Word production within sentence contexts: the role of cognitive ageing

Naveen Mujtaba Dar
(formerly Naveen Hanif)

Doctor of Philosophy

University of York

Psychology

May 2025

Abstract

Although the process of word production is seemingly effortless for most healthy adults, word retrieval declines with age (e.g., Condret-Santi et al., 2013). Words are usually produced within sentences of ongoing speech produced by speakers and their conversational partner(s). Sentence contexts can influence word retrieval in younger adults during production, causing facilitation when they predict the word to retrieve, but potential interference when speakers must retrieve an unexpected word instead. In this thesis, I studied these sentence-context effects, and the mechanisms underlying them, in healthy younger and older adults. Furthermore, I studied how potential age-related changes in word production might relate to other cognitive mechanisms argued to change with ageing, focusing on semantic control.

Through the empirical work in this thesis, I firstly examined how word production is influenced by sentence contexts predicting different types of target words in both younger and older adults (Chapter 2). I also examined the cognitive mechanisms that might relate to word production in ageing (Chapter 3). In Chapter 4, I further explored how the position of critical information (predicting the target word) in a sentence affects word production. In the final empirical chapter, I explored the effects of degree of unexpectedness on word production, while also manipulating the role of semantic competition.

Collectively, the findings suggest older adults' word production continues to benefit from sentence context. Preserved semantic networks might allow older adults to keep using priming and/or prediction in sentence contexts to aid their word retrieval. Although some declines in older adults' semantic and general cognitive control were observed, older adults did not show greater difficulties with unexpected or otherwise highly demanding sentence contexts than their younger counterparts. The assimilation of the traditional picture-naming paradigm within sentence contexts was an important step towards understanding how older adults' language changes within daily-life speech contexts.

Table of Contents

List of Figures

List of tables

Acknowledgements

وما توفيقي إلا بالله

My success is only by Allah (SWT)

As for all blessings gifted to me in this life, utmost gratitude and credit for the feat of completing this thesis undoubtedly belongs to Allah SWT; my Creator, my Provider, my Sustainer and (best) Guide and Teacher.

Heartfelt thanks are also due to my brilliant supervisors Dr Angela de Bruin and Professor Beth Jefferies for giving me the priceless opportunity to learn from them over the course of the last five years. You are both incredibly inspirational and fantastic in your own ways; and learning from you in such proximity has set a golden example for me of how academic research should be done. Angela, thank you for your dedication as a supervisor. Thank you even more for the laughs; the random chats and the last-minute emails and words of encouragement before a talk or big presentation. You supervised me through some of the most monumental transitions of my life, continuously providing solutions to help me balance my PhD with all my other roles outside of academia. for that, I will always be grateful. Beth, thank you for *everything*. Most of all, thank you for your kindness and warmth which you responded with every single time I saw you, or sent a draft of writing your way.

I would also like to extend my gratitude to my thesis advisory panel (Professors Sven Mattys and Lisa-Marie Henderson) and progression chair (Professor Gareth Gaskell) for their invaluable feedback on my research projects during my tap and progression meetings.

Credit is also due to both my parents (who nurtured my love for language from the very beginning of my life); from getting me my first library card, to buying me all the books I ever asked for. These actions were paramount in helping me to unearth my inner voice and my desire to connect with words *at* such an intricate level. It is only now, years later through my connection to the field of Psychology that I have come to appreciate the true impact that these seemingly small actions in my early years have likely had on shaping my life's path, and the person I am today. My dear husband Mujtaba who is a pillar of support for me, in everything I do. Maier Wali, my heart, my little joy, my first- born who taught me that women really can do anything (finishing a PhD with a toddler is not for the weak). Saad and Chelsea, and Zunaira

and Hamza, for above all, gifting me my precious nieces and nephew, each of whom added a little bit of magic to the rough days of this endeavour. I pray that this achievement will serve as a source of inspiration to them (and all our future babies) tomorrow, to have and chase their own dreams with determination, grit and a tough spirit.

Finally, gratitude is owed to the participants who took the time out to participate in my research; without them my thesis would simply be a big collection of ideas, of ‘what if’s’ and ‘maybe’s’.

Author's Declaration

I, Naveen Mujtaba Dar (formerly Naveen Hanif), declare that this thesis is a presentation of original work, carried out under the supervision of Dr Angela de Bruin and Professor Elizabeth Jefferies, and that I am the sole author. This work has not previously been presented for a degree or other qualification at this University (University of York, UK) or elsewhere. All studies included in this thesis were conducted in accordance with the ethical standards of the department of Psychology at the University of York. All sources are acknowledged as references. All studies were supported by a departmental scholarship. Please note that the General Introduction and Discussion chapters are written in first-person singular form ("I") while the empirical chapters are written in first-person plural form ("we").

The work presented in Chapters 2 and 3 have been published jointly in a peer-reviewed journal article:

Hanif, N., Jefferies, E., & de Bruin, A. (in press). Naming speed during language production in younger and older adults: Examining the effects of sentence context. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218241309602

Chapter 1: General Introduction

1.1 Background

The topic of cognitive ageing is highly relevant given that the ageing population is larger than it has ever been. For example, in England alone, the number of people aged 50 years and over has increased by 47% in the last 40 years (Centre for Ageing Better, 2023). Difficulties relating to word production (specifically difficulties in finding the words they wish to use within conversation) are commonly reported by healthy older adults (e.g., Condret-Santi et al., 2013; Lovelace & Twohig, 1990). In addition to word finding difficulties, older adults also experience other difficulties during everyday conversations. For instance, they experience difficulties in adhering to conversational topics (e.g., Hoffman et al., 2018). Compared to healthy younger adults, they also exhibit more frequent tip of the tongue states (TOTs) a phenomenon whereby speakers are able to recall partial information about a word they need to produce such as its meaning, or other words with a similar meaning/word form, but are not able to recall the word itself ; e.g., Burke et al., 1991; Cohen & Faulkner 1986; Evrard, 2002; Heine et al., 1999; James et al., 2018; Rastle & Burke, 1996; Salthouse & Mandell, 2013; Shafto et al., 2007). Finally, older adults also have been shown to make more speech errors such as omissions (i.e., missing out segments of spoken words such as producing '**pan**' instead of '**pans**', or '**bake fluid**' instead of '**brake fluid**'; e.g., MacKay & James, 2004). Given that language ability is central to many interactions that form part of our rich social experiences (e.g., Keller-Cohen et al. 2006), it is paramount to try to understand these age-related changes in language production, as well as how they present in interactions with other speakers.

Although language changes with age have been studied in many forms (including the examples mentioned above), *speed* of word retrieval during production might be especially sensitive to changes with ageing from a relatively younger age (e.g., Kavé & Goral, 2017; Verhaegen and Poncelet, 2013). However, many studies within the language production literature looking at naming times have focused on studying how words are produced in isolation (e.g., in tasks asking people to name individual pictures without any conversational context). This is despite an emerging literature showing that a strong contributing factor to ease of word retrieval and production efficiency is the sentence context within which words are to be produced (Bannon et al., 2025; Shao & Rommers, 2020). A key open question is how these sentence contexts influence word retrieval during production in *older adults,*

particularly in comparison to younger adults. In this thesis, I examine how context-target word relationships that are present in a conversational partner's speech affect one's own ongoing/subsequent production mechanisms. This will be done with a particular focus on *matched* contexts (ones that align with and support the retrieval of target words) as well as on *mismatched* contexts (ones that predict a different word than the target). Furthermore, age-related changes in semantic processing and in the availability of cognitive resources might impact how words are retrieved within these contexts (e.g., Burke et al., 1991, 2004; Burke & Shafto, 2004; Hoffman et al., 2018). Therefore, I explore language production operations within the framework of healthy ageing and examine the role of language, semantic, and cognitive abilities during language production.

1.2 The phenomenon of word production

Words are usually produced relatively effortlessly as part of fluent conversation. For example, on average, in conversations, adults produce around two-three words per second (Levelt et al., 1999). Furthermore, the mental lexicon is thought to expand over the course of one's lifespan. While a twenty-year-old native speaker of American-English is estimated to know 42,000 words, a sixty-year-old is estimated to know 48,000 (Brysbaert et al., 2016). Within a conversation, required words often need to be retrieved quickly. Usually a conversation consists of turn-taking between two or more speakers. Each turn lasts for around 2 seconds (Levinson, 2016), and the gap between turns only lasts for around 200ms (Heldner & Edlund, 2010; Stivers et al., 2009). With language production processes taking at least 600ms (Levinson & Torreira, 2015), this means speakers likely need to process their conversation partner's speech while also already preparing their own upcoming response. The quick and accurate selection and retrieval of required words in the short time window contained within conversational turns therefore is an extremely perplexing feat.

The semantic system is a store of our conceptual world knowledge and is believed to be where word production begins. For example, it is postulated that in the earliest stages of word production, the conceptual features of the word that a speaker needs to produce become active in their mind (Dell et al., 1997, Roelofs, 1992, 1997, 2003). If a speaker wishes to say '**cat**', semantic nodes representing related features become activated (e.g., whiskers, paws, tail, house pet). Following this, in Levelt's influential model of word production (Levelt et al., 1999), the next step in the word production process involves 'lemma selection'. This

occurs through active semantic concepts triggering the activation of items they are connected to at the lemma level. Taking the example of the word '**cat**', following activation of the conceptual features of a cat, the lemma of this target becomes active (i.e., the lemma '**cat**' comprises all variations of it, including '**cat**' and '**cats')**. Together, semantic activation and lemma selection compose the first part of word production. The second part of this process includes 'form encoding', which comprises code retrieval: retrieving the morphological and phonological codes of the target word (e.g., '**cat**' includes the morpheme /**cat**/ and the phonemes /**kæt**/); prosodification and syllabification (chunking the morpho-phonological code into syllables); and phonetic encoding and articulation, which involve preparing the appropriate articulatory code needed for producing the required word. Finally, speakers articulate the word using the laryngeal and supra-laryngeal apparatus (Levelt, 2001). These outlined stages of word production are core parts of influential models such as Dell's two step interactive activation model (Dell et al., 1997) and Roelofs's Weaver network model (Roelofs, 1992, 1997, 2003).

