bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Healthy Aging and Sentence Production:

Disrupted Lexical Access in the Context of Intact Syntactic Planning

Sophie M. Hardy

Katrien Segaert

Linda Wheeldon

University of Birmingham, UK

Univeristy of Agder, Norway

Author Note

Sophie M. Hardy, Katrien Segaert, School of Psychology, University of Birmingham; Katrien Segaert, Centre for Human Brain Heath, University of Birmingham; Linda Wheeldon, Department of Foreign Languages and Translation, University of Agder.

This research was partly supported by an Economic and Social Research Council (ESRC) studentship awarded to Sophie M. Hardy from the University of Birmingham Doctoral Training Centre (grant number: ES/J50001X/1). The authors gratefully acknowledge the help of Denise Clissett, the coordinator of the Patient and Lifespan Cognition Database.

Correspondence concerning this article should be addressed to Sophie M. Hardy, School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK. Email: smh649@bham.ac.uk.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Fluent sentence production requires rapid syntax generation and word retrieval. We investigated how healthy aging affects these processes in two timed picture description tasks. In Experiment 1, young and older adults produced a syntactically related or unrelated prime prior to a target sentence (e.g., *"the bell and the glove move up"*). Both groups displayed significant facilitatory effects of priming on sentence onset latencies. In Experiment 2, participants produced sentences with initial coordinate or simple noun phrases (e.g., *"the owl and the car move above the harp" / "the owl moves above the car and the harp"*). On half the trials, the second picture (car) was previewed; critically, this previewed picture only fell within the initial phrase in the coordinate condition. Without preview, both age groups were slower to initiate sentences with larger coordinate phrases, suggesting a similar phrasal planning scope. However, age group differences did emerge in the preview conditions. Young adults displayed speed benefits of preview both within and outside the initial phrase. Whereas, older adults only displayed speed preview benefits within the initial phrase, and preview outside the initial phrase caused them to become significantly more error-prone. Thus, while syntactic planning scope appears unaffected by age, older adults do appear to encounter problems with managing the activation and integration of lexical items into syntactic structures. Taken together, our findings indicate that healthy aging disrupts the lexical, but not the syntactic, processes involved in sentence generation.

Keywords: aging, sentence production, priming, syntactic planning, lexical retrieval.

Word count: 7,992

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Healthy Aging and Sentence Production:

Disrupted Lexical Access in the Context of Intact Syntactic Planning

The ability to produce and understand language is essential to human well-being (Greene & Burleson, 2003), and a decline in language abilities with age can lead to increased social withdrawal and loneliness (Burke & Shafto, 2008; Mick, Kawachi, & Lin, 2014). Producing a coherent sentence is a complex task involving both the retrieval of lexical items and the generation of an appropriate syntactic structure (Levelt, 1989). We have conducted two novel experiments to investigate age-related changes in lexical retrieval and syntactic planning during sentence production in order to better understand the mechanisms that underlie changes in language ability with old age.

Current research indicates that healthy aging impacts several aspects of language production (for reviews, Abrams & Farrell, 2011; Burke & Shafto, 2008). For example, older adults experience increased tip-of-the-tongue states in which they cannot access the phonology associated with a particular word (e.g., Burke, MacKay, Worthley, & Wade, 1991; Segaert et al., 2018; Shafto, Burke, Stamatakis, Tam, & Tyler, 2007) and are more error-prone in picture naming tasks (see Feyereisen, 1997, for a meta-analytical review). There is also an age-related decrease in the production of complex syntactic structures, such as embedded clauses, coupled with an increase in syntactic errors, such as the use of the incorrect tense (Kemper, 1987; Kemper, Greiner, Marquis, Prenovost, & Mitzner, 2001; Kemper & Sumner, 2001; Rabaglia & Salthouse, 2011). Moreover, older adults are even more error-prone when forced to use more complex syntactic structures in constrained production tasks, such as when incorporating more words into a sentence (Kemper, Herman, & Lian, 2003) or imitating sentences with initial relative clauses (Kemper, 1986). It has been suggested that the language decline experienced by older adults is caused by other emerging

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cognitive deficits, such as a decline in working memory capacity (Kemper & Sumner, 2001), inhibitory control (Hasher & Zacks, 1988) and processing speed (Salthouse, 1993).

Most studies investigating language production and aging have primarily used off-line measures, involving the assessment and coding of sentences after they have been produced; however, such measures do not provide information about the time-course of the underlying sentence generation processes. We have therefore employed on-line sentence production paradigms that measure speech onset latencies, as well as accuracy. Moreover, we measured individual differences in cognitive ability and physical health to investigate whether changes in these functions impact upon sentence generation processes in old age. Individual differences arise from a complex interplay between cognitive systems and the environment: understanding the role of these differences can lead to a better understanding of the mechanisms that underlie language (Kidd, Donnelly, & Christiansen, 2018). In the following introduction, we will first review current models of sentence generation before discussing the two on-line paradigms employed in this study to investigate lexical and syntactic processing during sentence production in healthy aging.

Modelling Sentence Generation Processes

In order to produce a fluent and coherent sentence, several complex processes must be executed quickly and efficiently. A speaker must first form a conceptual representation of the information that they wish to convey – the *message* (Levelt, 1989). This then triggers the formulation stage in which the message is turned into linguistic representations, involving both the rapid retrieval of lexical items and the generation of an appropriate syntactic structure, and these must be integrated correctly to convey the appropriate message. More traditional models of sentence production propose that grammatical encoding is lexically-driven such that lemmas (representations of the syntactic and semantic properties of a word) are first selected and assigned grammatical roles (e.g., subject or object), which then drive the

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

generation of a syntactic structure (Bock & Levelt, 1994; Garrett, 1980; Levelt, Roelofs, & Meyer, 1999; Pickering & Branigan, 1998). Alternatively, computational models postulate that there is a complete dissociation between syntax generation and lexical retrieval, such that syntactic structure is derived solely from conceptual structure (i.e., thematic roles) with lexical access occurring independently (Chang, 2002; Chang, Dell, & Bock, 2006; Chang, Dell, Bock, & Griffin, 2000).

While there remains debate about the exact relationship between syntax generation and lexical retrieval (see Wheeldon, 2011, for a review of the evidence for both lexically-mediated and lexically-independent models), it is widely agreed that sentence production occurs incrementally, such that only a small amount of planning occurs prior to articulation and that planning continues to unfold after speech onset for the remaining sentence (Kempen & Hoenkamp, 1987; Levelt, 1989, 1992). An incremental sentence production system is beneficial as chunks can be rapidly released when planning is complete for that part of the sentence; this enables the processing load to be spread effectively across multiple components and reduces the demand on memory as production is spread across time (Levelt, 1989; Wheeldon, 2013).

Nevertheless, there is less agreement about the scope of the planning that must occur before sentences are initiated. Although Bock and Levelt (1994) proposed that incremental planning occurs in clausal units, evidence indicates that planning likely occurs at a more minimal phrasal level (e.g., Martin, Crowther, Knight, Tamborello, & Yang, 2010; Martin, Yan, & Schnur, 2014; Smith & Wheeldon, 1999; Wheeldon, Ohlson, Ashby, & Gator, 2013) or even at a highly incremental word-by-word level (e.g., Griffin, 2001; Zhao & Yang, 2016). It is possible however that planning scope is not rigidly fixed as it can vary due to multiple factors including ease of syntactic processing (Konopka, 2012; Konopka & Meyer, 2014), task complexity (Ferreira & Swets, 2002; Wagner, Jescheniak, & Schriefers, 2010) and

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cognitive abilities, such as working memory and production speed (Martin, Miller, & Vu, 2004; Slevc, 2011; Swets, Jacovina, & Gerrig, 2014; Wagner et al., 2010). In the present study, we investigated the effect of old age on on-line sentence generation at the syntactic and lexical level of processing.

Investigating On-line Sentence Generation

The amount of time that a speaker takes to begin a sentence is indicative of the amount of planning that has occurred prior to speech onset, in terms of both the retrieval of lexical items and the generation of syntax (Levelt, 1989). Consequently, picture descriptions paradigms and the measurement of speech onset latencies can be used to examine specific sentence generation mechanisms and aspects of incremental planning. Below we review the two paradigms that we have employed, and explain how these can be used to test different hypotheses about age-related changes in sentence planning processes.

Syntactic priming. Syntactic priming refers to the facilitation of syntactic processing that occurs when a syntactic structure is repeated across otherwise unrelated prime and target trials (Bock, 1986; Pickering & Ferreira, 2008). Such facilitation can surface at both the selection and planning level of sentence production. *Choice syntactic priming* is the phenomenon whereby speakers are more likely to produce a dispreferred syntactic alternative, such as a passive sentence, if they have just processed a passive prime compared to an active prime (see Mahowald, James, Futrell, & Gibson, 2016, for a meta-analytical review). *Onset latency syntactic priming* is the facilitated speed of syntactic processing that occurs when a syntactic structure is repeated across a prime and target (Corley & Scheepers, 2002; Segaert, Menenti, Weber, & Hagoort, 2011; Segaert, Weber, Cladder-Micus, & Hagoort, 2014; Segaert, Wheeldon, & Hagoort, 2016; Wheeldon & Smith, 2003). For example, Smith and Wheeldon (2001) demonstrated that when a speaker must produce a given syntactic structure on a target trial (1a), recent production of the structure on a previous

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

trial (1b) increased the speed with which the target was subsequently reproduced, compared to if a different structure (1c) had just been produced.

(1a) Target: "the spoon and the car move up"

(1b) Related prime: "the eye and the fish move apart"

(1c) Unrelated prime: "the eye moves up and the fish moves down"

Further experiments by Smith and Wheeldon (2001) ruled out alternative explanations for the effect relating to visual perception, lexical access, phonological planning and clausal differences (the effect persists when the both the related and unrelated prime feature the same number of clauses as the target). Only a few studies have examined the effect of old age on syntactic priming, all using off-line choice methodologies. While two studies found preserved priming of passives in older adults (Hardy, Messenger, & Maylor, 2017; Heyselaar, Wheeldon, & Segaert, 2017), others have not (Heyselaar, Segaert, Walvoort, Kessels, & Hagoort, 2017, footnote 2; Sung, 2015).[1] To date, no study has investigated age effects in onset latency syntactic priming; this is a noticeable gap as a complete model of sentence generation must account for age effects on both how a syntactic structure is chosen, as well the processes by which it is planned and produced.

Most models have primarily focused on just explaining choice syntactic priming by use of residual activation (Pickering & Branigan, 1998) or implicit learning mechanisms (Chang et al., 2006). However, Segaert et al. (2016) proposed a two-stage competition model that explains the effect of syntactic priming on both choices and onset latencies (see also Segaert et al., 2011, 2014). According to the model, alternative syntactic structures (e.g., active vs. passive) are represented by nodes with competing activation levels, and there are two sequential stages of production: first selection of one of the syntactic alternatives; and second planning, during which the selected syntax is incrementally planned and produced. While syntactic choice is determined solely at the selection stage, production speed is

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

determined by the time taken to complete both stages. Consequently, when the choice element is removed (as in Smith & Wheeldon, 2001), onset latencies are largely determined by processing at the planning stage with very minimal processing required at the selection stage as there are no competing syntactic alternatives. In this study we therefore investigated the effect of age on onset latency syntactic priming effects. We removed the choice element as this allowed us to tap more directly in the processes involved in sentence planning. The magnitude of the onset latency syntactic priming effects observed in the older adults will be informative about age-related changes in syntactic planning and facilitation that occur during real-time sentence production.

