three-year doctoral scholarship from the Max Planck International Research Network on Aging (MaxNetAging) in 2011. The MaxNetAging Research School started in January 2012 with a six-month training at the Max Planck Institute for Demographic Research in Rostock, Germany. Afterwards, Cornelia joined the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands, where she conducted the major part of her research. From 2014 to 2015, Cornelia spent eight months at the Linguistics Department of Northwestern University, USA, where she collaborated with Professor Matt Goldrick for one of her research studies.

As of May 2016, Cornelia has been working with Nuance Communications in Ulm as a Natural Language Understanding engineer.

Publications

Moers, C., Meyer, A. S., & Janse, E. (2016). Effects of Word Frequency and Transitional Probability on Word Reading Durations of Younger and Older Speakers. *Language and Speech*, 60(2), 289-317. doi: 10.1177/0023830916649215

Moers, C., Janse, E., & Meyer, A. S. (2015). Probabilistic reduction in reading aloud: A comparison of younger and older adults. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.), *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*. London: International Phonetics Association.

Moers, C., Möbius, B., Rosanowski, F., Nöth, E., Eysholdt, U., & Haderlein, T. (2012). Vowel- and text-based cepstral analysis of chronic hoarseness**.** *Journal of Voice*, 26 (4), 416-424. doi: 10.1016/j.jvoice.2011.05.001

Haderlein, T., Moers, C., Möbius, B., & Nöth, E. (2012). Automatic Rating of Hoarseness by Text-based Cepstral and Prosodic Evaluation. *Proceedings of Text, Speech and Dialogue, 15th International Conference* (TSD 2012), Brnow, Czech Republic, 573-580.

Haderlein, T., Moers, C., Möbius, B., Rosanowski, F., & Nöth, E. (2012). Intelligibility Rating with Automatic Speech Recognition, Prosodic, and Cepstral Evaluation. *Proceedings of Text, Speech and Dialogue, 14th International Conference* (TSD 2011), Pilsen, Czech Republic, 195-202.

MPI Series in Psycholinguistics

1. *The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing*. Miranda van Turennout
2. *The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography.* Niels O. Schiller
3. *Lexical access in the production of ellipsis and pronouns*. Bernadette M. Schmitt
4. *The open-/closed-class distinction in spoken-word recognition*. Alette Haveman
5. *The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach*. Kay Behnke
6. *Gesture and speech production*. Jan-Peter de Ruiter
7. *Comparative intonational phonology: English and German*. Esther Grabe
8. *Finiteness in adult and child German*. Ingeborg Lasser
9. *Language input for word discovery*. Joost van de Weijer
10. *Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe*. James Essegbey
11. *Producing past and plural inflections*. Dirk Janssen
12. *Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea*. Anna Margetts
13. *From speech to words*. Arie van der Lugt
14. *Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language*. Eva Schultze-Berndt
15. *Interpreting indefinites: An experimental study of children’s language comprehension*. Irene Krämer
16. *Language-specific listening: The case of phonetic sequences*. Andrea Weber
17. *Moving eyes and naming objects*. Femke van der Meulen
18. *Analogy in morphology: The selection of linking elements in Dutch compounds*. Andrea Krott
19. *Morphology in speech comprehension*. Kerstin Mauth
20. *Morphological families in the mental lexicon*. Nivja H. de Jong
21. *Fixed expressions and the production of idioms*. Simone A. Sprenger