Focussing on the semantic/conceptual level of word production, as mentioned previously, activation at this level triggers the activation of the lexical item that the speaker wishes to produce (e.g., whiskers, paws, and house pet activate the lemma '**cat**'). Crucially, activation in semantic representations also leads to the activation of other similar lexical items that share or have connections with those semantic features (e.g., '**dog**', '**mouse**', and '**gerbil**') through the mechanism of spreading activation (Dell, 1986). The phenomenon of spreading activation can give rise to two polar effects, depending on the type of task and processing required. The first is *semantic priming* wherein the retrieval/activation of one word can facilitate access to/retrieval of related words. For example, healthy younger and older adults are faster to produce presented target words that are related to a previously presented prime word (e.g., furniture – '**sofa***'*) compared to target words that are unrelated to it (e.g., furniture – '**cow**'; Balota et al., 1999; Faust et al., 2004). This facilitation arises through spreading activation to a small number of representations, making them easier to access. Facilitation through semantic relationships can also occur in connected speech/language contexts. For example, neural data obtained from language comprehension studies have reported that the processing of target words that are semantically related to words that listeners/readers expect to occur within specific contexts is facilitated through the semantic overlap between the two (e.g., Broderick et al., 2021; Federmeier & Kutas, 1999a; Federmeier, 1999b, Federmeier et

al., 2002, Federmeier, 2007, and Federmeier and Kutas, 1999). For example, healthy younger and older adults' processing of the word '**salt**' is facilitated after listening to *"I take my coffee with cream and.."* due to salt being a close semantic neighbour of '**sugar**'. The other effect of spreading activation between semantic representations is competition between co-activated words. For example, if '**cat'** and '**dog**' are active at the same time but only one word can be produced, the competition between the two can make the retrieval of required words slower, potentially due to the need for unrequired representations to be inhibited first. Evidence for this account is derived from the picture-word interference paradigm wherein the presence of semantically related distractor words creates delays in the production of target words (e.g., target: '**ball'**, distractor: '**frisbee'**) compared to when semantically unrelated distractors are present (unrelated distractor: '**hammer'**; (Taylor & Burke, 2002; but also see Glaser & Düngelhoff, 1984, Hantsch et al., 2009; La Heij et al., 1990; La Heij & Vermeij, 1987; Lupker, 1979; Schriefers et al., 1990). Furthermore, words that have a larger number of semantic neighbours are associated with delayed picture naming RTs as well as poorer naming accuracy (e.g., Fieder et al., 2019; Mirman, 2011, but cf. Hameau et al., 2019), likely since the coactivation of multiple representations interferes with targeted word retrieval.

Thus, the semantic connections between words have been argued to be a key part of word retrieval during language production. Such connections might play an even larger role during word production in context and in interaction with other speakers, where the relationships between and with preceding words can influence subsequent responses (Broderick et al., 2021; Federmeier & Kutas, 1999a; Federmeier, 1999b, Federmeier et al., 2002, Federmeier, 2007, and Federmeier & Kutas, 1999). This highly intricate process underlying fast and efficient word retrieval is believed to be influenced by a myriad of semantic, cognitive, and contextual factors that are subject to changes as age advances (see Chapter 3 for a review). Below, I first review studies examining age-related changes in language production, specifically the production of individual words without context, before further discussing age-related changes *in* context.

1.3 Declines in word retrieval speed in older age: naming studies

Traditionally, word production research has focussed on the retrieval of single words from memory (employing methods such as picture naming, verbal fluency, naming to definition, and question-answer tasks). One commonly used task is picture naming, which assesses

participants' production of pictured objects or actions. Many of these studies have shown declines in picture naming performance with age (e.g., Connor et al., 2004; Evrard, 2002; Gertel et al., 2020; Gordon & Kindred, 2011; Gordon & Kurczek, 2014; Kavé & Goral, 2017; Tsang & Lee, 2003, Wei et al., 2024; although cf. Alves et al., 2023). The advantage of these tasks is that they not only provide information about words that older adults can or cannot retrieve; but can also tell us about the speed of word retrieval during production when the word form is correctly retrieved. Studying both naming latencies and accuracy together can therefore provide deeper insights to language production changes across the lifespan.

Individual picture-naming studies have reached a range of heterogeneous conclusions. For example, while some have shown that older adults exhibit both poorer naming accuracy and longer naming latencies than younger adults (Gordon & Kurczek, 2014; Tsang & Lee, 2003), others have reported poorer accuracy but equivalent naming latencies to younger adults (Gordon & Kindred, 2011), while yet others have reported longer naming latencies, but similar accuracy to younger adults (Boudiaf et al., 2018; Ferre et al., 2020; Hoyau et al., 2017). Conversely, some studies have reported *higher* naming accuracy in older adults than in younger adults. This finding has been attributed to the type of pictures that are named, with older adults potentially benefiting from greater familiarity with rarer words such as "**yoke**", "**trellis**", "**palette**", and "**abacus**" than younger adults (Poon & Fozard, 1978; Schmitter-Edgecombe et al., 2000). Alternatively, other findings suggest that while accuracy may be poorer in older adults during the initial phase of picture naming, this can improve as the task progresses, suggesting that this group may just require more time to adjust to the task (Gertel et al., 2020).

In a review of 25 picture-naming studies with older adults, Goulet et al. (1994) concluded that this diversity in findings was due to variations in participant age ranges, study designs, task demands, and statistical techniques used in different studies. In addition to this, other factors have been linked to older adults' picture naming performance, including naming agreement (words with high name agreement are usually easier to produce, LaGrone & Spieler, 2006); the type of noun being produced (TOTs are more likely to be experienced for proper than object names or abstract words, Burke et al., 1991); and whether a word has near semantic neighbours (e.g., '**jam**' and '**jelly'**) as the existence of near semantic neighbours leads to slower production times and poorer accuracy for target words (Britt et al., 2016).

Collectively, however, the findings from picture naming data have demonstrated that in terms of accuracy, the most prominent declines occur after the age of 70 years (Feyereisen, 1997; Gordon & Kindred; 2011; Krethlow et al., 2024; Wen & Dong, 2023). In those with lower levels of education, this decline can be greater and begin earlier (at the age of 60 years, Wen & Dong, 2023). Word production latencies might be more sensitive to earlier changes in the ease and speed of word production. For example, picture naming latencies are shown to decline earlier, from 50 years of age (Verhaegen and Poncelet, 2013). These findings indicate that perhaps age-related declines in word production begin in middle adulthood (affecting production speed first), and as they progress into old age, coincide with an additional decline in accuracy.

1.4 Mechanisms underlying word retrieval changes with advancing age

Various cognitive and language mechanisms have been proposed to change with age, including processing speed (Salthouse, 2000); inhibitory control and semantic control (Hoffman, 2018); semantic knowledge and vocabulary (Kavé & Halamish, 2015); and working memory (Bopp & Verhaeghen, 2005). Some of these mechanisms will be discussed in more detail in the relevant Chapters (i.e., in Chapters 3 and 4) and below (section 1.5.4). Here, we will focus on how connections within the language system, semantic knowledge and vocabulary, and semantic control might change with age and can influence language production, as these proposed mechanisms are most relevant for my studies on word production in different types of semantic contexts.

The Transmission Deficit Hypothesis is one prominent account that captures some of the changes associated with ageing that affect word retrieval mechanisms (Burke et al., 1991, 2004; Burke & Shafto, 2004). It posits that with age, the connections between representations in the mental lexicon weaken, resulting in more frequent word retrieval problems. This theory further argues that the connections facilitating the retrieval of regularly accessed/ higher frequency words are strengthened by this repeated activation, while connections to less frequently recalled items are weaker, and are therefore more vulnerable to retrieval errors. In keeping with this, the Transmission Deficit Hypothesis has been particularly useful in explaining age-related increases in the frequency of TOTs (e.g., Brown and Nix, 1996), as well as the existence of stronger word frequency effects in old age (e.g., Burke et al., 1991; Burke et al., 2004).

The framework of the Transmission Deficit Hypothesis postulates that word retrieval is underpinned by connectivity between nodes comprised within three levels: semantic, lexical, and phonological. It argues that ageing weakens connections to phonological representation levels in particular, since *single* nodes connect lexical items to each of the phonological components within their structure (Burke et al., 1991, 2004; Burke & Shafto, 2004). The sparsity/weakening of these connections to phonological representations is argued by the Transmission Deficit Hypothesis to underlie the increase of tip of the tongue states in old age. In support of this notion, the processing of phonologically related words or syllables has been shown to improve/aid the verbal retrieval of target words which younger and older participants find difficult to retrieve, arguably through facilitating access to required phonological representations (James & Burke, 2000). While phonological connections might weaken, the vastness of the semantic system has been shown to grow with age, as a result of older adults' greater language experiences (e.g., Carrol, 2023; Kavé & Yafé, 2014; Verhaeghen, 2003). For example, older adults outperform younger adults on semantic knowledge tasks, which require them to select the synonyms (e.g., Hoffman, 2018; Hoffman et al., 2018) or meanings (e.g., Kavé & Halamish, 2015) of presented words. During word retrieval, older adults might therefore be able to leverage intact and richer semantic relationships to help them access the words they require. For example, a more developed semantic system could facilitate older adults' continued use of priming processes within conversational contexts. In doing so, concepts relating to those being discussed would also become active, potentially making it easier to retrieve words likely to be needed in those contexts. Indeed, some studies have suggested older adults benefit more from semantic priming than younger adults (cf. Laver & Burke, 1993). Additionally, interconnectivity between preserved semantic connections and lexical representations might help older adults to overcome the age-related weakening of connections within other parts of the lexical system, if excitation within semantic nodes creates a strong enough level of excitation within subsequent levels of processing.