Planning scope. The second paradigm that we employed specifically probes incremental sentence production and can therefore provide insight into age-related changes in the integration of syntactic and lexical information. In the *planning scope* paradigm, picture displays are used to elicit sentences with different syntactic structures and speech onset latencies are used as a measure of the amount of advanced planning that has occurred prior to articulation (see Wheeldon, 2013, for a review). For example, Smith and Wheeldon (1999) found that participants took longer to produce sentences with larger initial coordinate phrases (2a) compared to smaller initial simple phrases (2b), suggesting that planning scope occurs in phrasal units: when the first phrase is larger, speakers need longer to plan the syntax and retrieve the second lexical item before speech onset (see also Levelt & Maassen, 1981; Martin et al., 2004; Wheeldon et al., 2013).

(2a) "[the dog and the hat move] above the fork"

(2b) "[the dog moves] above the hat and the fork"

Martin et al. (2010, 2014) ruled out an alternative explanation for this effect relating to the visual array (i.e., the grouping of objects moving together) as they found the same phrasal planning scope using stationary pictures arrays (e.g., *"the drum and the package are*

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

below the squirrel"). Moreover, the phrasal planning effect cannot be attributed to the fact that in English the second content word in the simple initial phrase (always the verb 'moves'; 2b) may be easier to retrieve than in the coordinate initial phrase (always the second lexical item; 2a) as the effect has been demonstrated in Japanese, a verb-final language in which the subject and the complement take the first two positions in the a sentence regardless of initial phrase type (Allum & Wheeldon, 2007, 2009).

The paradigm is therefore a useful tool for investigating incremental planning scope in older adults. Given the proposed importance of working memory to sentence planning (Slevc, 2011; Swets et al., 2014), it may be that older adults perform differently to young adults. Indeed, Martin et al. (2004) found that an aphasic patient with a semantic working memory deficit displayed a greater phrasal complexity effect than controls (i.e., a markedly greater difference in the speed of production of larger, compared to smaller, initial phrases); the authors attributed this to the patient attempting to plan both nouns in the initial phrase, but having difficulty doing so because of deficits at the lexical-semantic level. To our knowledge, no study to date has directly examined the effect of old age on incremental planning scope.[2] It is possible that due to emerging deficits in working memory, older adults will display a larger phrasal complexity effect than young adults (although this is unlikely be as exaggerated as that seen in aphasic patients). Alternatively, older adults may unconsciously choose to engage in a more extreme word-by-word incremental strategy. Ferreira and Swets (2002) found that when time pressure was applied, speakers engaged in considerably less advanced planning, suggesting that incremental planning can be strategically controlled by the speaker. Therefore, older adults may adopt a more minimal sentence planning strategy than young adults in order to maintain an acceptable speed of speech output.

In order to directly investigate lexical processing during sentence production, some planning scope studies have also included a picture preview element. Wheeldon et al. (2013)

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

required participants to produce sentences similar to (2a) and (2b), but on some trials there was a preview of one of the upcoming pictures. They found that previewing the second to-be-produced lexical item (*hat* for the examples shown in 2) decreased onset latencies more when it fell within, rather than outside of, the initial phrase (see Allum & Wheeldon, 2009, for a similar study in Japanese using stationary visual displays). This suggests that the retrieval of lexical items within the first phrase is prioritised prior to speech onset. Nevertheless, the preview benefit was not reliably maintained when the phrase consisted of three nouns and participants previewed the third lexical item (*"[the drum, the star and the* ***hat*** *move] above the crab"*). Thus, it appears that advanced lexical planning only encompasses a subset of the required nouns. Wheeldon et al. (2013) speculated that this is because attempting to retrieve and hold all three nouns within an initial phrase prior to articulation may lead to problems with buffering and maintaining a linearized output. For young adults, their preferred scope of lexical encoding appeared to be two lexical items; however, given that older adults are known to have a reduced memory buffer for holding linguistic information (e.g., word and sentence span; Waters & Caplan, 2003) their preferred limit may be even less. In particular, if older speakers typically only encode the first lexical item within a phrase prior to articulation, then, unlike young adults, they may not display the preview benefit of the second lexical item within a larger initial phrase.

The Present Study

We report two experiments designed to measure age-related changes in sentence planning using on-line measures of sentence production. In Experiment 1, young and older adults completed an onset latency syntactic priming task (similar to Smith & Wheeldon, 2001) to test for age-related changes in syntactic facilitation effects. In Experiment 2, participants completed a planning scope task with a picture preview element (similar to Wheeldon et al., 2013) to test for age-related changes in syntactic and lexical planning scope.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In addition, participants completed seven cognitive and physical measures that are used as markers of the 'healthy aging phenotype' (Lara et al., 2013), and have been related to individual differences in the magnitude of age-related language decline. We aimed to relate participants' scores on the aging markers to individual differences in performance in the sentence production tasks.

Method

Participants

The same participants completed Experiments 1 and 2. We recruited 50 young adults (36 female) aged 18-25 from the University of Birmingham student population (compensated with course credits) and 56 older adults (37 female) aged 64-80 from the Patient and Lifespan Cognition Database (compensated monetarily). All older adults scored above 26 out of 30 (*M* = 27.4; *SD* = 1.3) on the Montreal Cognitive Assessment (Nasreddine et al., 2005), indicating that they were currently experiencing healthy aging (scores < 26 indicate significant risk of mild cognitive impairment or dementia; Smith, Gildeh, & Holmes, 2007). All participants were native English speakers with normal or corrected-to-normal vision, and did not report any language disorders. There was no significant difference in education between age groups (Table 1). The study was approved by the University of Birmingham Ethical Review Committee and all participants provided written informed consent.

Experiment 1: Syntactic Priming Task

Design. We used a 2 X 2 mixed design with one between-participant variable of age (young vs. old) and one within-participant variable of prime type (syntactically related vs. syntactically unrelated). Hence, there were two experimental conditions (Figure 1A).

Materials. To create the experimental items, we used 80 simple photographic pictures of everyday concrete objects. All picture names were mono- or disyllabic, and care was taken

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to ensure that the objects could be identified and named quickly and easily. Forty of the pictures were used to create the 40 picture pairs for the target trials; each picture appeared in two different pairs (once each in the left and right position). Using the same constraints, we constructed 40 picture pairs from another 40 pictures for the prime trials. We then paired each target pair with a prime pair to generate 40 experimental items. We ensured that there was no phonological or conceptual overlap between any of the four pictures within each experimental item.

The movement of each picture pair was controlled using E-Prime (Schneider, Eschman, & Zuccolotto, 2002). In all target trials, both pictures moved in the same vertical direction (either up or down). Participants were instructed to describe the picture movements from left to right using specific sentences; hence, the target trials elicited a coordinate noun phrase (*"the A and the B move up/down"*). In the *related* prime condition, the pictures moved in opposing horizontal directions which elicited a sentence that was syntactically related to the target trials (*"the C and the D move together/apart"*). In the *unrelated* prime condition, the pictures moved in opposing vertical directions which elicited a sentence that was syntactically unrelated to the target trials (*"the C moves up/down and the D moves down/up"*). We then created two item lists that each contained the same 40 target sentences, but the prime condition matched to each target was rotated such that there were 20 related and 20 unrelated primes per list. Each participant was randomly assigned to one of the two lists and therefore completed 20 experimental items (prime plus target pairs) from each condition (Table 2A).

Lastly, we used a further 54 pictures to construct 120 filler trials designed to increase the variety of syntactic structures produced by the participant and minimise the risk of them noticing the priming manipulation. We created 96 filler trials that elicited phrases such as: *"there is an X and a Y"* (no picture movement); *"the Xs move up"* (two repeat pictures move

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

simultaneously) and *"there are no pictures"* (screen is blank). We also created 24 filler trials that elicited phrases that were syntactically similar to the experimental trials; without such 'decoy' fillers, experimental trials would always occur in pairs (i.e., prime and corresponding target) which may enable the participant to predict the upcoming movement of a target trial. All 120 fillers were added to each of the two items lists. We then divided each list into four blocks that each contained 5 *related* experimental items, 5 *unrelated* experimental items and 30 filler items. The distribution of items within each block was pseudorandomized with the constraint that two experimental items never occurred consecutively. The ordering of the blocks was rotated across participants.

Procedure. Each participant was tested individually in a sound-attenuating booth facing the screen of a 17 inch *Dell* monitor, in front of which was a *Sony* microphone connected to an amplitude voice key that recorded his/her responses and onset latencies. Figure 1B illustrates the sequence of stimuli presentation per trial. To begin, there were 50 practice trials; the sentences elicited resembled those in the experimental and filler trials and featured all 80 experimental pictures once. The task then continued until all four blocks had been completed. The experimenter listened from outside the booth via headphones and noted down any errors made by the participant. Errors included: incorrect picture naming (e.g., 'fish' instead of 'shark'); use of a difference sentence structure (e.g., *"the pig moves towards the leaf"* instead of *"the pig and leaf move together"*); and disfluencies, such as stuttering and pausing.

Experiment 2: Planning Scope Task

Design. We used a 2 X 2 X 2 mixed design with one between-participant variable of age (young vs. older) and two within-participant variables of preview (no preview vs. preview) and initial phrase type (coordinate vs. simple). Hence, there were four experimental conditions (Figure 2A). Critically, the previewed picture (always of the second upcoming

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

lexical item) fell within the initial phrase in the coordinate condition, but outside of the initial phrase in the simple condition.

Materials. To create the experimental items, we used 80 photographic pictures of everyday concrete objects (these were different to those used in Experiment 1, but meet the criteria). We created 80 experimental items that each consisted of three different pictures that were conceptually and phonologically distinct: each of the 80 pictures appeared in three different experimental items (once in the left, central and right position). The sentence descriptions of the items were again elicited by controlling the movement of the pictures (using E-prime) and participants were instructed to describe the picture movements from left to right using specific sentences. In the *simple* conditions, only the left picture moved (either up or down) and the other two pictures remained stationary (*"the A moves above/below the B and the C"*). In the *coordinate* conditions, both the left and the central picture moved simultaneously (either up or down) and only the right picture remained stationary (*"the A and the B move above/below the C"*). In the preview trials, the preview was always of the central upcoming picture (i.e., object *B*). We created four item lists by evenly rotated the experimental condition assigned to each of the 80 experimental items. Each participant was randomly assigned to one of the four lists and therefore completed 20 experimental items per condition (Table 2B).

Lastly, we used a further 106 pictures to create 220 filler items designed to prevent the participant from anticipating the location of the preview picture and building expectations to guide their response. The fillers elicited some experimental-type sentences and other sentences that differed from the experimental items in terms of the number of pictures and the type of movement, such as: *"there is an X, a Y and a Z"* (no picture movement); *"the Xs move up"* (three repeat picture move simultaneously); and "*there are no pictures"*. Importantly, we also varied the position of the preview pictures within the fillers, such that

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

across all the experimental and filler items each screen position was previewed an equal number of times. All 220 filler items were added to each of the four item lists. We then divided each list into five blocks that each contained 44 fillers and 16 experimental items (4 per condition), and pseudorandomized the order of items using the same constraints as Experiment 1. The ordering of the blocks was rotated across participants.

