

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

Language and Speech
2017, Vol. 60(2) 289–317

© The Author(s) 2016

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0023830916649215

journals.sagepub.com/home/las



Cornelia Moers

Max Planck Institute for Psycholinguistics, The Netherlands

Antje Meyer

Max Planck Institute for Psycholinguistics, The Netherlands
Radboud University Nijmegen, The Netherlands

Esther Janse

Radboud University Nijmegen, The Netherlands
Max Planck Institute for Psycholinguistics, The Netherlands

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 neighbouring 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.

Keywords

Word frequency, transitional probability, reading aloud, aging, lexical access

Corresponding author:

Cornelia Moers, Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands.

Email: cornelia.moers@mpi.nl

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.

1.1 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.

1.2 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; Reville & 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, Mielliet, 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 neighbouring 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.

1.3 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; 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 neighbouring word and from the following neighbouring 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.

1.4 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 Sections 3 to 5 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 3 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 4 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 5, 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 6.

2 General methods

2.1 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 Cucchiari, 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.

2.2 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 neighbour. The calculations of these measures are described with the following formulae:

- Relative frequency: $P(W_i) = F(W_i) / N$
- FTP¹: $P(W_i | W_{i-1}) = F(W_{i-1} W_i) / F(W_{i-1})$
- BTP¹: $P(W_i | W_{i+1}) = F(W_i W_{i+1}) / F(W_{i+1})$

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 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.

2.3 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

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

Variable name and description	Type	Transformation
Word duration (in ms)	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
Forward TP (TP with previous word)	predictor	natural log
Backward TP (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
Forward TP	0.0000040	1.00	0.04	0.08
Backward TP	0.0000038	1.00	0.08	0.17

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 6), 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 3 to 5 can be found in Appendix B.

3 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.

3.1 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

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	β	SE	<i>t</i>	<i>p</i> <
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.

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	β	SE	<i>t</i>	<i>p</i> <
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.

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.

3.2 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.

3.3 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.

4 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).

4.1 Results

All β -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

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	β	SE	<i>t</i>	<i>p</i> <
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.

words were more reduced than infrequent words. None of the interactions between age and the probabilistic variables (frequency, and predictability) were significant.

4.2 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.

5 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 3.2) and 64 older adults in the age range of 62 to 95 years (cf. Section 4). 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).

5.1 Results

A summary of the model fitted for the age group comparison, including β -coefficients, standard errors, *t*-values and significance levels, can be found in Table 6. The model included random

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	β	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.

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.

5.2 Discussion

The models we obtained in this age group comparison speak to our research questions in two ways. First, as shown in Sections 3 and 4, 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.

6 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 5 (more experienced older adults versus less experienced adolescents). We used the POS tags provided in the JASMIN annotations to

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.

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.

We fitted separate mixed-effects models for both word classes following the exact same procedure described above (in Section 5 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 neighbouring 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

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	β	SE	<i>t</i>	<i>p</i> <
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	β	SE	<i>t</i>	<i>p</i> <
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.

predictability by age group interactions, such that older adults had smaller predictability effects on word durations than adolescents (as in Section 5). 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.

7 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 neighbouring 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.

7.1 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 behaviour 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, neighbouring 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).

7.2 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 5) 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 5) 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 3). 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 (Rodrigues-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.

7.3 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 behavioural 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, 2015; Bybee, 2006; Gahl, 2008; Jacobs, Yiu, Watson, & Dell, 2015).

8 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.

Funding

This work was supported by a scholarship from the Max Planck International Research Network on Aging (MaxNetAging).

Note

1. 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 transitional probability (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.

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*(2), 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, UK: 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. *Journal of the Acoustical Society of America*, *113*(2), 1001–1024. <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. Livensedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 643–662). Oxford, UK: Oxford University Press.
- Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer, Version 5.3. Retrieved from <http://www.fon.hum.uva.nl/praat/>
- 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. (2015). 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, UK: Oxford University Press.
- Calfee, R., & Drum, P. (1986). Research on teaching reading. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 804–849). New York, NY: 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. *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.
- Cucchiari, 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>
- 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, 66*(5), 843–863. 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., Baker, H. R., 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 Linguistic Society*, *35*, 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](https://doi.org/10.3758/s13428-011-0148-y).
- Hand, C. J., Miell, S., O'Donnell, P. J., & Sereno, S. C. (2010). The frequency–predictability interaction in reading: It depends where you're coming from. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1294–1313. doi: [10.1037/a0020363](https://doi.org/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). Retrieved from http://www.mirjamernestus.nl/Ernestus/publications/Hanique_Ernestus_2011_Interspeech.pdf
- 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). Retrieved from http://www.mirjamernestus.nl/Ernestus/publications/Hanique_Schuppler_Ernestus_2010_Interspeech.pdf
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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, The Netherlands: 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(1/2), 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](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, et al. (Eds.), *Proceedings of the 18th International Congress of Phonetic Sciences [ICPhS 2015]*. London, UK: 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, 1571. 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, The Netherlands: 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.
- Pluymackers, 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.
- Pluymackers, M., Ernestus, M., & Baayen, R. H. (2005b). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*, 118(4), 2561–2569. <http://dx.doi.org/10.1121/1.2011150>.

- 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](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. Retrieved from <http://idiom.ucsd.edu/~rlevy/papers/smith-levy-2011-cogsci.pdf>
- 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*(4), 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*(3), 221–235. doi:10.1177/002383099403700301.
- Wurm, L. H., & Fiscaro, 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.

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	β	SE	<i>t</i>	<i>p</i> <
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	β	SE	<i>t</i>	<i>p</i> <
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 dataset SECTION 3.1–Children.

	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 dataset SECTION 3.2–Adolescents.

	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 dataset SECTION 4–Older adults.

	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 dataset SECTION 5–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.