Age-related increments in semantic knowledge might thus facilitate word retrieval processes. However, they could also *hinder* word retrieval processes if the accumulation of semantic representations results in the activation of multiple lexical items during word production and *increases competition* with the required word. Managing these co-active representations could be even more difficult for older adults considering the well-

documented age-related declines in semantic specific *and* more general inhibitory control processes (e.g., Hoffman, 2018; Hoffman et al., 2018; Spieler et al., 1996; West & Alain, 2000, also see Chapter 3 for a detailed review of these findings). These abilities facilitate the targeted retrieval of representations from the semantic store, as well as inhibiting competing representations or tasks during operations like word production (Hasher & Zacks, 1988; Jefferies, 2013). Within single-word production, some studies have also shown larger semantic interference in older than younger adults (e.g., Taylor & Burke, 2002). Declines in control processes paired with greater semantic knowledge might make it harder for older adults to efficiently retrieve the words they need while managing co-activated words (e.g., Ramscar et al., 2014), as doing so might first require the inhibition of the interfering representations. Furthermore, *accurate* lexical selection might be harder if there are multiple active items, increasing the likelihood of selecting the wrong word. Providing some support for these ideas, Hoffman and colleagues (2018) found a negative relationship between semantic knowledge and coherence in connected speech. However, a large-scale study suggested that greater semantic knowledge *aids* word production processes across the lifespan (although these mechanisms were found to weaken with increasing age; Shafto et al., 2016). This argues against the hypothesis proposed by Ramscar and colleagues (2014) that word-finding difficulties in older adults are due to increases in verbal knowledge creating more interference during word retrieval (at least during production, as tested in Shafto et al., 2016.

Within single word production tasks such as picture naming, the priming (facilitation) and competition (interference) outlined above may arise between target words and their semantic neighbours. Within everyday speech contexts though, priming and competition could arise from multiple sources. For instance, the words produced and/or processed by a speaker during ongoing conversation will trigger the activation of related words (through priming mechanisms), which may facilitate word retrieval of expected words but also compete with the word(s) a speaker is required to produce when those are unexpected. Furthermore, sentence contexts can allow speakers to predict upcoming words, again potentially resulting in facilitation when the expected word has to be produced but interference when another word is needed.

In the next part of this review, I will therefore focus on studies that have investigated language production changes across the lifespan when words are produced within connected speech tasks. This will be followed by a review of language comprehension studies that have

specifically focused on semantic relationships between words in context, finishing with a review of language production in alternation with comprehending another speaker's utterances.

1.5 Language production within connected speech

1.5.1 Age effects during language production within connected speech tasks

Moving on to the effects of age on language production processes within connected speech, Kavé and Goral (2017) reviewed 27 such studies. These studies used different types of tasks eliciting connected speech, which included participants' descriptions of personal events such as family vacations (Gould & Dixon, 1993), descriptions of their friends and family (Glosser & Desser, 1992); conversations with a partner (Arbuckle et al., 2000); or descriptions of short stories or series of pictures (Fergadiotis et al., 2011). In contrast to the studies reviewed above looking at the production of individual words, their review did not find strong support for declines in language production abilities in old age. Across the studies reviewed, word production was assessed in a range of different ways including word productivity (number of words produced; e.g., Arbuckle & Gold, 1993), lexical diversity (variety of words produced, and type-token ratio; Dennis & Hess, 2016; Fergadiotis et al., 2011), and word retrieval failures (Heller & Dobbs, 1993). Only two studies showed a reduction in the number of words produced by older adults in their speech (Heller & Dobbs, 1993; Kemper & Sumner, 2001), whilst fourteen studies contradicted this finding by showing that word productivity was not lower in older adults (Capilouto et al., 2016; Castro & James, 2014; Cooper, 1990; Kavé et al., 2009; Shewan & Henderson, 1988), or that word productivity was actually greater in older than in younger adults (Arbuckle et al., 2000; Arbuckle & Gold, 1993; Bortfeld et al., 2001; Dennis & Hess, 2016; Horton et al., 2010; James et al., 1998; Juncos-Rabadán et al., 2005; Saling, et al., 2012; Trunk & Abrams, 2009). However, increased word productivity could reflect easier/faster retrieval but could also reflect difficulties retrieving a specific word (therefore relying more on descriptions rather than specific words). Indeed, elsewhere in the literature, older adults have been shown to produce tangential, off-topic utterances during speech (Arbuckle and Gold, 1993; Glosser and Deser, 1992) and provide irrelevant information when telling a story (Juncos-Rabada´n et al., 2005; Marini et al., 2005) or describing an object (Long et al., 2018).

With regards to lexical diversity, Capilouto et al. (2016) found that when describing pictures, older adults (aged 70-89) produced a narrower selection of words compared to younger adults (aged between 20-69 years). Similar findings were reported in a recent study wherein younger and older participants produced verbal descriptions of the 'Cookie Theft' picture (Cho et al., 2021). In contrast, studies measuring type-token ratio have generally found that older adults produce a *greater* variety of words compared to younger adults (Fergadiotis & Wright, 2011; Horton et al., 2010; Kemper & Sumner, 2001). Again, these findings suggest mixed evidence for word-retrieval difficulties, with a greater variety of words potentially also being related to older adults' increased vocabulary and semantic knowledge (e.g., Hoffman, 2018) and/or reflecting strategies to avoid difficult-to-retrieve words.

There is also mixed evidence on age-related increases in word retrieval failures. Some studies have reported increases in measures that could reflect retrieval difficulties, including substitutions or erroneous word selection (Heller & Dobbs, 1993; Shewan & Henderson, 1988); word reformulations (defined as making changes to already produced words; Schmitter-Edgecombe et al., 2000); more use of fillers (Cho et al., 2021); and an increase in the use of pronouns, which may be indicative of difficulty in producing specific nouns (Cho et al., 2021; Heller & Dobbs, 1993; Mackenzie, 2000; Ulatowska et al., 1986). In addition, in Cho et al. (2021) older adults favoured words that were easier to access/produce (such as those which were more frequent and familiar to them) In contrast, other studies included in Kavé and Goral's (2017) review found no such age-group differences on related measures (e.g., Schmitter-Edgecombe et al., 2000), in terms of circumlocutions or digressions from the main topic (e.g., Cooper, 1990; Mackenzie, 2000; Schmitter-Edgecombe et al., 2000), or self-corrections (Cooper, 1990).

However, studies measuring local coherence (the degree to which consecutive sentences relate to each other in a meaningful way) and global coherence (the degree to which utterances relate to the wider conversational topic) have often highlighted declines in both with age (Glosser & Deser, 1992; Kemper et al., 2010; Marini et al., 2005; Wright et al., 2014). In addition to lower coherence perhaps reflecting strategic choices to avoid words that are difficult to retrieve, coherency during conversation might also reflect the use of inhibitory mechanisms, which are also needed for managing the competition between lexical representations during word retrieval. As a result, poorer coherence in older adults' speech may be linked to their lexical retrieval efficiency. Furthermore, somewhat in line with RT data

from picture naming studies, word retrieval *efficiency* has also been shown to decline within connected speech tasks with age. For example, older adults' speech rate is slower during discourse production comparative to younger adults' (Marini et al., 2025; Hilviu et al., 2025; Leeper & Culatta, 1995).

While the connected speech studies discussed above have not strongly supported the presence of word production difficulties in older adults, this may be related to the diversity and the nature of the variables studied. The many different variables studied make it difficult to compare findings across studies. Furthermore, several measures (e.g., producing a larger number of words) could reflect easier word retrieval/ production *or* strategies to avoid having to produce a specific word. Such strategies are only possible in connected speech studies, and not when participants must retrieve a specific word in response to, for example, a picture. Finally, naming times are a particularly sensitive measure of (early) age-related changes in word retrieval/ production, and this is typically not studied in connected speech studies. This further hinders comparisons across the connected speech and picture naming literatures.

The different findings observed across studies highlight the need for studies comparing age-related differences in word retrieval with and without context. However, beyond the presence of context (e.g., the difference between individual picture naming and connected speech), the content of the context might also play a crucial role. I will therefore now focus on the different types of sentence contexts that are of relevance to this thesis. These contexts vary in terms of their predictability of and match with the target word.

1.5.2 Sentential predictability: matching contexts

Sentential predictability/constraint refers to the degree to which a preceding sentence or speech utterance matches with/ prepares a speaker for a specific upcoming word(s). An example of a matched context is the question *"What is he woken up by every morning?"* since it predicts/ primes a single answer: "**alarm clock**", making it easier for the speaker to retrieve (Shao & Rommers, 2020; also see Chapter 2). On the other hand, a mismatching context primes/ predicts a different word to what the speaker needs to produce (e.g., *"What did the father read to his daughter before bed?"*- "**bible**", expected answer: "**fairytale**").

Matching sentence contexts can affect both language comprehension (e.g., Laszlo & Federmeier 2009; Kutas & Federmeier, 2011) and production (Bannon et al., 2025; Hustá et al., 2021; Papoutsi & Piai, 2023; Shao & Rommers, 2020). For example, Shao and Rommers

(2020) found that younger adults were faster to produce spoken words in matched contexts than in neutral contexts (which do not prime a specific word, e.g., *"What did he hear yesterday?"*- "**alarm clock**"). Furthermore, Bannon et al. (2025) found that younger adults were faster to produce words in matched contexts (e.g., *"Thanksgiving dessert consisted of ice cream and a pumpkin"*- "**pie**") than in neutral contexts (e.g., *"The artist painted a picture of a single"*- "**apple**").

The role of priming and prediction

Matched contexts are believed to facilitate the production of expected words through priming and/or prediction mechanisms. As discussed earlier, semantic priming is largely believed to be an automatic process through which words that are related to those that are encountered within a context rapidly become activated through spreading activation (e.g., Meyer & Schvaneveldt, 1971). Prediction on the other hand is a mechanism through which speakers form more specific expectations regarding upcoming words based on the context holistically (Corps et al., 2022, Experiment 1; Kamide et al., 2003, experiment 2; Pickering & Gambi, 2018). Prediction is argued to be closely related to, and use, automatic mechanisms such as priming (e.g., the "prediction-by-association" phase in Pickering & Gambi, 2018). However, researchers have argued that prediction also includes a largely non-automatic mechanism that might be cognitively demanding (an argument rooted in findings suggesting that prediction leverages cognitive resources such as working memory, e.g., Ito et al., 2018, experiment 2; Huettig and Janse, 2016; cf. Pickering & Gambi, 2018). This is also supported by the finding that participants become less efficient in implementing prediction when they are tasked with other cognitively demanding tasks such as speech encoding, while spreading activation processes are not affected by this (Ito et al., 2016). However, the exact nature of prediction remains an open question, with some findings suggesting that prediction *can be* automatic. For example, even when participants are asked to ignore sentence contexts, they cannot control the resulting facilitatory or interfering effects of them on the processing/production of subsequent words (Fischler & Bloom, 1979).