Procedure. Each participant was tested using the same equipment set-up described in Experiment 1. Figure 2B illustrates the sequence of stimuli presentation per trial. To begin, there were 40 practice trials; the sentences elicited resembled those in the experimental and filler trials and featured all 80 experimental pictures once. The task then continued until all five blocks had been completed. Using the same criteria described in Experiment 1, the experimenter noted down any errors made by the participant.

Aging Markers

Each participant completed eight additional measures, designed to provide an indicator of how healthily he/she was aging across cognitive, physical and physiological domains (Lara et al., 2013). Details of these measurements for each aging marker are summarised in Table 3.

Data Preparation and Analyses

We excluded the data of participants whose error rates were above 50% on the experimental trials. This resulted in exclusion of five older adults in Experiment 1, and one older adult in Experiment 2.[3] Of the 4040 target responses in Experiment 1, we excluded trials in which the participant made an error on the corresponding prime (170 (8.5%) of young and 301 (14.7%) of older adult trials), and for which the target onset latency was below 300 ms, above 3000 ms or more than 2.5SD above/below the participants' mean per experimental condition (discarding 53 (2.9%) young and 49 (2.8%) older adult trials). We applied to same onset latency exclusion criteria to the 8400 experimental trials in Experiment

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2 (discarding of 124 (3.1%) young and 166 (3.8%) older adult trials). For both experiments, all remaining trials were used in the error analyses, but only correct responses (87.4% and 81.7% of trials in Experiments 1 and 2) were used in onset latency analyses.

The data from Experiments 1 and 2 were analysed separately, but following the same method. All data were analysed in R (R Core Team, 2015) using mixed-effects models (*lme4* package; Bates, Mächler, Bolker, & Walker, 2014); this was the most suitable way to analyse the datasets as there were repeated observations for participants and items (Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013; Jaeger, 2008). We fitted a logit mixed-effects model to the error data as the dependent variable was categorical (correct = 0, incorrect = 1), and a linear mixed-effects model to the onset latency data as the dependent variable was continuous. We used a maximal random effects structure as this allowed us to include per-participant and per-item adjustments to the fixed interprets (random intercepts) with additional random adjustments to the fixed effects (random slopes). For the Experiment 1 models, we entered age group (young vs. old) and prime type (related vs. unrelated) as fixed effects. For Experiment 2, we entered age group (young vs. older), initial phrase type (coordinate vs. simple) and preview type (no preview vs. preview) as fixed effects. In all models, we included random intercepts for participants and items, as well as by-participant random slopes for within-participant fixed effects and by-item random slopes for within-item fixed effects.

Prior to analysis, the fixed effects were sum-coded and transformed to have a mean of 0 and a range of 1. When a model did not converge with the maximal random effects structure, we simplified the random slopes, removing interactions before main effects in order of least variance explained until the model converged (Barr et al., 2013). Significance *p* values for the linear mixed-effects model were calculated using the *car* package (Fox & Weisberg, 2011). We modelled the complete data sets for both experiments as well as for the

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

young and older adult data separately as we had *a priori* hypotheses about age group differences.

Analysis of the aging markers. In order to understand the influence of individual differences on the participants' experimental task behaviour, we also entered the following aging markers into all mixed-effects models as continuous predictors: processing speed; short-term memory; long-term memory; vocabulary; inhibition; and lung capacity. We did not enter handgrip and working memory as we did not find the typical age-related declines in these measures (Table 1); hence, we considered that including them would not be informative about the effect of age-related changes in these domains on the experiment performance.[4] Before entering the predictors into the model, we converted the raw scores into age-scaled scores to enable us to compare group scores from different normal distributions within the same model (Howell, 2010): we converted raw scores into standardised z-scores within age groups (and gender group for lung capacity) for all predictors except processing speed, for which we used the existing age-adjusted scores in the WAIS manual. The Variance Inflation Factor (VIF), a measure of the size of correlations between different predictors, of all models was < 2.8 indicating that there was limited multicollinearity between predictors (VIF values < 5 are acceptable; Jaeger, 2011).

We began with all aging markers as continuous predictors in the model (centred prior to analysis), and then simplified the model using a stepwise "best path" reduction procedure, removing predictor interactions and then predictor main effects to locate the simplest model that did not differ significantly from the full model in terms of variance explained (Barr et al., 2013). To do this we used the *drop1* function of the *lme4* package that compares the Akaike's Information Criterion (AIC) values of the full model to a model with one interaction or main effect removed (see Schoot, Heyselaar, Hagoort, & Segaert, 2016, for a similar analysis).

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Thus, we reached the simplest model that was best able to explain the data for each experiment.

Results

Experiment 1: Examining the Effect of Aging on On-line Syntactic Priming

Figure 3 summarises the target error rates and the onset latencies across the two prime conditions for young and older adults.

Error rates. The best-fitting model of the error data is reported in Table 4A. Although older adults were significantly more error-prone than young adults (9.1% vs. 16.1%, $p <$ .001), there was no main effect of prime type (p = .432) and no interaction between age group and prime type (p = .487). Further modelling of the young and older adult data separately confirmed that prime type did not significantly affect error rates in either age group (ps > .6; Tables 4B and 4C).

Onset latencies. The best-fitting model of the onset latency data is reported in Table 5A. As expected, older adults were significantly slower than young adults (898 ms vs. 1060 ms, $p < .001$).[5] There was a main effect of prime type ($p < .001$), such that target responses were produced significantly quicker following related primes (953 ms) than following unrelated primes (994 ms), indicating an overall syntactic priming effect of 41 ms. Most interestingly, there was no interaction between age group and prime type (p = .635). Although null-effects should always be interpreted with caution (Altman & Bland, 1995), this suggests that young and older adults were experiencing similar onset latency priming. Indeed, the priming effect remained significant in both young (36 ms, 3.9% benefit, p = .002) and older (49 ms, 4.5% benefit, p = .004) adults when analysed separately (Tables 5B and 5C).

Aging markers. The model of the error data (Table 4) contained a significant interaction between prime type, age group and inhibition (p = .029); this indicates that

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

participants' inhibitory control may have affected their error rates in the prime conditions, but that this was different for young and older adults. The interaction between prime type and inhibition did not maintain significance when the young adults were considered separately (p = .276), but did for the older adults (p = .033). One way to understand this interaction is to perform a median split of participants' inhibition scores: this revealed that older adults with poorer inhibition produced more errors in the unrelated than the related prime condition (17.9% vs. 13.1%), but those with better inhibitory control were more error-prone in the related prime condition (15.5% vs. 18.0%). Moving to the onset latency data, the model (Table 5) included a main effect of long-term memory (p = .045): a median split revealed that participants with better long-term memory skills were quicker overall than those with poorer skills (951 ms vs. 990 ms). However, this effect was no longer significant when the age groups were analysed separately (ps < .09).

Experiment 2: Examining the Effects of Aging on On-line Planning Scope

Figure 4 summarises the error rates and the onset latencies across the four experimental conditions for young and older adults.

Error rates. The best-fitting model of the error data is reported in Table 6A. As in Experiment 1, older adults were significantly more error-prone than young adults (12.5% vs. 23.5%, p < .001). While there were no main effects of preview (p = .308) or initial phrase type (p = .098), there was a significant interaction between the two variables (p = .040): the presence of the preview resulted in a 1.6% decrease in participants' errors when producing sentences with initial coordinate phrases, but a 2.9% increase in errors when producing sentence with initial simple phrases. Although the interaction between preview, initial phrase type and age group did not reach significance (p = .285), modelling of the age groups separately indicated the experimental conditions may have had a greater effect on older adults' production of errors (Tables 6B and 6C). There were no main effects or interactions in

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the young adult analysis (all *p*s < .3), indicating that their error rates were fairly stable across conditions. However, the older adult analysis revealed a significant interaction between preview and initial phrase type ($p = .016$): the presence of the preview resulted in an opposing patterns of errors when it fell within and outside of the initial phrase (as can be clearly seen in Figure 4). Further investigation revealed that the preview caused a significant 5.3% increase in older adults' errors in the simple condition ($\chi^2(1) = 7.08$, $p = .016$), but a 1.8% decrease in errors in the coordinate condition, although this difference was not statistically significant ($\chi^2(1) = 0.53$, $p = .468$).[6]

Onset latencies. The best-fitting model of the onset latency data is reported in Table 7A. As in Experiment 1, older adults were significantly slower than young adults (843 ms vs. 991 ms, $p < .001$). There was a main effect of initial phrase type, such that sentences with initial simple phrases were produced significantly quicker than sentences with initial coordinate phrases (895 ms vs. 935 ms, $p < .001$), indicating an overall phrasal planning effect of 40 ms. Furthermore, the interaction between initial phrase type and age group was not significant (p =. 991), indicating that the incremental planning effect was similar for both young (40 ms, 4.6% benefit, $p < .001$) and older (41 ms, 4.0% benefit, $p < .001$) adults, as was confirmed by the modelling of the age group separately (Tables 7B and 7C).

The model of the complete data set also revealed a main effect of preview (Table 7A), such that sentences were produced significantly quicker following preview of the second upcoming lexical item compared to no preview (940 ms vs. 980 ms, $p < .001$). Moreover, there was a significant interaction between preview and initial phrase type ($p < .001$): the overall preview benefit was significantly greater when the preview picture feel within the initial phrase (coordinate condition; 74 ms, $\chi^2(1) = 57.13$, $p < .001$) compared to outside of it, although this benefit was still significant (simple condition; 26 ms, $\chi^2(1) = 7.53$, $p = .006$). While the interaction between age group, preview and initial phrase type did not reach

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

significance (p = .250), the separate age group analyses did reveal some potentially interesting differences (Tables 7B and 7C). The interaction between preview and initial phrase type maintained significant for young (p = .011) and older adults ($p < .001$); however, inspection of Figure 4 suggests that the interaction may be representing something different for the two age groups. For young adults, while the preview benefit was greater when it fell within the initial phrase, the effect of preview was still significant in both the coordinate (81 ms (8.9%), $\chi^2(1) = 36.02$, $p < .001$) and simple (45 ms (5.3%), $\chi^2(1) = 12.90$, $p < .001$) conditions. By contrast, the difference in onset latencies between preview conditions was only significant for the older adults when it fell within the initial phrase (67 ms (6.4%) preview benefit; $\chi^2(1) = 20.57$, $p < .001$), but not outside of it (2 ms (0.2%) preview benefit; $\chi^2(1) = 0.43$, $p = .513$).

Aging markers. The model of the error data did not include any aging predictors (Table 6); however, the onset latency model did include the predictor of short-term memory (Table 7). The young adult analysis revealed a significant interaction between preview and short-term memory (p = .025): a median split indicated that those who scored above average on the short-term memory task displayed a greater preview effect overall than those who scored below average (93 ms vs. 36 ms). The older adult analysis revealed a significant 3-way interaction between preview, initial phrase type and short-term memory (p = .045). Closer inspection of the data would suggest that older adults with above-average short-term memory displayed a larger preview benefit within the initial phrase than those with below-average short-term memory (97 ms vs. 25 ms).