22. *The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria)*. Birgit Hellwig
23. *Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies*. Fermín Moscoso del Prado Martín
24. *Contextual influences on spoken-word processing: An electrophysiological approach*. Daniëlle van den Brink
25. *Perceptual relevance of prevoicing in Dutch.* Petra M. van Alphen
26. *Syllables in speech production: Effects of syllable preparation and syllable frequency*. Joana Cholin
27. *Producing complex spoken numerals for time and space*. Marjolein Meeuwissen
28. *Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction*. Rachèl J. J. K. Kemps
29. *At the same time...: The expression of simultaneity in learner varieties*. Barbara Schmiedtová
30. *A grammar of Jalonke argument structure*. Friederike Lüpke
31. *Agrammatic comprehension: An electrophysiological approach*. Marlies Wassenaar
32. *The structure and use of shape-based noun classes in Miraña (North West Amazon)*. Frank Seifart
33. *Prosodically-conditioned detail in the recognition of spoken words*. Anne Pier Salverda
34. *Phonetic and lexical processing in a second language*. Mirjam Broersma
35. *Retrieving semantic and syntactic word properties*. Oliver Müller
36. *Lexically-guided perceptual learning in speech processing*. Frank Eisner
37. *Sensitivity to detailed acoustic information in word recognition*. Keren B. Shatzman
38. *The relationship between spoken word production and comprehension*. Rebecca Özdemir
39. *Disfluency: Interrupting speech and gesture*. Mandana Seyfeddinipur
40. *The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation*. Christiane Dietrich
41. *Cognitive cladistics and the relativity of spatial cognition*. Daniel B.M. Haun
42. *The acquisition of auditory categories*. Martijn Goudbeek
43. *Affix reduction in spoken Dutch*. Mark Pluymaekers
44. *Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence*. Valesca Kooijman
45. *Space and iconicity in German Sign Language (DGS)*. Pamela Perniss

46. *On the production of morphologically complex words with special attention to effects of frequency*. Heidrun Bien
47. *Crosslinguistic influence in first and second languages: Convergence in speech and gesture*. Amanda Brown
48. *The acquisition of verb compounding in Mandarin Chinese*. Jidong Chen
49. *Phoneme inventories and patterns of speech sound perception*. Anita Wagner
50. *Lexical processing of morphologically complex words: An information-theoretical perspective*. Victor Kuperman
51. *A grammar of Savosavo, a Papuan language of the Solomon Islands*. Claudia Wegener
52. *Prosodic structure in speech production and perception*. Claudia Kuzla
53. *The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension*. Sarah Schimke
54. *Studies on intonation and information structure in child and adult German*. Laura de Ruiter
55. *Processing the fine temporal structure of spoken words*. Eva Reinisch
56. *Semantics and (ir)regular inflection in morphological processing*. Wieke Tabak
57. *Processing strongly reduced forms in casual speech*. Susanne Brouwer
58. *Ambiguous pronoun resolution in L1 and L2 German and Dutch*. Miriam Ellert
59. *Lexical interactions in non-native speech comprehension: Evidence from electro-encephalography, eye-tracking, and functional magnetic resonance imaging*. Ian FitzPatrick
60. *Processing casual speech in native and non-native language.* Annelie Tuinman
61. *Split intransitivity in Rotokas, a Papuan language of Bougainville*. Stuart Robinson
62. *Evidentiality and intersubjectivity in Yurakaré: An interactional account*. Sonja Gipper
63. *The influence of information structure on language comprehension: A neurocognitive perspective*. Lin Wang
64. *The meaning and use of ideophones in Siwu*. Mark Dingemanse
65. *The role of acoustic detail and context in the comprehension of reduced pronunciation variants*. Marco van de Ven
66. *Speech reduction in spontaneous French and Spanish*. Francisco Torreira
67. *The relevance of early word recognition: Insights from the infant brain*. Caroline Junge
68. *Adjusting to different speakers: Extrinsic normalization in vowel perception*. Matthias J. Sjerps

69. *Structuring language. Contributions to the neurocognition of syntax*. Katrien R. Segaert
70. *Infants' appreciation of others' mental states in prelinguistic communication: A second person approach to mindreading*. Birgit Knudsen
71. *Gaze behavior in face-to-face interaction*. Federico Rossano
72. *Sign-spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing space*. Conny de Vos
73. *Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning*. Attila Andics
74. *Lexical processing of foreign-accented speech: Rapid and flexible adaptation*. Marijt Witteman
75. *The use of deictic versus representational gestures in infancy*. Daniel Puccini
76. *Territories of knowledge in Japanese conversation*. Kaoru Hayano
77. *Family and neighbourhood relations in the mental lexicon: A cross-language perspective*. Kimberley Mulder
78. *Contributions of executive control to individual differences in word production*. Zeshu Shao
79. *Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing*. Patrick van der Zande
80. *High pitches and thick voices: The role of language in space-pitch associations*. Sarah Dolscheid
81. *Seeing what's next: Processing and anticipating language referring to objects*. Joost Rommers
82. *Mental representation and processing of reduced words in casual speech*. Iris Hanique
83. *The many ways listeners adapt to reductions in casual speech*. Katja Poellmann
84. *Contrasting opposite polarity in Germanic and Romance languages*: *Verum Focus and affirmative particles in native speakers and advanced L2 learners*. Giuseppina Turco
85. *Morphological processing in younger and older people: Evidence for flexible dual-route access*. Jana Reifegerste
86. *Semantic and syntactic constraints on the production of subject-verb agreement*. Alma Veenstra
87. *The acquisition of morphophonological alternations across languages*. Helen Buckler
88. *The evolutionary dynamics of motion event encoding*. Annemarie Verkerk
89. *Rediscovering a forgotten language*. Jiyoun Choi