During prediction, speakers are believed to draw upon available linguistic and non-linguistic information, including their knowledge of their previous language experiences, and characteristics of their partner (such as their habits, likes and dislikes). If a conversational partner said: "*I noticed it was raining so I put up my* "; processing of the content word "**raining**"

would prime multiple related words such as: "**cloud**", "**puddles**", and "**umbrella**"; however, the listener would likely expect the word "**umbrella**" to occur next in this context. This is based on the listener likely hearing this word occurring within this context before. Furthermore, additional information such as the verb "*put up*", and perhaps having previously seen their partner using an umbrella, could collectively support the speaker in predicting this word.

One way in which predictions during comprehension are thought to form is *through* the language *production* system (Pickering & Gambi, 2018). It is argued that when listening to speech or processing written discourse, listeners/readers create a representation of the linguistic form of the utterance they hear/sentence they read, which also includes information about the speaker's communicative intent. The representation is then run through the listener's/reader's own language production system, setting into motion the activation of the relevant pathways. Consequently, this triggers lexical items that are usually used in those contexts, resulting in prediction of those items. If prediction during comprehension truly does operate through the language production system, this could affect subsequent language production processes (i.e., participants' planning and formulation of their own response for and during their subsequent conversational turn). The need to both comprehend and produce language within natural conversation, whilst drawing on potentially effortful prediction mechanisms, could strain key resources. With these cognitive resources argued to decline with age, older adults might continue to benefit from semantic priming but perhaps not engage in predictive mechanisms as effectively as younger adults do (cf. Federmeier et al., 2002, as discussed further below).

Age and matched contexts

Most of the existing research that has studied potential age-related differences in terms of how matched contexts influence word processing has done so within language comprehension studies (Kliegl et al., 2004; Kutas & Federmeier, 2000; Kutas & Federmeier, 2011; Laszlo & Federmeier, 2009; Wlotko, et al., 2012). The effect of age has presented differently across studies, however, (as is demonstrated below) for behavioural studies (presented first) and EEG studies (presented subsequently).

Behavioural studies have generally demonstrated a preservation/increase in the use of matched sentence contexts in old age. In Haigh et al. (2022), we found that both younger and older adults read target words in matched contexts more quickly (e.g., "*the man watched*

the lava erupt from **the volcano**") than in neutral sentence contexts (e.g., "they went to see **the volcano**"). Age group did not interact significantly with regards to the match effect (the mean difference in reading times between these conditions), suggesting that older adults leveraged a similar degree of benefit from the matched contexts as younger adults did. In a related study, Kliegl et al. (2004) found that sentential predictability facilitated younger *and* older adults' eye movements, but in different ways. For example, younger adults skipped words which were predicted by the context more frequently (than they did in low predictable contexts), while older adults fixated on them less frequently (in high than low predictable contexts). Other findings suggest that older adults harness matched contexts to a *greater* degree than younger adults do (Pichora-Fuller et al., 1995; Rayner et al., 2006). For example, older adults derived greater benefit from matched contexts in identifying sentence final words when listening to speech in babble, compared to younger adults (Pichora-Fuller et al., 1995). Rayner et al. (2006) also found that older adults' eye movements were more strongly influenced by sentence predictability.

Conversely, studies measuring neural activity using EEG (electroencephalogram) suggest older participants are *less* effective in deriving benefit from matched contexts in comparison to younger adults. For example, in some studies older adults have been shown to exhibit a *reduced* effect of matched contexts on participants' ERPs (Event Related Potentials; Federmeier et al., 2002, Federmeier & Kutas, 2005; Gunter et al., 1992; Hamberger et al., 1995; Wlotko & Federmeier, 2012; Woodward et al., 1993). For example, younger adults' N400 response (the negative deflection which occurs in the EEG waveform around 400ms after the onset of a stimulus, Kutas & Federmeier, 2000) is significantly reduced when processing words predicted by/ matched with the sentence context. In older adults' this response is smaller and slower (Ford et al, 1996; Wlotko & Federmeier, 2012). In addition, while younger adults also demonstrate facilitation (i.e., reduced N400s) for the processing of words semantically related to predicted words, older adults do not always show this (Federmeier et al. 2002), suggesting that they might also utilise semantic priming mechanisms less efficiently during language processing (these findings are discussed in more detail in Chapter 5).

1.5.3 Sentential predictability: mismatching contexts

Unlike matching sentence contexts, which facilitate word production mechanisms, *mismatching* contexts (that predict another word than the target) might have the opposite

(interference) effect on word production efficiency. The reasoning behind this is that such contexts might conflict with words predicted/primed by the context. Such conflict might arise if a speaker expects to produce a particular word based on the sentence context, but instead they are required to produce a different, unexpected word. Delays in word production in such contexts might be caused by speakers needing to inhibit the active/expected word, as well as retrieving the required one. For example, consider the sentence: *"These china cups are beautiful, please could you pour me some"- "***lemonade**" The most expected ending of this sentence is: '**tea**', however, realising tea isn't available on the menu, a speaker may decide to ask for some lemonade instead. To do this, the speaker would need to juggle the demands of inhibition (i.e., "**tea**") and activation (i.e., "**lemonade**").

In contrast to the literature surrounding matched contexts, explorations of the effects of mismatched sentence contexts have returned mixed findings (in younger adults). While some studies have shown interference effects (e.g., Bannon et al., 2025), other studies show no impact of unexpected sentence-target relationships on word processing (e.g., Haigh et al., 2022; Luke & Christianson, 2016). Recently, investigating word production in younger adults, Bannon et al. (2025) reported that picture naming latencies following mismatched contexts were affected by the degree of mismatch between a presented sentence context and the naming target. Slower naming times were observed following strongly mismatched sentences (e.g., ***"****the frontal lobe is an important part of the"*- **"cake**") than moderately mismatched ones (e.g., "*the children all sang happy birthday before they cut the"- "***pie**", expected answer = '**cake**'), with responses being fastest in neutral sentences (e.g., "*the artist painted a picture of a single"- "***apple**"). One question that currently remains to be addressed is how do mismatched contexts affect *older adults'* word production?

Age effects and mismatched contexts

In terms of age-group comparisons, in Haigh et al. (2022) we did not observe significant delays in younger nor older adults' reading times of words placed in mismatching contexts (e.g., "*the swimmer dived into"- "***the volcano**") in comparison to neutral contexts ("*they went to see"- "***the volcano**"). Similarly, Yoon et al. (2015) asked younger and older adults to judge the plausibility of presented sentences. Both groups exhibited equally poorer performance for mismatched/implausible constructions (i.e., "*because the ceiling light is on, the room is* ***dark***")

than matched/ plausible constructions (e.g., "*because the ceiling light is off, the room is* ***dark***").

ERP data obtained from younger participants has shown that the degree of match/mismatch between sentence contexts and their endings influences younger participants' N400 responses in a graded manner. The smallest N400 amplitudes are observed for strongly/the most predictable endings, and increase as a result of decreasing predictability/match with the context (Federmeier et al., 2002; Federmeier & Kutas, 1999b; Kutas & Hillyard, 1984, Thornhill & van Petten, 2012). With regards to the effects of age, Federmeier et al. (2002) found that older adults too exhibited larger N400s when listening to mismatched contexts (e.g., "*They wanted to make the hotel look more like a tropical resort. So, along the driveway they planted rows of* **pines/tulips**") than when contexts were paired with more expected endings, (e.g., '**palms**'). However, as discussed previously, this effect is smaller and somewhat delayed in comparison to the younger adult group.

One key point to consider, however, is that many (comprehension ERP) studies discussed here have typically compared word processing times in matched and mismatched contexts without a neutral baseline. This makes it difficult to disentangle the potential facilitatory and inhibitory effects of different types of contexts, as well as any age-group differences in terms of facilitation (through priming and/or prediction) and interference (through potential changes in control and inhibition).

In sum, some (matched) sentence contexts facilitate the retrieval of specific words. Leveraging benefit from such contexts during conversation is believed to be underpinned by the mechanisms of prediction and/or priming. Some research findings suggest that similarly to younger adults, older adults may be able to derive support from such contexts in helping them to produce primed/predicted words (since semantic priming mechanisms remain intact in old age). In contrast, prediction (which is more cognitively demanding) may be more vulnerable to the effects of ageing, since older adults may seek to preserve vital resources for use in other essential conversational processes. Consequently, older adults may derive *less* benefit from matched sentence contexts than younger adults. Indeed, the comprehension studies reviewed generally suggest older adults benefit from matching sentence contexts too, but vary in their conclusions about ageing strengthening or weakening such context effects.

In contrast to matched sentence contexts, *mismatched* contexts may give rise to competition by priming/ predicting specific words conflicting with the words a speaker needs

to produce. Due to declines in their semantic/ cognitive control, older adults may be expected to be hindered more by such contexts. However, previous research has reached differing conclusions regarding whether such contexts hinder the production/ processing of target words in either younger or older adults.

Much of the work looking at different types of sentence contexts has focused on comprehension rather than production. Furthermore, it has not always been able to tease apart facilitation (matching) and interfering (mismatching) sentence contexts in the absence of baseline conditions. Therefore, further work is needed to understand when such contexts create facilitation or interference within language production, if older adults are indeed more affected by them, and which cognitive and language abilities might contribute to sentence contexts. This therefore will be the focus of Chapters 2, 3, and 5.