Discussion

Using two on-line experiments, we investigated age-related changes in the syntactic and lexical processes involved in sentence generation. In Experiment 1, both young and older

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

adults produced target sentences quicker following syntactically-related primes, demonstrating that speed benefits of syntactic priming are preserved with age. In Experiment 2, both young and older adults initiated sentences quicker with smaller, compared to larger, initial phrases, suggesting that planning scope, at least at the syntactic level, is unaffected by healthy aging. Age differences did emerge, however, in the preview conditions; whereas young adults displayed speed benefits of picture preview when the pictured word fell both within and outside the initial phrase, older adults only displayed speed benefits from the previewed picture when it fell within the initial phrase, and preview outside of the initial phrase caused them to become more error-prone. This suggests age differences in the flexibility of lexical retrieval during sentence planning and the ability to integrate lexical information into syntactic structures. Taking both experiments together, our study therefore suggests age effects of lexical, but not syntactic, processes on the speed and accuracy of sentence production.

Our robust finding of preserved effects of syntactic priming on onset latencies in older adults is in line with previous research that has found no evidence for age effects on choice syntactic priming (Hardy et al., 2017; Heyselaar, Wheeldon, et al., 2017). Syntactic priming therefore facilitates both older adults' syntactic choices and speed of sentence planning. This is consistent with evidence that older adults maintain the benefit from priming in other areas of language processing. For example, morphological priming effects for regularly-inflected verbs are preserved with age (Clahsen & Reifegerste, 2017; Reifegerste, Elin, & Clahsen, 2018), as well as for transparent compounds (Duñabeitia, Marín, Avilés, Perea, & Carreiras, 2009). Critically, older adults did not show a greater magnitude of facilitation than younger adults, despite being significantly slower and more error-prone. This stability in the syntactic priming effect with age is consistent with age-related preservation of the implicit learning and/or residual activation processes underlying syntactic priming (Chang et al., 2006;

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Pickering & Branigan, 1998). Whichever the underlying mechanisms, our experiment provides the first evidence that onset latency priming effects are preserved with age in a task specifically designed to tap into syntactic planning processes.

In Experiment 2, the pattern of onset latencies observed is similarly consistent with an age-related preservation of syntactic planning skills as we found robust evidence of a phrasal scope of planning in both age groups (i.e., speakers took longer to initiate sentences with larger initial phrases). This finding replicates previous research in young adults (e.g., Martin et al., 2010, 2014; Smith & Wheeldon, 1999), and suggests that both age groups prioritised the generation of syntax within the first phrase prior to articulation. It is notable that older speakers did not experience disproportionate difficulty in planning the larger initial phrases (as has been observed in aphasia patients; Martin et al., 2004) or engage in a more extreme word-by-word planning strategy (if this was the case, onset latencies would have been similar for simple and coordinate initial phrases). This indicates that older adults still maintain sufficient cognitive capacity to support the planning of an initial phrase containing at least two nouns. Notably, evidence of intact syntactic processing in older adults has been found in other studies in which lexical items are also presented on screen and participants are asked to formulate the sentence (Altmann & Kemper, 2006; Davidson, Zacks, & Ferreira, 2003); however, in these studies the lexical items are presented in written word form meaning that the speaker does not need to independently encode the name of lexical item prior to sentence production. Our study, in which the participant is presented with photographic images of the lexical items, is therefore the first to demonstrate preserved syntactic planning when the speaker must independently access the lexical properties of all words, as well as generate the syntactic structure.

Nevertheless, it remains unclear whether older adults would continue to engage with a phrasal planning scope when producing a sentence containing a more complex initial phrase.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Wheeldon et al. (2013) found that young adults continued to plan at a phrasal level when the initial phrase consisted of three nouns. However, it is possible that three or more nouns may disrupt older adults' sentence planning as age-related effects did emerge due to the picture preview manipulation, suggesting that lexical planning mechanisms may be disrupted by old age. In Experiment 2, half the experimental trials were preceded by a picture of the upcoming second lexical item. When the previewed picture fell within the initial phrase (*"[the owl and the **car** move] above the harp"*), both young and older adults were quicker to initiate the sentence compared to when there was no preview, suggesting that the prior retrieval of the lexical item was significantly benefiting their sentence planning at the lexical encoding level (Allum & Wheeldon, 2009; Wheeldon et al., 2013). However, some interesting group differences did emerge in young and older adults' onset latencies and error rates when the previewed picture fell outside of the initial phrase (*"[the owl moves] above the **car** and the harp"*). Young adults continued to display speed benefits of preview outside the initial phrase, albeit to a lesser extent than when it fell within the phrase. This demonstrates that the young adults prioritised the retrieval of lexical items within the first phrase prior to speech onset, but they were also able to successfully manage the early activation of lexical items outside of their usual phrasal planning scope to benefit the overall speed of their sentence production. This evidence of adaptability within young adults' planning scope adds to the growing evidence that planning scope is flexible and can be influenced by the ease of syntactic and lexical processing (Ganushchak & Chen, 2016; Konopka, 2012; Konopka & Meyer, 2014; van de Velde & Meyer, 2014).

In contrast, older adults did not display any speed benefits of preview outside of the initial phrase, and the presence of the picture preview outside their preferred phrasal planning scope caused them to become significantly more error-prone. The onset latency and error data together therefore suggest that, unlike the young adults, the older adults did not benefit from

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

premature access to lexical information and that instead premature availability had a disruptive effect on their overall fluency. One explanation for this age difference is that older adults' planning scope may be more rigidly fixed to phrasal boundaries and so they are less adaptable when it comes to integrating new lexical information into syntactic structures. Indeed, older adults show less parafoveal preview effects across syntactic pauses than young adults during sentence comprehension, suggesting an age-related segmentation strategy designed to aid syntactic processing (Payne & Stine-Morrow, 2012, 2014; Stine-Morrow & Payne, 2016). This segmentation strategy may also apply to older speakers' sentence production, meaning that they are less able to successfully incorporate lexical information outside of the initial phrase into their sentence planning.

A second explanation for this age difference in lexical processing relates to the executive control required to successfully manage the premature access to lexical information. When there was a preview picture, the participant would have automatically accessed some lexical information about the item and this would then have been stored in their working memory. However, when the preview picture did not appear within the first phrase, the participant would have needed to temporarily inhibit this information to order to ensure that it did not interfere with their initial sentence planning (i.e., the retrieval of the first lexical item that was not previewed). Theoretical accounts propose that aging weakens the inhibitory processes that are responsible for regulating what information enters and leaves working memory, and a consequence of age-related declines in inhibition is increased interference effects (Hasher, Lustig, & Zacks, 2007; Hasher & Zacks, 1988). Indeed, deficits in inhibitory control have been used to explain other age effects of sentence processing, such as older adults experiencing increased difficulty ignoring visually-distracting information while reading (Connelly, Hasher, & Zacks, 1991), and naming pictures in the presence of semantically-related distractors (Taylor & Burke, 2002). Deficits in inhibitory control may

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

also therefore provide a valid explanation of our findings as, if the older adults were less able to regulate the storage of information relating to the preview picture during sentence planning, this would have led to increased problems with buffering and maintaining a linearized output, resulting in increased errors. Nevertheless, further work is needed to fully understand the effect of inhibition on age-related changes in syntactic and lexical processing, particularly as to what extend activation of different lexical items causes interference during sentence production.

Lastly, to consider the individual variation in performance on our experimental tasks as guided by the aging markers we used to measure variability in cognitive, physical and physiological aspects of healthy aging. We found some evidence for the role of short-term memory (measured using an immediate word recall task) on the preview effect in Experiment 2, such that participants with above-average short-term memory displayed a larger preview benefit. This is consistent with an influence of short-term memory on language processing (Baddeley, 2003; MacDonald, 2016); specifically, speakers with greater short-term memory capacity can maintain more lexical items prior to articulation and are therefore better able to benefit from the preview. In Experiment 1, we also found evidence that older adults with poorer inhibitory control (measured using a stop-signal task) produced more errors on the target trials following syntactically-unrelated primes. This may relate to the participant having to inhibit the unrelated prime syntax in order to correctly generate the target syntactic structure: those with poorer inhibition may have been less able to do this causing them to become more error-prone. Beyond these two findings however, we did not find the aging markers to have a great deal of influence on language task performance. This may be in part due to the fact that we did not use of comprehensive battery of measures for each aging marker, and that measuring individual differences within a factorial experimental design is acknowledged to be inherently difficult (Hedge, Powell, & Sumner, 2017; Kidd et al., 2018).

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Nevertheless, despite the challenges, it is important to continue to integrate measures of individual variability into experimental methods as the association between linguistic performance and cognitive factors can uncover unknown properties about underlying language mechanisms (Ehrman, Leaver, & Oxford, 2003; Kidd et al., 2018; Swets, 2015).

In summary, our study is the first to specifically examine on-line sentence production in older adults; our findings should therefore be considered in parallel with off-line studies of language and aging in order to gain a more complete understanding of the effect of old age on language processing. Specifically, our study provides evidence for the age-related preservation of syntactic processing during on-line sentence production, but an increased difficulty with lexical retrieval and management with age. We attribute this apparent age-related decline in lexical processing to a decline in the flexibility of sentence planning processes. This may be related to older speakers' stronger preference for segmentation at phrasal boundaries when planning a sentence and/or to declines in executive control that mean older speakers are less able to cope with premature lexical activation beyond the first phrase.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Abrams, L., & Farrell, M. T. (2011). Language processing in normal aging. In J. Guendouzi, F. Loncke, & M. . J. Williams (Eds.), *The handbook of psycholinguistic and cognitive processes: Perspectives in communication disorders* (pp. 49–74). New York: Psychology Press. https://doi.org/10.4324/9780203848005.ch3

Allum, P. H., & Wheeldon, L. (2007). Planning scope in spoken sentence production: The role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 791–810. https://doi.org/10.1037/0278-7393.33.4.791

Allum, P. H., & Wheeldon, L. (2009). Scope of lexical access in spoken sentence production: Implications for the conceptual–syntactic interface. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1240–1255. https://doi.org/10.1037/a0016367

Altman, D. G., & Bland, J. M. (1995). Absence of evidence is not evidence of absence. *British Medical Journal*, *311*, 485. https://doi.org/10.1136/bmj.311.7003.485

Altmann, L. J. P., & Kemper, S. (2006). Effects of age, animacy and activation order on sentence production. *Language and Cognitive Processes*, *21*(1–3), 322–354. https://doi.org/10.1080/0169096054400006

Baayen, H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

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*,

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

68(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

Bock, K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word-finding failures in young and older adults. *Journal of Memory and Language*, *30*, 542–579. https://doi.org/https://doi.org/10.1016/0749-596X(91)90026-G

Burke, D. M., & Shafto, M. A. (2008). Language and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 373–443). New York: Psychology Press. https://doi.org/10.1080/03610738508259280

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, *26*(5), 609–651. https://doi.org/10.1016/S0364-0213(02)00079-4

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, *29*(2), 217–229. https://doi.org/10.1023/A:1005101313330

Cho-Reyes, S., Mack, J. E., & Thompson, C. K. (2016). Grammatical encoding and learning in agrammatic aphasia: Evidence from structural priming. *Journal of Memory and Language*, *91*, 202–218. https://doi.org/10.1016/j.jml.2016.02.004

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Clahsen, H., & Reifegerste, J. (2017). Morphological processing in old-age bilinguals. In M. Libben, M. Goral, & G. Libben (Eds.), *Bilingualism: A framework for understanding the mental lexicon* (pp. 217–248). Amsterdam: John Benjamins Publishing Company.