90. *The road to native listening: Language-general perception, language-specific input*. Sho Tsuji
91. *Infants' understanding of communication as participants and observers*. Gudmundur Bjarki Thorgrímsson
92. *Information structure in Avatime*. Saskia van Putten
93. *Switch reference in Whitesands*. Jeremy Hammond
94. *Machine learning for gesture recognition from videos*. Binyam Gebrekidan Gebre
95. *Acquisition of spatial language by signing and speaking children: a comparison of Turkish sign language (TID) and Turkish*. Beyza Sümer
96. *An ear for pitch: on the effects of experience and aptitude in processing pitch in language and music*. Salomi Savvatia Asaridou
97. *lncrementality and Flexibility in Sentence Production*. Maartje van de Velde
98. *Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture*. Edwin van Leeuwen
99. *The request system in Italian interaction*. Giovanni Rossi
100. *Timing turns in conversation: A temporal preparation account*. Lilla Magyari
101. *Assessing birth language memory in young adoptees*. Wencui Zhou
102. *A social and neurobiological approach to pointing in speech and gesture*. David Peeters
103. *Investigating the genetic basis of reading and language skills*. Alessandro Gialluisi
104. *Conversation Electrified: The Electrophysiology of Spoken Speech Act Recognition*. Rósa Signý Gisladottir
105. *Modelling Multimodal Language Processing*. Alastair Smith
106. *Predicting language in different contexts: The nature and limits of mechanisms in anticipatory language processing*. Florian Hintz
107. *Situational variation in non-native communication*. Huib Kouwenhoven
108. *Sustained attention in language production*. Suzanne Jongman
109. *Acoustic reduction in spoken-word processing: Distributional, syntactic, morphosyntatic, and orthographic effects*. Malte Viebahn
110. *Nativeness, dominance, and the flexibility of listening to spoken language*. Laurence Bruggeman
111. *Semantic specificity of perception verbs in Maniq*. Ewelina Wnuk
112. *On the identification of FOXP2 gene enhancers and their role in brain development*. Martin Becker

113. *Events in language and thought: The case of serial verb constructions in Avatime*. Rebecca Defina
114. *Deciphering common and rare genetic effects on reading ability*. Amaia Carrión Castillo
115. *Music and language comprehension in the brain*. Richard Kunert
116. *Comprehending Comprehension: Insights from neuronal oscillations on the neuronal basis of language*. Nietzsche H.L. Lam
117. *The biology of variation in anatomical brain asymmetries*. Tulio Guadalupe
118. *Language processing in a conversation context*. Lotte Schoot
119. *Achieving mutual understanding in Argentine Sign Language*. Elizabeth Manrique
120. *Talking Sense: the behavioural and neural correlates of sound symbolism*. Gwilym Lockwood
121. *Getting under your skin: The role of perspective and simulation of experience in narrative comprehension*. Franziska Hartung
122. *Sensorimotor experience in speech perception*. Will Schuerman
123. *Explorations of beta-band neural oscillations during language comprehension: Sentence processing and beyond*. Ashley Glen Lewis
124. *Influences on the magnitude of syntactic priming*. Evelien Heyselaar
125. *Lapse organization in interaction*. Elliott Hoey
126. *The processing of reduced word pronunciation variants by natives and foreign language learners: Evidence from French casual speech*. Sophie Brand
127. *The Neighbors Will Tell You What To Expect: Effects of Aging and Predictability on Language Processing*. Cornelia Moers