1.5.4 Sentence processing

As discussed above, most studies looking at the effects of different types of (matching and mismatching) sentence contexts on word processing have focused on comprehension of spoken or written words. Production studies related to ageing have typically focused on either production of individual words, or on people's own connected speech. To examine the role of sentence context, I take a different approach in my thesis chapters by studying the influence of a preceding sentence produced by another speaker on older and younger adults' own word retrieval during production. This allowed me to use reaction times (onset of word production) as a measure of speed of word retrieval. As discussed above, one advantage is that reaction times might be sensitive to age changes from a relatively young age onwards. It furthermore allows us to study confrontational naming in a way that can be compared with the existing literature on picture naming. At the same time, this approach allowed me to manipulate the preceding sentence context (i.e., whether that context predicted the target or not), something that cannot easily be achieved with the same level of experimental control in traditionally used connected speech paradigms (Kavé & Goral, 2017).

One of the key differences between production of individual words versus having to retrieve those words in response to someone's question or sentence is that during the latter, speakers are simultaneously tasked with *comprehending* the speech produced by their conversational partner(s). Broadly speaking, successful sentence processing and comprehension is thought to depend on multiple streams of incoming information contained

in the verbal signal, relating to syntax, prosody, and visual and linguistic analysis (Ferreira & Clifton, 1986; Ferreira & Cokal, 2016; Frazier & Fodor, 1978; Rayner et al., 1983; Seidenberg & McClelland, 1989). Some of the processes associated with sentence processing might change with age, for example potential decreases in the cognitive resources (e.g., verbal short-term working memory) needed for navigating conversational interactions. A wealth of research has demonstrated age-related declines in verbal short-term working memory (Beese et al., 2017; Bopp & Verhaeghen, 2005; Salthouse, 1994). During speech processing, this can influence comprehension in various ways. Focussing on sentential parsing and the resolution of grammatical dependencies within speech signals, older adults experience difficulties in comprehending sentences that contain a distance (number of words; Gibson, 1998, 2000) between meaningful grammatical units needed for resolving grammatical dependencies. An example of this type of sentence is those containing relative clauses: clauses attached to an antecedent using a relative pronoun such as 'which' or 'whom' (e.g., *"He helped the man whom the wolf attacked."*). The Dependency Locality Theory (DLT; Gibson, 1998, 2000) posits that two working memory components are paramount for successful sentence comprehension: storage and integration. Storage involves holding recently processed syntactic information in the working memory store, until the remaining information needed for the grammatical dependency to be resolved is available. Furthermore, integration involves adding incoming information to this held information. In line with this, Liu and Wang (2019) found that both younger and older adults' comprehension accuracy declined when this distance within sentences increased. Furthermore, there was a significant interaction between distance and age group in comprehension accuracy. Distance disrupted older adults' comprehension more than younger adults.'

Literature showing that working memory is a key component of sentence processing, and that working memory declines with age, highlights the need for further exploration of how this decline can influence older adults' own production in response to other people's speech. In Chapter 4, I therefore address the question of whether placing greater demands on working memory resources during sentence comprehension (in the form of more complex sentence processing) affects subsequent language production processes.

1.5.5 Response planning during conversations

In addition to processing their partner's speech, within conversations speakers are also tasked with formulating/planning their own verbal responses. Crucially, this does not only require retrieval of words for one's own production but also timing when responses can be given (i.e., turn taking; De Ruiter et al., 2006; Magyari et al., 2014; Riest et al., 2015). There are currently two differing viewpoints on when speakers start planning their responses to other people's speech within conversational contexts. On the one hand, some findings suggest that speakers postpone speech planning right up until their conversational turn approaches (termed *late planning*). For instance, studies adopting dual task paradigms (wherein participants engage in conversational turn taking whilst taking part in a non-linguistic task) have shown participants' performance on the non-linguistic tasks declines close to when their turns occur in the conversation (Boiteau et al., 2014; Sjerps & Meyer, 2015). Boiteau et al. (2014) asked participants to track a visual target on a computer screen using the computer mouse whilst engaging in spontaneous conversation with a partner. They found that their performance on the tracking task decreased just prior to and during their conversational turns (cf. Sjerps & Meyer, 2015, for similar findings). These findings are taken as evidence that speakers don't start planning their responses up until just before they take their turns in conversation. This implies that comprehension and production occur sequentially; and thus, there is little overlap between the operations involved in both tasks.

In contrast, other research suggests that speakers begin preparing the speech for their upcoming turn as soon as they can (whilst their partner is still taking their turn in the conversation; termed *early planning)*. This notion is somewhat more aligned with how quickly natural conversation tends to unfold since planning speech earlier likely supports conversing partners in keeping up with the pace of the interaction. One mechanism which might aid speech planning is prediction; for example, Pickering and Garrod (2013) argue that *both* speakers and listeners during conversational turns predict both their own, and *each other's* upcoming utterances, helping them to achieve the smooth alternating transitions between the roles of speaker and listener that are typical of fluent conversation.

Evidence for early planning accounts is received from ERP studies showing that planning processes for the next conversational turn begins as soon as a person has acquired enough information from their partner's speech signal to formulate an appropriate response (Bogels et al., 2015a; Bogels et al., 2018). Bogels et al. (2015a) and Bogels et al. (2018)

compared participants' ERP responses while they answered questions comprising critical information that occurred at different places within the sentence context. For example, participants were asked "*which character, also called **007**, appears in the famous movies*?" or "*which character from the famous movies, is also called **007***?" expected answer: '**James Bond**'. Participants' ERP data indicated that they began planning their responses as soon as they had encountered the critical information (i.e., '**007'**). This was demonstrated by a large positivity starting around 500ms after the onset of the critical word in language production areas in the brain, as well as a decrease in alpha power occurring in occipital and parietal areas. While the former effect was interpreted as directly depicting response planning mechanisms, the latter was interpreted as an attention switch from the spoken input towards language production planning. Crucially, these processes started as soon as the critical information was presented, suggesting speakers plan their responses while still processing the input they are receiving.

If conversationalists truly do engage language comprehension and production mechanisms in parallel as this branch of research suggests, then this raises questions regarding how cognitive resources are managed/split between these two operations; and if performing both tasks at the same time poses a cost for production. Previous work has highlighted that language comprehension and production processes engage many of the same brain regions and processes (Menenti et al., 2011; Segaert et al., 2012); and that both require the allocation of central attention (Jongman, 2021; Shitova et al., 2017). Furthermore, previous work has indeed suggested that engaging in the complex mechanisms underlying ongoing language comprehension tasks can interfere with language production processes. For example, language production in younger adults has been found to slow down when distracting words are played in the background (e.g., Schriefers et al., 1990) and when having to undertake a visuospatial or verbal working memory task at the same time (e.g., Klaus et al., 2017). Although these studies use paradigms that differ from natural conversation wherein speakers are comprehending and integrating their partners speech, these studies suggest that the cognitive processes involved in comprehension can influence one's production. With these cognitive control and working memory abilities changing in old age, it is important to investigate how these changes manifest in language production processes in the latter portion of the lifespan. In Chapter 4, I therefore examine how information presented at different points in the preceding sentence context (allowing early or late planning) as well as in

sentences varying in their syntactic and processing demands influence word production in younger and older adults.

1.6 Rationale for empirical work and overview of thesis chapters

The literature reviewed above forms the basis for the research carried out in this thesis. The general overview from previous studies is that word retrieval/production difficulties (i.e., difficulties in the retrieval of individual words from the mental lexicon, as measured by for example picture naming performance) worsen with age. One of the key questions that remains to be addressed is: do these difficulties also present within older adults' day to day conversations? It is important to address this to shed light on the language experiences of older adults, and to better understand how to support them. This thesis takes a first step by examining effects of another speaker's sentence context on word retrieval during production in older and younger adults. Producing words in conversation might place more demands on vital cognitive resources due to the multifaceted nature of this task (i.e., formulating and planning one's own responses while processing and comprehending speech produced by one's partner). The underlying processes involved in producing words in context versus producing them in isolation could therefore also be considerably different and might be shaped by the nature of the preceding sentence context. Exploring how older adults manage word production during these simultaneous processes compared to younger adults could therefore provide insight into the various cognitive and language-related mechanisms that underly word retrieval and any age-related changes.

An important first step towards understanding the potential impact of these mechanisms on conversations is better understanding word retrieval in different types of sentence contexts that are produced by another speaker and that older and younger adults need to respond to. As is shown in the literature review, behavioural studies with younger adults have found that different types of sentence contexts can have significant impacts on word production efficiency. One of the key goals of this thesis then is to try to understand how different contexts affect word production within a framework of healthy ageing. Matched/predictive and mismatched/conflicting sentence contexts can have facilitatory and interfering effects on younger adults' production processes respectively. What remains to be answered is how older adults' production is affected by these contexts, considering the

changes they have been argued to undergo in terms of their (increasing) semantic knowledge and (decreasing) semantic and general cognitive control.

In this thesis, Chapter 2 therefore firstly examines the effects of different types of sentence contexts (predicting a specific target word or another word than the target response, as well as a neutral context) on word production in younger adults. I also compare these context effects to producing words in isolation, in groups of healthy younger and older adults. In this chapter, I used a variation of the picture naming paradigm wherein participants are naming a set of pictured objects, either in response to a series of questions or in isolation (no context). I compared three types of contexts: matched (predicting the target response), mismatched (predicting another word than the target), and a neutral baseline. This allowed me to study both facilitation and potential interference relative to a neutral baseline. While facilitation of producing primed words in matching contexts is believed to arise from intact semantic networks (which facilitate priming/prediction), interference from mismatching contexts is believed to arise from difficulties in managing co-active representations during target word retrieval. As these underlying skills change with age, this chapter studied how this affects the resultant facilitatory and inhibitory effects of sentence contexts during language production. This in turn will increase our understanding of if older adults continue to benefit from priming and prediction forming (underlying facilitation in matched contexts) as well as if they experience relatively more difficulty in suppressing unwanted representations (underlying interference in mismatched contexts). Furthermore, the comparison between neutral context and no context allowed me to address the question of whether word production processes differ within these two types of environments. In addition, it allowed me to bridge the gap between work looking at production in sentence contexts and the existing body of work focussing on word production in isolation (within picture naming studies).