Connelly, S. L., Hasher, L., & Zacks, R. T. (1991). Age and reading: The impact of distraction. *Psychology and Aging*, *6*(4), 533–541. https://doi.org/10.1037//0882-7974.6.4.533

Cooper, R., Hardy, R., Sayer, A. A., Ben-Shlomo, Y., Birnie, K., Cooper, C., … Kuh, D. (2011). Age and gender differences in physical capability levels from mid-life onwards: The harmonisation and meta-analysis of data from eight UK cohort studies. *PLos ONE*, *6*(11), e27899. https://doi.org/10.1371/journal.pone.0027899

Corley, M., & Scheepers, C. (2002). Syntactic priming in English sentence production: Categorical and latency evidence from an internet-based study. *Psychonomic Bulletin & Review*, *9*(1), 126–131. https://doi.org/10.3758/BF03196267

Davidson, D. J., Zacks, R. T., & Ferreira, F. (2003). Age preservation of the syntactic processor in production. *Journal of Psycholinguistic Research*, *32*(5), 541–566. https://doi.org/10.1023/A:1025402517111

de Rosario-Martinez, H. (2015). Analysing interactions of fitted models. *CiteSeer*, 1–30. https://doi.org/10.1007/s13398-014-0173-7.2

Duñabeitia, J. A., Marín, A., Avilés, A., Perea, M., & Carreiras, M. (2009). Constituent priming effects: Evidence for preserved morphological processing in healthy old readers. *European Journal of Cognitive Psychology*, *21*(2–3), 283–302. https://doi.org/10.1080/09541440802281142

Ehrman, M. E., Leaver, B. Lou, & Oxford, R. L. (2003). A brief overview of individual differences in second language learning. *System*, *31*(3), 313–330. https://doi.org/10.1016/S0346-251X(03)00045-9

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*(1), 57–84. https://doi.org/10.1006/jmla.2001.2797

Ferreira, V. S., Bock, K., Wilson, M. P., & Cohen, N. J. (2008). Memory for syntax despite memory amnesia. *Psychological Science*, *19*(9), 940–946. https://doi.org/10.1111/j.1467-9280.2008.02180.x.Memory

Feyereisen, P. (1997). A meta-analytic procedure shows an age-related decline in picture naming: Comments on Goulet, Ska, and Kahn (1994). *Journal of Speech, Language, and Hearing Research*, *40*(6), 1328–1333. https://doi.org/10.1044/jslhr.4006.1328

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression*. Thousand Oaks, CA: Sage.

Ganushchak, L. Y., & Chen, Y. (2016). Incrementality in planning of speech during speaking and reading aloud: Evidence from eye-tracking. *Frontiers in Psychology*, *7*, 33. https://doi.org/10.3389/fpsyg.2016.00033

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (pp. 117–220). London: Academic Press.

Greene, J. O., & Burleson, B. R. (2003). *Handbook of communication and social interaction skills*. Mahwah, NJ: Lawrence Erlbaum Associates.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*(1), 1–14. https://doi.org/https://doi.org/10.1016/S0010-0277(01)00138-X

Hardy, S. M., Messenger, K., & Maylor, E. A. (2017). Aging and syntactic representations: Evidence of preserved syntactic priming and lexical boost. *Psychology and Aging*, *32*(6), 588–596. https://doi.org/10.1037/pag0000180

Hartsuiker, R. J., & Kolk, H. H. J. (1998). Syntactic facilitation in agrammatic sentence

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

production. *Brain and Language*, *62*(2), 221–254. https://doi.org/10.1006/brln.1997.1905

Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (pp. 228–249). New York: Oxford University Press.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of Learning and Motivation - Advances in Research and Theory*, *22*, 193–225. https://doi.org/10.1016/S0079-7421(08)60041-9

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*. https://doi.org/10.3758/s13428-017-0935-1

Heyselaar, E., Segaert, K., Walvoort, S. J. W., Kessels, R. P. C., & Hagoort, P. (2017). The role of procedural memory in the skill for language: Evidence from syntactic priming in patients with amnesia. *Neuropsychologia*, *101*, 97–105. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2017.04.033

Heyselaar, E., Wheeldon, L., & Segaert, K. (2017). Structural priming is supported by different components of non-declarative memory: Evidence from priming across the lifespan. *bioRxiv*, *190355*. https://doi.org/10.1101/190355

Howell, D. C. (2010). *Statistical methods for psychology*. Belmont, CA: Wadsworth Cengage Learning.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jaeger, T. F. (2011). *Diagnosing collinearity in mixed models from lme4*. Retrieved from https://hlplab.wordpress.com/2011/02/24/diagnosing-collinearity-in-lme4/

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*(2), 201–258. https://doi.org/10.1016/S0364-0213(87)80006-X

Kemper, S. (1986). Imitation of complex syntactic constructions by elderly adults. *Applied Psycholinguistics*, *7*(3), 277–288. https://doi.org/10.1017/S0142716400007578

Kemper, S. (1987). Life-span changes in syntactic complexity. *Journal of Gerontology*, *42*(3), 323–328. https://doi.org/10.1093/geronj/42.3.323

Kemper, S., Greiner, L. H., Marquis, J. G., Prenovost, K., & Mitzner, T. L. (2001). Language decline across the life span: Findings from the Nun Study. *Psychology and Aging*, *16*(2), 227–239. https://doi.org/10.1037/0882-7974.16.2.227

Kemper, S., Herman, R. E., & Lian, C. H. T. (2003). Age differences in sentence production. *Journal of Gernotology: Psychological Sciences*, *58*(5), 260–269. https://doi.org/10.1093/geronb/58.5.P260

Kemper, S., & Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychology and Aging*, *16*(2), 312–22. https://doi.org/http://dx.doi.org/10.1037/0882-7974.16.2.312

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, *66*, 143–162. https://doi.org/10.1016/j.jml.2011.08.003

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology*, *73*, 1–40. https://doi.org/10.1016/j.cogpsych.2014.04.001

Lara, J., Godfrey, A., Evans, E., Heaven, B., Brown, L. J. E., Barron, E., … Mathers, J. C.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(2013). Towards measurement of the Healthy Ageing Phenotype in lifestyle-based intervention studies. *Maturitas*, *76*, 189–199. https://doi.org/10.1016/j.maturitas.2013.07.007

Lee, J., Yoshida, M., & Thompson, C. K. (2015). Grammatical planning units during real-time sentence production in speakers with agrammatic aphasia and healthy speakers. *Journal of Speech Language and Hearing Research*, *58*(4), 1182. https://doi.org/10.1044/2015_JSLHR-L-14-0250

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1–3), 1–22. https://doi.org/10.1016/0010-0277(92)90038

Levelt, W. J. M., & Maassen, B. (1981). Lexical search and order of mention in sentence production. In W. J. M. Levelt & W. Klein (Eds.), *Crossing the boundaries in linguistics* (pp. 221–252). Dordrecht, Netherlands: D. Reidel Publishing Company.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75. https://doi.org/10.1017/S0140525X99001776

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*(3), 295–327. https://doi.org/10.1037/a0035230

MacDonald, M. C. (2016). Speak, act, remember: The language-production basis of serial order and maintenance in verbal memory. *Current Directions in Psychological Science*, *25*(1), 47–53. https://doi.org/10.1177/0963721415620776

Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language*, *91*, 5–27.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

https://doi.org/10.1016/j.jml.2016.03.009

Martin, R. C., Crowther, J. E., Knight, M., Tamborello, F. P., & Yang, C. L. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition*, *116*(2), 177–192. https://doi.org/10.1016/j.cognition.2010.04.010

Martin, R. C., Miller, M., & Vu, H. (2004). Lexical-semantic retention and speech production: Further evidence from normal and brain-damaged participants for a phrasal scope of planning. *Cognitive Neuropsychology*, *21*(6), 625–644. https://doi.org/10.1080/02643290342000302

Martin, R. C., Yan, H., & Schnur, T. T. (2014). Working memory and planning during sentence production. *Acta Psychologica*, *152*, 120–132. https://doi.org/10.1016/j.actpsy.2014.08.006

Mick, P., Kawachi, I., & Lin, F. R. (2014). The association between hearing loss and social isolation in older adults. *Otolaryngology - Head and Neck Surgery*, *150*(3), 378–384. https://doi.org/10.1177/0194599813518021

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., … Chertkow, H. (2005). The Montreal cognitive assessment, MoCA: A brief screening for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*, 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Oberauer, K., Süß, H. M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, *31*(2), 167–193. https://doi.org/10.1016/S0160-2896(02)00115-0

Pathan, S. S., Gottesman, R. F., Mosley, T. H., Knopman, D. S., Sharrett, A. R., & Alonso, A. (2011). Association of lung function with cognitive decline and dementia: The Atherosclerosis Risk in Communities (ARIC) Study. *European Journal of Neurology*, *18*(6), 888–898. https://doi.org/10.1111/j.1468-1331.2010.03340.x

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

Payne, B. R., & Stine-Morrow, E. A. L. (2014). Adult age differences in wrap-up during sentence comprehension: Evidence from ex-gaussian distributional analyses of reading time. *Psychology and Aging*, *29*(2), 213–228. https://doi.org/10.1037/a0036282

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*(4), 633–651. https://doi.org/10.1006/jmla.1998.2592

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, *134*(3), 427–459. https://doi.org/10.1037/0033-2909.134.3.427

Rabaglia, C. D., & Salthouse, T. A. (2011). Natural and constrained language production as a function of age and cognitive abilities. *Language and Cognitive Processes*, *26*(10), 1505–1531. https://doi.org/10.1080/01690965.2010.507489

R Core Team. (2015). *A language and enviroment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Raven, J., Raven, J. C., & Court, J. H. (1988). *Manual for Raven's progressive matrices and vocabulary scales.* London: H. K. Lewis.

Reifegerste, J., Elin, K., & Clahsen, H. (2018). Persistent differences between native speakers and late bilinguals: Evidence from inflectional and derivational processing in older speakers. *Bilingualism: Language and Cognition*, 1–16. https://doi.org/10.1017/S1366728918000615

Salthouse, T. A. (1993). Speed and knowledge as determinants of adult age differences in verbal tasks. *Journal of Gerontology: Psychological Sciences*, *48*(1), 29–36. https://doi.org/10.1093/geronj/48.1.P29

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide*. Pittsburgh, PA: Psychology Software Tools Inc.