Benefiting from priming and/or prediction would require intact semantic knowledge and vocabulary. Similarly, difficulties with managing interfering information during language production is expected to be related to one's semantic and inhibitory control abilities. In Chapter 3, I therefore extend on the findings from Chapter 2. Participants who took part in the first study were invited to take part in a battery of cognitive tasks that measured: semantic knowledge and control, inhibitory control, letter and semantic fluency, and working memory. For the reasons discussed in the literature review, I was most interested in how semantic

knowledge and verbal fluency link to the facilitatory effect of matched contexts; how semantic and inhibitory control link to interference within mismatching contexts; and how working memory links to the difference in retrieval within neutral contexts compared to in isolation. Studying the role of these various cognitive and language abilities (in younger and older adults) allows us to examine the various mechanisms that have been proposed to underlie word production processes, and age-related changes in these processes.

Expanding on sentence contexts facilitating word production, in Chapter 4, I present the findings from an empirical study examining when speakers start planning their response and how they are influenced by the complexities of sentences produced by their conversation partner. Healthy younger and older adults produced individual words which they believed were the correct endings of a series of verbally presented sentences. The sentences participants heard were designed to prime/predict low frequency target words as previous studies (e.g., Burke et al., 2004; Cohen & Faulkner, 1984; Lovelace & Twohig, 1990) have shown older adults experience greater difficulties in retrieving these words. Since lower frequency words are typically less imageable than higher frequency ones (and thus less-suited to a picture naming paradigm), the sentence completion paradigm used provided the opportunity to target words more likely to be affected by ageing. In this chapter, I firstly examined how word production processes are affected in situations where the demands imposed by the context are higher (i.e., when the sentences are longer/more complex) compared to when they are lower (i.e., when contexts are shorter/simpler). This comparison allowed me to study a potential influence of sentence complexity on subsequent word production as well as the potential role of working memory, particularly in older adults. Second, in this chapter, I also investigated the differences in younger and older adults' word production in contexts where semantic primes critical for predicting upcoming words were presented earlier compared to when they were presented later in the preceding context. This examined when younger and older adults start planning their responses.

Chapter 5 is the final empirical chapter of this thesis. Adopting a similar picture naming paradigm to what was used in Chapter 2, I tested the effects of different types of mismatching context-target relationships on word production. Specifically, I was interested in exploring if the magnitude of the interference created by such a context is directly related to the degree of mismatch between the context and target word. I addressed whether a stronger/greater mismatch creates more interference compared to a smaller mismatch, in terms of word

production times. This was to explore whether a graded interference based on mismatching contexts, similar to what has been observed neurally (particularly within younger populations), also presents behaviourally and influences word production in older adults. I was also interested in whether an unexpected word's semantic relatedness to the context or to the expected word facilitates *or* interferes with the production of an unviable sentence ending. To better understand the underlying mechanisms and the role of semantic control during ageing, I also examined semantic control and its relationship with sentence-context interference costs during word production.

In Chapter 6, I present a summary of the key findings from the research presented in this thesis, contextualised within the wider body of literature in this area. This chapter also outlines some of the key contributions of the empirical work, as well as questions left open by it, which should be addressed to allow us to move towards an even more complete understanding of language production within naturalistic speech contexts.

1.7 Conclusion

At the heart of this thesis lies the aim of elucidating the effects of differing everyday sentence contexts on word production processes. The work presented is rooted in previous findings showing that word retrieval becomes harder with advancing age. However, I work towards trying to understand production processes within more naturalistic conversational contexts; keeping in mind the additional demands faced by the speaker during this process. I also take into consideration the influence of various language and cognitive abilities (that develop and change as people age), and investigate how these abilities affect how older adults are influenced by different types of sentence contexts in comparison to younger adults. Taken together then, through the empirical work I carry out in this thesis, I try to increase our understanding of how ageing affects language production processes within older adults' day to day interactions.

Chapter 2: Picture naming in conversational contexts

This chapter is adapted from the following published article:

Hanif, N., Jefferies, E., & de Bruin, A. (in press). Naming speed during language production in younger and older adults: Examining the effects of sentence context. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218241309602

Abstract

Word retrieval during speech production has been found to slow down with age. Usually, words are produced in sentence contexts. The current studies examined how different sentence contexts influence word retrieval/ production in younger and older adults. 48 younger and 48 older adults named pictures that were preceded by a matched context (which predicted that specific target word), a mismatched context (predicting another word), a neutral context (that did not predict one specific word), or no context. In comparison to the neutral context, both younger and older adults' word production was faster in matched contexts, suggesting both age groups benefited from sentence contexts facilitating the retrieval of predictable words. Neither age group was slowed down by the mismatched contexts (compared to the neutral contexts), suggesting these contexts did not create (sufficient) interference to hinder word retrieval/ production.

2.1 Introduction

The efficiency (speed) with which words are retrieved from memory declines with age (Verhaegen & Poncelet, 2013). In the existing literature, this phenomenon has most commonly been studied when words are produced in isolation (e.g., during picture naming). However, in everyday conversation, language is often used to respond to other people, for example when they ask you a question. The sentence context (e.g., the nature of the question asked by our conversational partner) might influence how quickly we can retrieve the words we need to respond to them. This chapter therefore focusses on word retrieval efficiency in younger and older adults, in response to different types of sentence contexts.

2.1.1 Ageing and word production

Picture naming tasks often show that older adults (typically defined as aged 65 years or older) have more difficulty naming objects compared to younger adults (e.g., Barresi et al., 2000; Connor et al., 2004; Goral et al., 2007; MacKay et al., 2005). While accuracy can sometimes be preserved in older adults (Boudiaf et al., 2018; Ferré et al., 2020; Hoyau et al., 2017; Schmitter-Edgecombe et al., 2000), or only show age-group differences after the age of 70 (Wen & Dong, 2023), naming times appear more sensitive to age-related changes at a relatively earlier age (e.g., Ferré et al., 2020; Hoyau et al., 2017). For instance, slower naming compared to younger adults has been observed from the age of fifty (e.g., Verhaegen & Poncelet, 2013). Indeed, while the vocabulary itself (i.e., the knowledge) remains intact or even increases with age (e.g., Hoffman, 2018), it is the (speed of) *retrieval* of words that appears impacted. This can affect low-frequency words in particular (e.g., Ferré et al., 2020).

These changes in word retrieval/ production can be explained through several cognitive changes generally observed with ageing, including general age-related slowing (e.g., Salthouse, 1996) and difficulties suppressing goal-irrelevant information (e.g., Hasher & Zacks, 1988). Focusing on language specifically, the Transmission Deficit Hypothesis explains these findings through weakened connections between representations at different levels within the ageing lexicon (Burke et al., 1991). While semantic information about words can remain intact, this account argues that weaker connections from the lexical to the phonological level can result in poorer word retrieval/ production in older adults.

2.1.2 Language production in sentence contexts

Daily-life word retrieval during production typically takes place in context, including sentences and interaction with other people. Studies examining connected speech in older and younger adults have shown mixed findings regarding age-related differences. Some studies report that older adults perform more poorly in connected speech tasks than younger adults do (e.g., older adults produce fewer words and make more word-choice errors, Heller & Dobbs, 1993). Other studies have reported similar performance in older and younger adults on several measures (e.g., in terms of the number of words produced and disfluencies during neutral picture descriptions; Castro & James, 2014) or show that older adults produce more words than younger adults during connected speech (e.g., Arbuckle et al., 2000; see Kavé & Goral, 2017 for a review). Based on their review, Kavé and Goral (2017, p.521) concluded that *"there*

is little evidence for significant word retrieval deficits in connected speech production in healthy aging." Furthermore, performance on picture naming and connected speech tasks does not always correlate (e.g., Saling et al., 2012; Schmitter-Edgecombe et al., 2000). This might in part be related to the different measures used. Connected speech studies have used a range of variables, including number of words produced, lexical diversity (variety in the words used), word substitutions and circumlocutions (suggesting failed target-word retrieval), coherence, and dysfluencies (e.g., Kavé & Goral, 2017; Hoffman et al., 2018). Picture naming tasks, however, often measure *speed* of word retrieval during production (naming times) and/or accuracy. These measures might tap into different aspects of language production and thus make it difficult to directly compare retrieval in context (connected speech) versus in isolation (picture naming). Naming times measured during picture naming might be most sensitive to detecting earlier and smaller changes in word retrieval during production (e.g., Verhaegen & Poncelet, 2013). Furthermore, some measures used in connected speech might partly assess *compensatory strategies* (Kavé & Goral, 2017), for example in the form of circumlocutions that can mask retrieval difficulties (e.g., Nicholas et al., 1985). The various measures used across studies hinder a direct comparison, making it difficult to evaluate whether age-related effects on word retrieval during production are truly influenced by the presence of context. In this chapter, we therefore assessed speed of picture naming (naming times) when pictures were presented in isolation and when they were preceded by a sentence context.

2.1.3 Matched and mismatched contexts

As well as the *presence* of context, the *type* of context might influence word retrieval during production. We now turn our attention to different types of sentence contexts and the (facilitatory or inhibitory) effects they might have on language production. Words are often retrieved faster when preceded by contexts with lexical-semantic information that is matched with the target word. For example, following the question "*what is he woken up by every morning*?" the picture of an '**alarm clock**' is named faster than after a semantically neutral question such as "*what did he hear yesterday*?" (Shao & Rommers, 2020). Upon hearing or seeing a word, activation of its semantic and/or lexical features can spread activation to neighbouring representations that share features or associations. This priming can facilitate production of related words, compared to unrelated words that do not share semantic

features. This has also been linked to prediction, with listeners argued to predict upcoming words, potentially through pre-activation of specific lexical features that are likely to appear (e.g., Federmeier et al., 2007). In contexts where those predictions are accurate, this could facilitate the speaker's own production.

However, in daily-life speech and language, we often also encounter words that are not highly predictable (Luke & Christianson, 2016). Contexts that are mismatched with a target word (i.e., where the prediction is incorrect) can result in processing costs (e.g., Federmeier et al., 2003, but cf. Luke & Christianson, 2016) due to them activating words that compete with required ones for selection. Within language production models, 'competitors' refer to lexical representations that are characterised by overlapping semantic or phonological connectivity to target words (e.g., words that share the same meaning as the target: 'sofa' – 'couch', 'jam'- 'jelly'; are close semantic neighbours of it such as 'cat' – 'dog'; or are phonologically similar such as 'cat'= 'cap'; Zhang et al., 2025). However semantic representations that are not direct competitors may still compete with eachother if the context strongly predicts them (e.g., consider the following: "*the children all sang happy birthday before they cut the* **pie**", expected answer = '**cake,**' Bannon et al., 2025). This competition can slow down language production, although a recent study (Bannon et al., 2025) interestingly suggested this might especially be the case when the unexpected words are not related to the predicted target word at all.