Schoot, L., Heyselaar, E., Hagoort, P., & Segaert, K. (2016). Does syntactic alignment effectively influence how speakers are perceived by their conversation partner? *PloS One*, *11*(4), e0153521. https://doi.org/10.1371/journal.pone.0153521

Scott, R. M., & Wilshire, C. E. (2010). Lexical competition for production in a case of nonfluent aphasia: Converging evidence from four different tasks. *Cognitive Neuropsychology*, *27*(6), 505–538. https://doi.org/10.1080/02643294.2011.598853

Segaert, K., Lucas, S. J. E., Burley, C. V, Segaert, P., Milner, A. E., Ryan, M., & Wheeldon, L. (2018). Fit to speak - Physical fitness is associated with reduced langauge decline in healthy ageing. *Scientific Reports*, *8*, 6715. https://doi.org/10.1038/s41598-018-24972-1

Segaert, K., Menenti, L., Weber, K., & Hagoort, P. (2011). A paradox of syntactic priming: Why response tendencies show priming for passives, and response latencies show priming for actives. *PLoS ONE*, *6*(10), e24209. https://doi.org/10.1371/journal.pone.0024209

Segaert, K., Weber, K., Cladder-Micus, M., & Hagoort, P. (2014). The influence of verb-bound syntactic preferences on the processing of syntactic structures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1448–1460. https://doi.org/10.1037/a0036796

Segaert, K., Wheeldon, L., & Hagoort, P. (2016). Unifying structural priming effects on syntactic choices and timing of sentence generation. *Journal of Memory and Language*, *91*, 59–80. https://doi.org/10.1016/j.jml.2016.03.011

Shafto, M. A., Burke, D. M., Stamatakis, E. A., Tam, P. P., & Tyler, L. K. (2007). On the tip-of-the-tongue: Neural correlates of increased word-finding failures in normal aging. *Journal of Cognitive Neuroscience*, *19*(12), 2060–2070.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

https://doi.org/10.1162/jocn.2007.19.12.2060

Slevc, L. R. (2011). Saying what's on your mind: Working memory effects on sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1503–1514. https://doi.org/10.1037/a0024350

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, *73*(3), 205–246. https://doi.org/10.1016/S0010-0277(99)00053-0

Smith, M., & Wheeldon, L. (2001). Syntactic priming in spoken sentence production – an online study. *Cognition*, *78*(2), 123–164. https://doi.org/10.1016/S0010-0277(00)00110-4

Smith, T., Gildeh, N., & Holmes, C. (2007). The Montreal Cognitive Assessment: Validity and utility in a memory clinic setting. *Canadian Journal of Psychiatry*, *52*(5), 329–332. https://doi.org/10.1177/070674370705200508

Speer, P., & Wilshire, C. E. (2013). What's in a sentence? The crucial role of lexical content in sentence production in nonfluent aphasia. *Cognitive Neuropsychology*, *30*(7–8), 507–543. https://doi.org/10.1080/02643294.2013.876398

Stine-Morrow, E. A. L., & Payne, B. R. (2016). Age differences in language segmentation. *Experimental Aging Research*, *42*(1), 83–96. https://doi.org/10.1080/0361073X.2016.1108751

Sung, J. E. (2015). Age-related changes in sentence production abilities and their relation to working-memory capacity: Evidence from a verb-final language. *PLoS ONE*, *10*(4). https://doi.org/10.1371/journal.pone.0119424

Swets, B. (2015). Psycholinguistics and planning: A focus on individual differences. In S. Fuchs, D. Pape, C. Petrone, & P. Perrier (Eds.), *Speech production and perception: Individual differences in speech production and perception* (pp. 89–122). Frankfurt, Germany: Peter Lang International Academic Publishers.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Swets, B., Jacovina, M. E., & Gerrig, R. J. (2014). Individual differences in the scope of speech planning: Evidence from eye-movements. *Language and Cognition*, *6*(2014), 12–44. https://doi.org/10.1017/langcog.2013.5

Taylor, J. K., & Burke, D. M. (2002). Asymmetric aging effects on semantic and phonological processes: naming in the picture-word interference task. *Psychology and Aging*, *17*(4), 662–676. https://doi.org/10.1037/0882-7974.17.4.662

United Nations. (2011). *International standard classification of education*. Montreal, Canada: UN Institute for Statistics. Retrieved from http://uis.unesco.org/en/topic/international-standard-classification-education-isced

van de Velde, M., & Meyer, A. S. (2014). Syntactic flexibility and planning scope: The effect of verb bias on advance planning during sentence recall. *Frontiers in Psychology*, *5*, 1174. https://doi.org/10.3389/fpsyg.2014.01174

Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 423–440. https://doi.org/10.1037/a0018619

Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 550–564. https://doi.org/10.3758/BF03195534

Wechsler, D. (1997). *Wechsler memory scale - third edition manual*. San Antonia, TX: Pearson.

Wechsler, D. (2008). *Wechsler adult intelligence scale - fourth edition*. San Antonio, TX: Pearson.

Wheeldon, L. (2011). Generating spoken sentences: The relationship between words and syntax. *Linguistics and Language Compass*, *5*(6), 310–321.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

https://doi.org/10.1111/j.1749-818X.2011.00281.x

Wheeldon, L. (2013). Producing spoken sentences: The scope of incremental planning. In S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.), *Speech production and perception: Speech planning and dynamics* (pp. 97–118). Frankfurt, Germany: Peter Lang International Academic Publishers.

Wheeldon, L., Ohlson, N., Ashby, A., & Gator, S. (2013). Lexical availability and grammatical encoding scope during spoken sentence production. *The Quarterly Journal of Experimental Psychology*, *66*(8), 1653–1673. https://doi.org/10.1080/17470218.2012.754913

Wheeldon, L., & Smith, M. (2003). Phrase structure priming: A short-lived effect. *Language and Cognitive Processes*, *18*(4), 431–442. https://doi.org/10.1080/01690960244000063

Zhao, L.-M., & Yang, Y.-F. (2016). Lexical planning in sentence production is highly incremental: Evidence from ERPs. *PLoS ONE*, *11*(1), e0146359. https://doi.org/10.1371/journal.pone.0146359

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Footnotes

[1] Note, some other studies have tested non-young adults as controls for clinical patients; however, the samples are small and the age ranges are large. While Ferreira, Bock, Wilson, and Cohen, (2008, *n* = 4 aged 50-58) and Cho-Reyes, Mack, and Thompson (2016; *n* = 13 aged 33-76) found evidence of choice syntactic priming in controls, Hartsuiker and Kolk (1998; *n* = 12 aged ~28-67) did not.

[2] We note that some previous studies of incremental sentence production in aphasic patients have used non-young adults as controls; however, again the samples are often small and the age ranges are large (e.g., Lee, Yoshida, & Thompson, 2015; Martin et al., 2004; Scott & Wilshire, 2010; Speer & Wilshire, 2013). It is therefore difficult to draw any firm conclusions about the effect of old age on incremental planning from these studies.

[3] Significant effects in the analyses remain if the participants excluded from Experiment 1 are also excluded from Experiment 2 and vice-versa.

[4] We speculate that the lack of age differences in the working memory may be because we only used one measure. Use of multiple measures can potentially provide a better indicator of overall working memory ability because different tasks target different facets of working memory (Oberauer, Süß, Wilhelm, & Wittman, 2003; Waters & Caplan, 2003).

[5] Due to the large speed differences between young and older adults, we also performed the modelling analysis with age-standardised onset latencies (using z-score adjustments within age groups). This produced the same effects (expect for the main effect of age) seen in the non-adjusted onset latencies analyses for both Experiments 1 and 2.

[6] The ‘testInteractions’ function in the *phia* package (de Rosario-Martinez, 2015) allows for the direct comparison of the contrasts specified within the model and adjusts *p* values for multiple comparisons using the Holm-Bonferroni correction.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1

Means and Standard Deviations of the Background Characteristics and Performance on the Aging Markers for Young and Older Adults, Including the Results of Comparisons Between the Age Groups (Independent Samples t-tests)

	Young		Older		Comparison	
Measure	*M*	*SD*	*M*	*SD*	*t*(104)	*p*
Age (years)	19.8	1.1	71.8	4.5	---	---
Education[a]	6.0	0.14	5.8	1.3	1.36	.178
Processing speed	80.90	14.77	67.61	14.62	4.65	< .001
Short-term memory	7.62	1.66	5.66	1.27	6.86	< .001
Long-term memory	5.14	1.60	3.16	1.51	6.54	<.001
Working memory	5.43	1.45	5.18	1.19	0.98	.331
Vocabulary	16.82	2.72	23.79	3.03	-12.39	<.001
Inhibition	207.7	57.1	255.4	57.2	-4.29	< .001
Handgrip	28.54	8.10	27.82	8.92	0.43	.667
Lung capacity	3.46	0.75	2.26	0.66	8.82	< .001

Note. All values represent raw non-standardised scores. Detailed descriptions of the aging marker are included in Table 3. [a]Education was scored according to the International Standard Classification of Education (United Nations, 2011), which classifies education on a scale of 0 (pre-primary school) to 8 (university doctorate).

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2

Overview of the Different Items Used in the Experiments 1 and 2. Number of Items and Example Stimuli Completed by Each Participant are Provided.

Item Type	N	Example
A: Experiment 1		
Related	20	Prime: "the pencil and the orange move together" Target: "the clock and the drum move up"
Unrelated	20	Prime: "the cow moves up and the broom moves down" Target: "the apple and the goat move up"
Filler	120	"There are two houses"
B: Experiment 2		
Preview Initial Coordinate	20	Preview: spoon "The trumpet and the spoon move above the crab"
No Preview Initial Coordinate	20	Preview: NA "The skirt and the bell move above the carrot"
Preview Initial Simple	20	Preview: snail "The balloon moves above the snail and the pear"
No Preview Initial Simple	20	Preview: NA "The spanner moves above the monkey and the toaster"
Filler	220	"There are three stars"

Note. The condition to which each experimental item was assigned was rotated across lists (e.g., the picture trio of trumpet-spoon-crab would also have appeared in the three other conditions in Experiment 2 in alternative lists). This meant that, across all participants, each item appeared an equal number of items in each condition; therefore, lexical factors of individual words, such as age of acquisition, were not a concern.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3.

Descriptions of the Measurement and Procedure Used for Each Aging Marker.

Aging Marker	Measurement	Procedure details
Processing speed	Coding task (WAIS-IV; Wechsler, 2008)	A written task in which participants must match as many numbers as possible to arbitrary symbols (following a key) in a two minute time period.
Short-term and long-term memory	Immediate and delayed word recall (WMS; Wechsler, 1997)	Immediate recall of 12 unrelated words read aloud by the experimenter (short-term memory). Then recall of the words again after a five minute interval (long-term memory).
Vocabulary	Mill Hill vocabulary test (Raven, Raven, & Court, 1988)	Multi-choice task in which the participant must select the correct definition of a word; there were 34 questions that increased in difficulty.
Working memory	Backward digit span (Waters & Caplan, 2003)	Reverse recall of number sequences read aloud by the experimenter (ranging from 3-8 span lengths). A participant's score was defined at the span length at which he/she could correctly recall the digit in reverse order on three out of five trials.
Inhibition	Stop-signal task (Logan & Cowan, 1984)	Participant had to respond to a 'go' stimulus as quickly as possible, but to without their response if a stop-signal appeared (the delay between the 'go' and 'stop' signal was varied dynamically). The stop signal reaction time (SSRT) was calculated by subtracting a participant final stop-signal delay from their average response time to the 'go' stimulus; a smaller SSRT score indicated better inhibitory control.
Handgrip	*Jamar* hand dynamometer (Cooper et al., 2011)	Participant held the dynamometer upwards and then moved their arm down while squeezing with maximal effort for three seconds. The highest value across six trials (three per hand) was used for analysis.
Lung capacity	Forced expiratory volume in one second (FEV1; Pathan et al., 2011).	Participant blew as hard as possible into the *Vitalograph In2itive* spirometer until their lungs were empty (aiming for blow for at least six seconds). The highest value across three trials was used for analysis.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 4

Summary of the Best-Fitted Mixed-Effects Models for the Experiment 1 Error Data.

Predictor	Coefficient	SE	Wald Z	*p*
A: all data				
Intercept	2.33	0.16	14.71	< .001
Prime type	-0.11	0.14	-0.79	.432
Age group	0.76	0.20	3.77	< .001
Inhibition	0.06	0.08	0.77	.443
Prime type * Age group	-0.16	0.23	-0.70	.487
Prime type * Inhibition	-0.07	0.12	-0.63	.526
Age group * Inhibition	0.22	0.16	1.43	.153
Prime type * Age group * Inhibition	0.51	0.23	2.19	.029
B: young adults				
Intercept	2.70	0.20	13.83	< .001
Prime type	-0.09	0.23	-0.40	.691
Inhibition	0.17	0.12	1.44	.149
Prime type * Inhibition	0.19	0.18	1.09	.276
C: older adults				
Intercept	1.96	0.18	10.90	< .001
Prime type	-0.14	0.18	-0.79	.430
Inhibition	-0.05	0.10	-0.48	.632
Prime type * Inhibition	-0.33	0.15	-2.13	.033

Note. The final model of the complete data set did not differ significantly from the full model in terms of variance explained (full model AIC = 2476; final model AIC = 2450; $p = .567$). The model converged with random intercepts for participants and items with an additional by-participant random slope for the main effect of age group, and by-item random slopes for the main effects of prime type and age group.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 5

Summary of the Best-fitted Mixed-effects Models for the Experiment 1 Onset Latency Data.

Predictor	Coefficient	SE	*t*-value	*p*
A: all data				
Intercept	982.97	23.47	41.88	< .001
Prime type	39.56	9.98	3.96	< .001
Age group	-173.23	42.63	-4.06	< .001
Long-term memory	-42.42	21.24	-2.00	.045
Prime type * Age group	-9.21	19.41	-0.47	.635
B: young adults				
Intercept	898.49	27.17	33.07	< .001
Prime type	35.05	11.55	3.03	.002
Long-term memory	-31.74	26.36	-1.20	.228
C: older adults				
Intercept	1074.01	35.81	29.99	< .001
Prime type	42.63	14.70	2.90	.004
Long-term memory	-55.57	33.53	-1.66	.097

Note. The final model of the complete data set did not differ significantly from the full model in terms of variance explained (full model AIC = 42178; final model AIC = 42143; $p = .985$). The model converged with a fully-expressed random intercepts and slopes structure.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 6

Summary of the Best-fitted Mixed-effects Model for the Experiment 2 Error Data.

Predictor	Coefficient	SE	Wald Z	*p*
A: all data				
Intercept	2.02	0.15	13.62	< .001
Preview	-0.07	0.07	-1.02	.308
Initial phrase type	0.12	0.07	1.66	.098
Age group	0.89	0.16	5.70	< .001
Preview * Initial phrase type	-0.28	0.14	-2.05	.040
Preview * Age group	0.14	0.14	1.05	.292
Initial phrase type * Age group	-0.04	0.14	-0.32	.747
Preview * Initial phrase type * Age group	0.29	0.27	1.07	.285
B: young adults				
Intercept	2.50	0.17	14.65	< .001
Preview type	-0.06	0.12	-0.51	.607
Initial phrase type	0.14	0.12	1.16	.245
Preview * Initial phrase type	-0.20	0.21	-0.93	.352
C: older adults				
Intercept	1.59	0.17	9.18	< .001
Preview type	-0.12	0.09	-1.43	.154
Initial phrase type	0.12	0.09	1.39	.163
Preview * Initial phrase type	-0.41	0.17	-2.51	.016

Note. The final model of the complete data set did not differ significantly from the full model in terms of variance explained (full model AIC = 6411; final model AIC = 6356; $p = .751$). The model converged with random intercepts for participants and items with additional by-participant random slopes for the main effects of preview and initial phrase type, and a by-item random slope for the main effect of age group.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 7

Summary of the Best-fitted Mixed-effects Model for the Experiment 2 Onset Latency Data.

Predictor	Coefficient	SE	*t*-value	*p*
A: all data				
Intercept	924.27	17.03	54.30	< .001
Preview	-52.68	8.68	-6.07	< .001
Initial phrase type	-39.80	5.55	-7.18	< .001
Age group	-150.02	32.18	-4.66	< .001
Short-term memory	-5.15	16.06	-0.32	.662
Preview * Initial phrase type	49.43	10.80	4.57	< .001
Preview * Age group	-22.97	16.74	-1.37	.170
Initial phrase type * Age group	-0.28	11.05	-0.03	.991
Preview * Short-term memory	-12.28	8.57	-1.43	.162
Phrase type * Short-term memory	-0.55	5.55	-0.10	.947
Age group * Short-term memory	-19.45	32.11	-0.61	.611
Preview * Initial phrase type * Age group	-24.50	21.52	-1.14	.250
Preview * Initial phrase type * Short-term memory	13.97	10.81	1.29	.211
Preview * Age group * Short-term memory	-9.87	17.20	-0.57	.555
Initial phrase type * Age group * Short-term memory	17.39	11.10	1.57	.117
Preview * Initial phrase type * Age group * Short-term memory	-36.79	21.61	-1.70	.089
B: young adults				
Intercept	849.18	19.41	43.74	< .001
Preview	-63.06	10.62	-5.94	< .001
Initial phrase type	-41.19	7.77	-5.30	< .001
Short-term memory	-13.47	18.82	-0.72	.239
Preview * Initial phrase type	38.03	15.01	2.53	.011
Preview * Short-term memory	-23.30	10.46	-2.23	.025
Phrase * Short-term memory	7.97	7.70	1.04	.301

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Preview * Initial phrase type * Short-term memory	-3.45	14.42	-0.24	.881
C: older adults				
Intercept	998.68	26.20	38.12	<.001
Preview	-41.30	13.53	-3.05	.003
Initial phrase type	-40.64	8.51	-4.78	< .001
Short-term memory	4.22	25.44	0.17	.809
Preview * Initial phrase type	61.87	16.59	3.73	< .001
Preview * Short-term memory	-7.86	16.59	3.73	.568
Phrase * Short-term memory	-9.32	8.14	-1.14	.253
Preview * Initial phrase type * Short-term memory	32.38	16.14	2.01	.045

Note. The final model of the complete data set did not differ significantly from the full model in terms of variance explained (full model AIC = 90826; final model AIC = 90784; p = .536). The model converged with random intercepts for participants and items with additional by-participant random slopes for the main effects of preview and initial phrase type, and by-item random slopes for the main effects of preview and age group.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

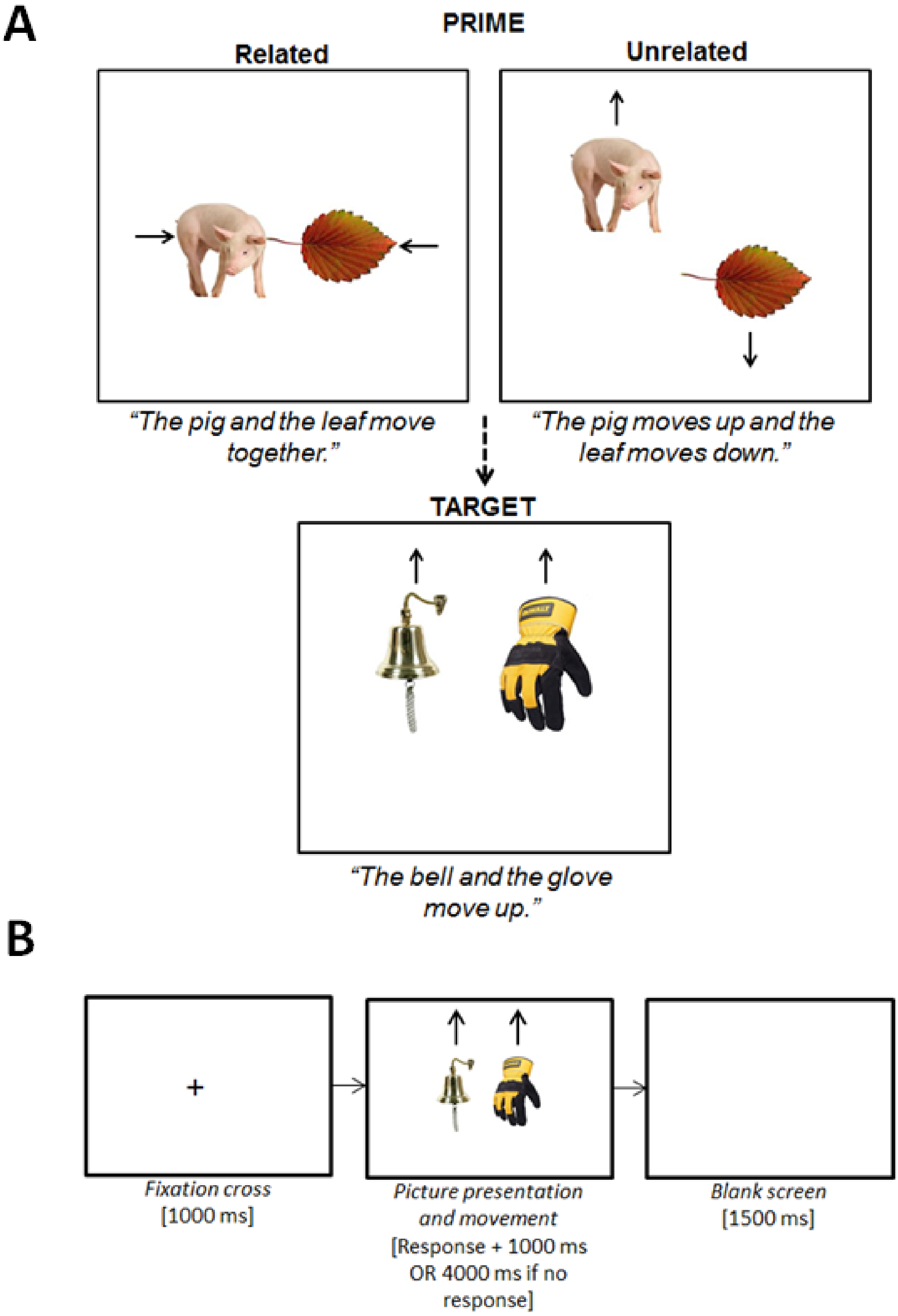

Figure 1. Experiment 1 syntactic priming task design (A) and stimuli presentation events per trial (B). The participant was instructed to begin describing the picture movement as soon as possible using specific sentence types. The stimuli presentation sequence was the same for prime and target trials, and primes were always immediately followed by the corresponding target (i.e., we used a 0-lag delay). Speech latencies on the target trials were recorded from the onset of the pictures to the participant beginning to speak.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