2.1.4 Ageing of cognitive and semantic processes

The semantic relationships between sentence contexts and words that need to be produced can modulate age-related differences in word retrieval as older adults have shown poorer semantic control, the mechanism through which intended representations (semantic knowledge) are retrieved from the semantic store while competing representations are supressed (Hasher & Zacks, 1988; Jefferies, 2013). Older adults' semantic knowledge is often reported to be comparable to, or even larger than, younger adults' knowledge (e.g., Carrol, 2023; Hoffman, 2018; Hoffman et al., 2018; Kavé & Halamish, 2015; Kavé & Yafé, 2014; Verhaeghen, 2003). In contrast, semantic control and inhibitory control diminish with age (e.g., Hasher & Zacks, 1988; Hoffman, 2018; Spieler et al., 1996). The term "inhibitory control" is often used as a general term to refer to control over different types of information, both linguistic/verbal and non-linguistic/non-verbal. Semantic control specifically refers to control

over irrelevant semantic representations (e.g., distractor words that are semantically related to the target while having to make a size judgement). Preserved semantic knowledge may allow older adults to continue leveraging matched contexts to aid the retrieval of primed/predicted words. On the other hand, age-related weakening of semantic and/or inhibitory control can create difficulties in suppressing responses and information irrelevant to the task at hand. Both semantic and inhibitory control have been found to predict some aspects of word production (Arbuckle & Gold, 1993; Hoffman et al., 2018; Yin & Peng, 2016; but see Higby et al., 2019). Age-related changes in semantic networks and control processes may thus have implications for word retrieval processes in everyday conversation.

2.1.5 Ageing and matched contexts

Previous literature has mostly assessed the role of sentence context in older adults through comprehension studies, with little work assessing production. Production has mostly been studied through priming paradigms manipulating the relationship between an individual prime word and target word (e.g., **doctor** - **nurse**). Both younger and older adults show faster naming when target words are preceded by a semantic prime compared to a neutral prime (Balota et al., 1999; Faust et al., 2004), with some findings suggesting older adults may benefit even more from semantic priming than younger adults (see Laver & Burke, 1993 for a review). The *comprehension* studies that have looked at word processing in sentence contexts have reached a range of conclusions. Some have found that both younger and older adults use sentence context to facilitate retrieval of upcoming words to the same extent (e.g., Kliegl et al., 2004, using eye tracking). Other findings suggest that older adults can utilise semantic cues to a *greater* degree than younger adults (e.g., Pichora-Fuller et al., 1995; Rayner et al., 2006), lending support to the hypothesis that older adults can utilise semantic context to help them to overcome age-related declines (see also Rayner et al., 2006; Speranza et al., 2000). However, other studies suggest that older adults do not use semantic information within sentential contexts as effectively as younger adults do. These findings are often shown through EEG data examining N400 effects, a negative-going wave peaking approximately 400ms after stimulus onset. This N400 effect is often reduced in amplitude in congruent sentence contexts, but such sentence context effect has been observed to be smaller or delayed for older than younger adults (e.g., Federmeier et al., 2002; Wlotko et al., 2012).

Together, these findings suggest older adults can continue to use semantically congruent information during language processing, although it remains unknown whether they can benefit more or less than younger adults. These findings are based on studies looking at language comprehension (e.g., sentence processing), leaving it also largely unknown how congruent ('matched") sentence contexts can influence language production in older adults.

2.1.6 Ageing and mismatched contexts

While the focus has been on facilitation stemming from semantically congruent information, in contexts that are incongruent with a target word, older adults might experience more interference based on diminished inhibitory and/or semantic control. Studies looking at semantic interference at the individual word-level (e.g., naming the picture of a '**ball**' while seeing the word '**frisbee**') have shown both younger and older adults show slower naming in the presence of this distractor than when they see a neutral word. This semantic interference effect is often greater within older adults than in younger adults (e.g., Taylor & Burke, 2002), suggesting that older adults have more difficulty in inhibiting semantic distractors. However, this is not found across all studies, with Lorenz et al. (2019) showing a similar impact of semantic distractors on younger and older adults. Some ERP studies, finally, have suggested that older adults might be influenced less by incongruent sentence contexts as a consequence of predicting upcoming words less or less successfully (e.g., Wlotko et al., 2012). However, these studies often compare unexpected, incongruent words to expected, congruent words without a neutral baseline. This makes it difficult to disentangle effects of (potentially facilitating) congruent contexts and (potentially interfering) incongruent contexts.

2.1.7 Current study

The current literature thus has shown mixed effects regarding age-related differences in terms of word production in context. It has furthermore focused on comprehension and, in the absence of a neutral baseline, often does not allow for a comparison between semantically congruent (matched) and incongruent (mismatched) contexts. In this study we used a picture naming paradigm similar to Shao and Rommers (2020), who showed faster naming in younger adults when naming a picture after a matching question that was related to that target word, compared to a neutral question. We examined word production in older and younger adults in different sentence contexts that varied in their semantic relationships with the target word

to be named. The preceding sentence was either neutral (not semantically related to the target), predicted the target word ("matched", e.g., '**mountain**' was predicted by "*what did the alpinist climb*?"), or predicted a different target ("mismatched", e.g., '**ladder**' rather than the target '**mountain**' was predicted by "*what did the construction worker climb*?"; see Table 2.1 for example stimuli).

In terms of our hypotheses, we first expected a "Match effect", with faster picture naming after a matched than neutral sentence. Based on previous literature, it was unclear if older adults' Match effect would be similar to that of younger adults. Preserved or increased semantic knowledge in older adults (cf. Hoffman, 2018) might help both age groups equally to retrieve words in matched sentence contexts, or might help older adults even more than younger adults to compensate for slower word retrieval/production (larger Match effect). However, less efficient use of semantic information (for example due to slower transmission between representations or less prediction forming, e.g., Dell, 1986; Federmeier et al., 2010) may result in a smaller or no Match effect for older adults.

Furthermore, we expected a "Mismatch effect" with slower naming in semantically mismatched contexts than in neutral contexts. In line with the inhibition deficit hypothesis (Hasher & Zacks, 1988) and decreased semantic control in older age (Hoffman, 2018), semantically mismatched information may be more likely to interrupt older adults' retrieval of upcoming words (larger Mismatch effect). A similar Mismatch effect in younger and older adults would suggest that mechanisms used to inhibit competing words are not negatively affected by ageing. Finally, if older adults do not use semantic information to predict upcoming words to the same extent as younger adults, they might experience less interference (smaller Mismatch effect).

Finally, we aimed to examine whether context in general (producing picture names in response to a neutral question) can influence word retrieval/production, relative to no context (producing individual picture names). This was done in an attempt to overcome the issues faced previously when trying to compare word production across different tasks employing differing measures (Kavé & Goral, 2017). Age effects might be exacerbated in a relatively artificial task asking participants to produce individual words without the typical syntactic and lexical connections between words in context. Faster retrieval *within* a sentence context (compared to an isolated word) would suggest that those syntactic and lexical connections between words (as is common in daily-life speech) can aid word retrieval, even if the context

does not provide clear semantic predictions. If older adults rely more heavily on contextual support to aid word retrieval, they may exhibit a larger context effect. On the other hand, listening and responding to another speaker may impose greater working memory demands than producing words without context. If older adults have more difficulty managing the working memory demands of producing words within conversation, they may be delayed by context (cf. Kemtes & Kemper, 1997; Murphy et al., 2000).

Table 2.1 Example stimuli

Target picture	Matched	Mismatched	Neutral	No context
Egg	*What did the chef crack?*	*What did the baby shake?*	*What did the father ask his son to bring?*	No question
Rattle	*What did the baby shake?*	*What did the chef crack?*	*What did the father ask his son to bring?*	No question

The first three columns relate to the three types of sentence contexts used in the study, including example sentences. The question always preceded the presentation of a target picture, which participants had to name. In the 'no context' condition, participants just named the picture (without hearing a question beforehand).

2.2 Methods

This study was pre-registered on the Open Science Framework: https://osf.io/8qexr/. The data corresponding to Chapters 2 and 3 are also available on that OSF page.

2.2.1 Participants

Ethical approval was obtained from the Department of Psychology at the University of York and participants provided informed consent at the start of the study. The final sample included 96 native English-speaking monolinguals. Forty-eight older adults (aged 65-77 years old) were recruited through prolific.co (*n*=34) and through our departmental database (*n*=14). Forty-eight younger adults (aged 18-35) were recruited through SONA (the university's internal participant recruitment system; *n*=6) and Prolific (*n*=42). Participants received either

monetary compensation, Amazon vouchers, or course credit for their participation. The groups of younger and older adults were matched on sex ratio, number of years of formal education received, and the number of participants within each age group who had completed at least an undergraduate degree (see Table 2.2).

Participants were first asked to complete a series of checks, including testing their microphone and playing audio files. We checked these responses before inviting them to the full study. Three participants were not invited as they could not complete the pre-study checks; one participant did not respond to the invitation; and three participants completed the study but were not included as they did not follow the instructions or because their naming-task recordings were empty. Our final sample size of 96 participants (as pre-registered, and after exclusion) was based on a GPower analysis. It was not possible to retrieve previous effect sizes from the literature as comparable designs/tasks had not been used previously with younger and older adults. We therefore conducted a power analysis using a medium effect size (f = 0.25) for an interaction between age and sentence context. This suggested our sample size yielded over 95% power to detect a medium-sized effect.