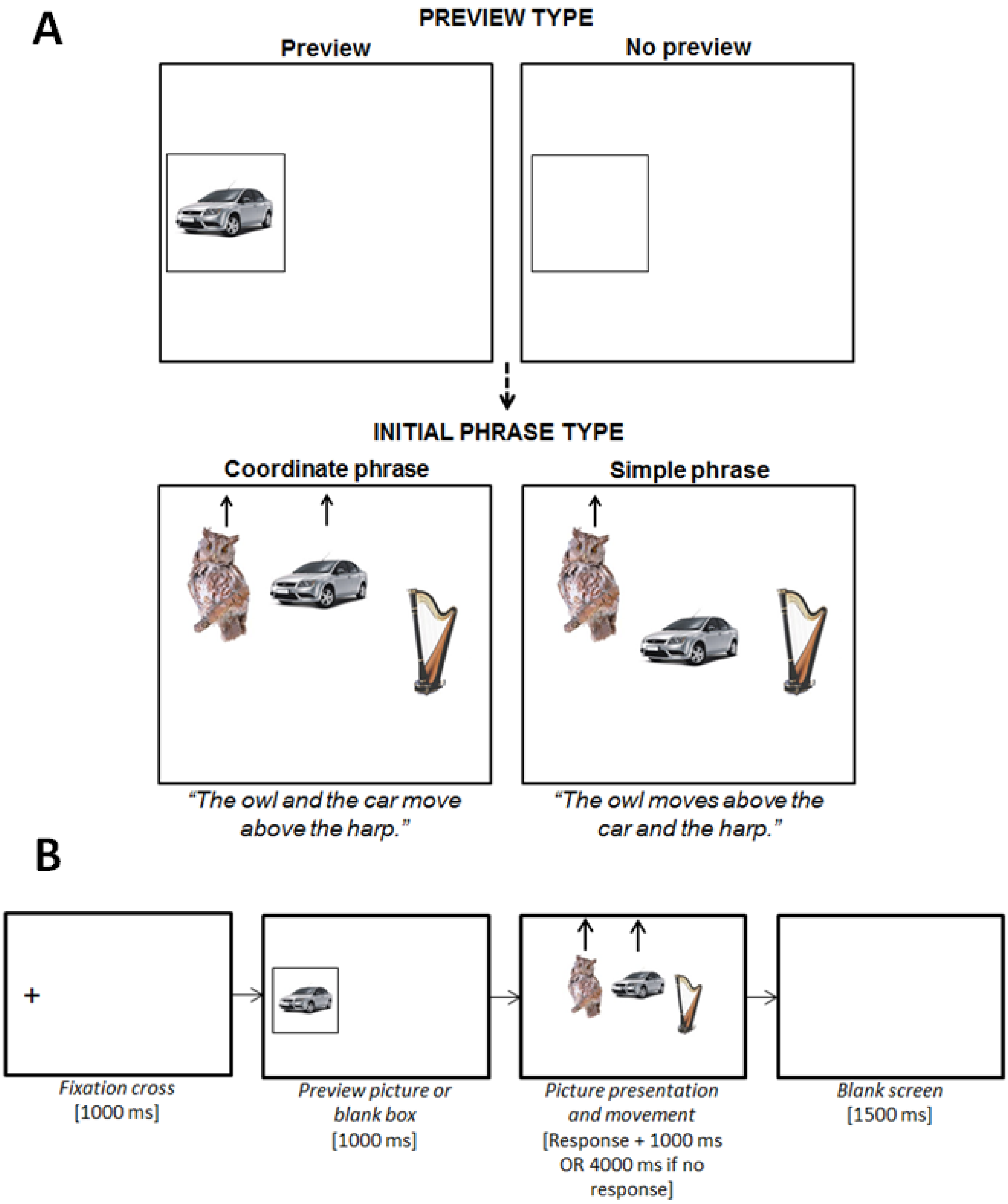

Figure 2. Experiment 2 planning scope design (A) and stimuli presentation events per trial (B). The participant was instructed to pay attention to the preview because it would appear in the upcoming trial, but not to name it aloud. The three pictures then appeared aligned centrally in the horizontal plane (importantly, the leftmost picture did not appear where the preview picture had just been, but in a more right-adjusted position). The participant was instructed to begin describing the picture movement as soon as possible using specific sentence types. Speech latencies were recorded from the onset of the pictures to the participant beginning to speak.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

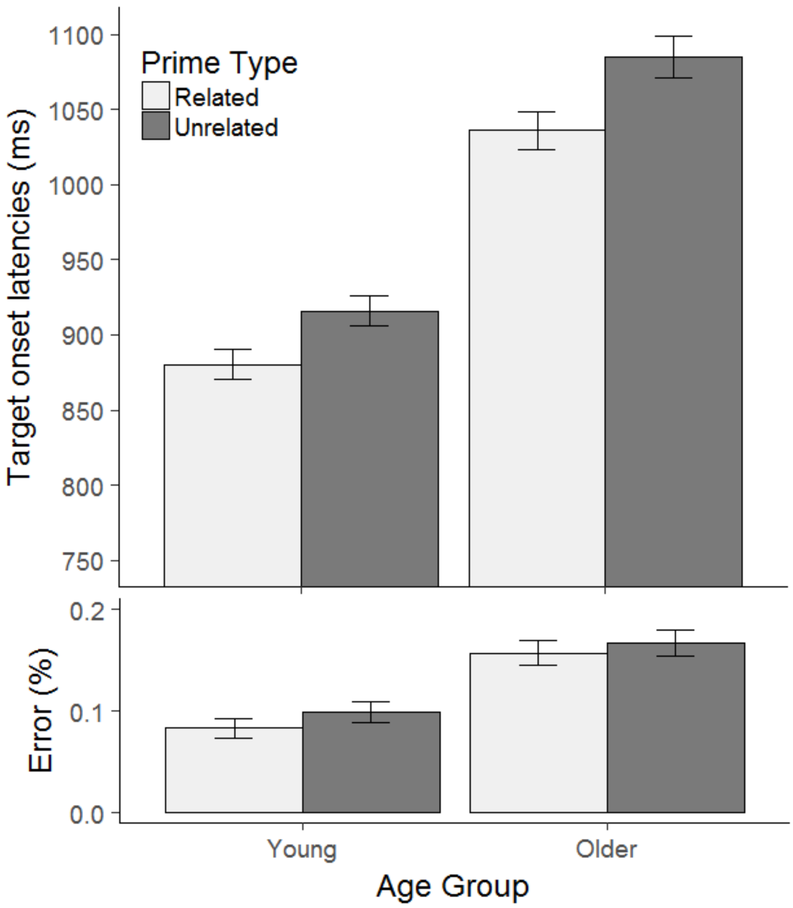

Figure 3. Experiment 1 mean target error rates and onset latencies for young and older adults following syntactically related and unrelated primes. Error bars denote ±1 the standard error of the mean.

bioRxiv preprint doi: https://doi.org/10.1101/327304; this version posted October 10, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

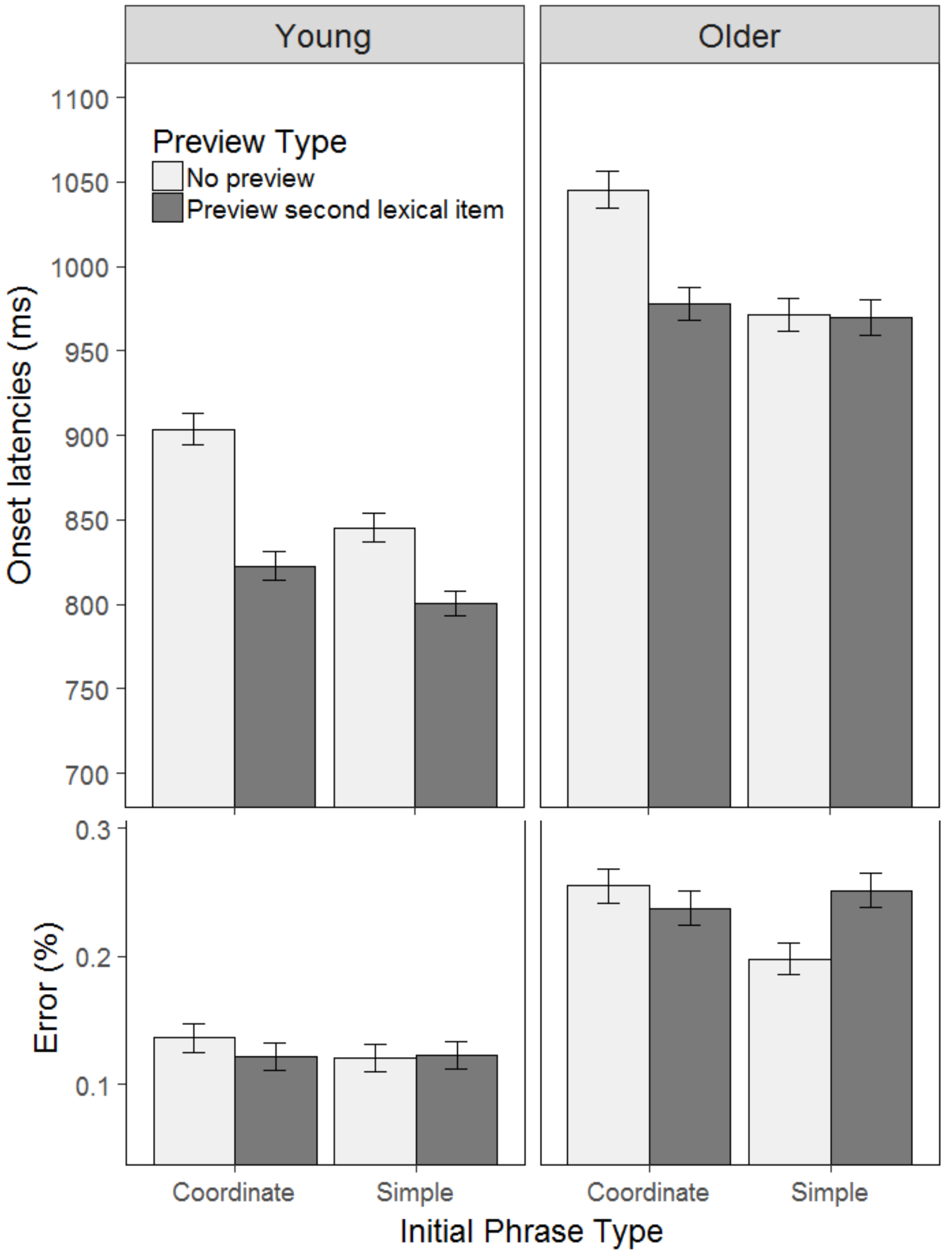

Figure 4. Experiment 2 mean error rates and onset latencies for young and older adults when producing sentences within initial coordinate and simple phrases following no preview or a preview of the second upcoming lexical item. Error bars denote ±1 the standard error of the mean.