All included participants furthermore confirmed meeting the following eligibility requirements: they did not use a hearing aid, had (corrected-to-)normal vision, were not colour blind, had not been using medication affecting their concentration in the past three months, did not have a language/reading disability, and had not been diagnosed with a neurodegenerative disease or cognitive impairment. Given that the study was conducted online, we were not able to use an assessment of cognitive functioning such as the ACE-III. In addition to asking participants to confirm each eligibility point, where possible, we also used existing screening criteria to only invite participants without a history of head injury, cognitive impairment, or dementia.

Table 2.2 Participant details

	N	Age (years)	Formal education (years)	Graduate education	Sex		Handedness	
					M	**F**	**Left**	**Right**
Younger	48	24.25 (4.6) 18-35	16.5 (2.8)	29	25	23	10	38
Older	48	69 (4.1) 65-77	15.6 (3.4)	31	25	23	4	44

Age and formal education (mean number of years, and with standard deviations in parentheses, and age range below); graduate education (total number of participants who had completed at least an undergraduate degree); sex and handedness (total number of participants belonging to each category).

2.2.2 Design

Participants completed a picture naming task with the within-participant independent variable Context. This had four levels (see Table 2.1 for examples): Matched, Mismatched, Neutral, or No Context. Age group was a between-subject variable. The dependent variable was picture naming times (ms), defined as onset of naming relative to picture presentation.

2.2.3 Materials

All target pictures for the naming task were presented in greyscale and were sourced from the Multipic database (Dunabeitia et al., 2018) or from Google images. Pictures were preceded by a spoken question or presented without context. The questions were recorded by a female English speaker, reflecting natural speech as much as possible. They were pre-processed using Praat (Boersma & Weenink, 2022) to add 50ms to the beginning and end of each recording and to scale all recordings to 60dB. Background noise was also reduced using Audacity® version 3.0.0.

We created 76 matched question-answer pairs, in which the question was strongly predictive of the upcoming picture (see Table 2.1). Each matched pair was combined with another matched pair to create a duo (e.g., in Table 2.1, '**egg**' and '**rattle**' form a duo). Duos were formed on the basis that the sentence formed a match with one target word but a mismatch with the other target in the duo. We created mismatch sentences in which the

target word was unlikely to follow but not impossible, to avoid unrealistic scenarios that would never happen in real-life conversations. Each duo was also assigned a neutral question, which did not strongly prime a specific word.

Each participant named each picture four times: three times within context (once per matched, mismatched, and neutral question) and once without context. We ensured that participants only heard each question once so that they could not use previous exposure to predict a word. Using the example presented in Table 2.1, half of the participants named '**egg**' four times in the four conditions while the other half named "'**rattle**' four times in the same contexts. Matched and mismatched contexts were therefore the same questions across participants. Neutral questions were matched to the matched/mismatched questions in terms of overall sentence length (number of words) and syllable length and frequency of the key words. The full list of stimuli, with further details about the sentence characteristics and matching, is provided in the Appendices.

To make sure the stimuli functioned as intended (i.e., target words were most likely in matched contexts and least likely in mismatched contexts), we ran three pilot studies, as described below. The pilot studies were completed online with our initial set of stimuli. Participants were recruited through SONA, Qualtrics, and Prolific for all three pilots.

The pilot studies included 47 sets of initially prepared stimuli (comprising *n* = 47 each of matched, mismatched and neutral questions). Changes to the stimuli were made based on the pilot responses. The final set of stimuli was also evaluated through a likeliness rating task in the main study. This confirmed that target words were most likely in matched contexts and least likely in mismatched contexts (see "Results" for an analysis of these ratings).

One pilot study was a short written-picture naming task and was completed by five older (*M* Age = 63.6 years, range = 61-68) and six younger adults (*M* Age = 18.7 years, range = 18-20) to make sure all pictures could be recognised and named easily. We replaced pictures where this was not the case. Two further pilot studies were conducted to examine suitability of the stimulus materials in the different contexts. In the first, 21 older (*M* Age = 65.7 years, range = 60-75 years) and 20 younger adults (*M* Age = 19.85 years, range = 18-31) completed a cloze probability task and a likeliness rating task. In the cloze probability task, they viewed each question individually and were asked to generate their first three single-word answers in response. We computed cloze probabilities (i.e., the proportion of times the target word was given as part of that "top three") through by-item means rather than by-participant means,

and below report scores only including items that were kept in the same form in the actual experiment. Cloze probability was highest for the matched condition in both age groups (Younger: *M* = 80.52%, *SD* = 24.22; Older *M* = 84.56%, *SD* = 19.77). This confirmed the matched target responses were indeed good answers to the questions. In contrast, as we wanted, the target words were almost never given in the mismatched context (Younger *M* = 2.38%, *SD* = 6.80; Older *M* = 2.78%, *SD* = 9.15) and neutral context (Younger *M* = 1.72%, *SD* = 5.51; Older *M* = 1.48%, *SD* = 5.65).

Table 2.3 Pilot 1 Likeliness ratings

	Younger adults	**Older adults**
Likeliness rating (1-5)		
Matched	4.89 (0.13)	4.88 (0.26)
Neutral	2.70 (0.87)	3.12 (0.83)
Mismatched	2.27 (0.79)	2.55 (0.91)

Mean likeliness ratings for matched, neutral and mismatched question-answer pairs obtained from younger and older adults. Note that contrary to the main experiment, the pilot used a 1-5 rating scale (1 = very unlikely, 5 = very likely).

Participants in the pilot also completed a likeliness ratings task. They viewed each question-answer pair and were asked to rate on a scale of 1-5 how *likely* the presented answer was to follow the preceding question. Ratings (only including items that were kept in similar form in the actual experiment) were highest for matched pairs, followed by neutral pairs, and were lowest for mismatched pairs (see Table 2.3). A 2x3 ANOVA confirmed that there was a significant effect of Condition ($F(1.663,84.831) = 199.591$, $p < .001$), with a significant difference between Matched and Mismatched ($p < .001$), Matched and Neutral ($p < .001$), and between Mismatched and Neutral ($p = .015$) likeliness ratings in the expected direction. Older adults overall provided slightly higher ratings ($F(1,51) = 38.832$, $p < .001$), which interacted with Condition ($F(1.718,87.615) = 18.964$, $p < .001$). This reflected that while younger and older adults rated likeliness of matched pairs similarly, older adults' ratings of neutral and mismatched pairs were slightly higher than the younger adults' ratings. However, analyses by

age group confirmed that likeliness ratings were highest for the matched and lowest for the mismatched question-answer pairs in each age group individually.

Finally, we asked participants to indicate whether the scenario depicted in each question-answer pair was possible (i.e., whether it was something that could happen in real life, as we wanted the mismatched sentences to be unlikely but not impossible) and whether the question-answer pairs were grammatically correct. For both questions, the answer options were "yes" (can happen in real life/grammatical) or "no" (cannot happen in real life/not grammatical). Both plausibility and grammaticality judgements (inferred from the mean number of participants who agreed that the scenarios could happen in real life and that they were grammatically correct) were high for all question-answer pairs we included.

After making modifications to the stimuli based on the results from the first pilot studies, we then conducted another pilot study with another five younger (*M* Age = 22.8 years, range = 19-29) and older adults (*M* Age = 63 years, range = 60-67). Scores for the modified stimuli again, in line with pilot described above, confirmed that the targets were most likely to follow matched questions and least likely to follow mismatched questions.

2.2.4 Procedure

The experiment was conducted using Gorilla.sc (Anwyl-Irvine et al., 2020). Participants first read the information sheet and provided informed consent. They then completed a background questionnaire (see "Participants"). Next, they completed a sound check to ensure that they could record audio files through their browser and to adjust their device's volume so that they were able to clearly hear the sentences. For the naming task, participants were allocated to one of twelve experiment lists. Half of the participants named the Set 1 targets and the other half the Set 2 targets (see Appendices). Furthermore, half of the participants named the pictures with context first, while the other half named the pictures without context first. Participants named 38 target words four times each, once without context and once in each of the three question contexts (114 trials, with a break in the middle). The presentation order of the stimuli was pseudo-randomised in the context blocks so that the same word was not repeated twice in a row and there were no more than three consecutive trials of the same type of context.

Participants first completed a picture familiarisation task in which they saw the target pictures and words, asking them to read the word aloud and use it during the task. This phase

was included to make sure all participants recognised the pictures when naming them in the study. Given that pictures were repeated within the task across the four conditions, including a familiarisation phase ensured participants did not see the picture for the first time within the main task, which could have affected condition comparisons. In the naming tasks, participants were instructed to name the pictures as quickly and accurately as possible. Participants first saw three practice trials. In the Context blocks, participants first viewed a fixation cross (500ms) followed by a blank screen while they heard a pre-recorded question. This was followed by another fixation cross (presented for 500ms). Next, the article "the" was presented on the screen (500ms), and then another fixation cross (300ms) was presented before the picture was presented. The picture remained on screen for 2500ms, regardless of when a response was given. In the No-Context block, participants viewed a single fixation cross of the mean duration of the sentence recordings in the Context blocks, plus the duration of the fixation crosses (total of 3853.8ms). The rest of the trial was identical to the Context trials (article followed by picture).

We also assessed participants' subjective experienced workload using the NASA-TLX (Task Load Index, Hart & Staveland, 1988). This task was used to examine (potential) differences in younger and older adults' experienced subjective demands (rated on a scale of 1 (very low) to 100 (very high)) during the naming task. This allowed us to assess potential age-group differences not just in terms of objective performance (i.e., RTs) but also in terms of experienced workload, which might be higher for older than younger adults. After each naming block, participants provided ratings evaluating how mentally demanding, physically demanding, and temporally demanding (pace of the block) they found the task, as well as their performance (how successful they felt in terms of following the task instructions), effort (how hard they had to work), and their frustration level. We also assessed 'overall workload' by asking participants to complete the full NASA-TLX after finishing the full naming task. This again asked participants to complete the same ratings (listed above), but we now also asked participants which aspect (e.g., "effort" versus "mental demand", asking this question for each combination of the six experiences) they found more important when describing the experienced workload. This allowed us to compute scores reflecting the participants' experienced workload per part of the task, as well as an overall score that also took into consideration that different aspects of workload experiences vary in how important they are for individual participants.

Finally, participants completed a likeliness-rating task in which they rated on a scale of 1-7 how likely the target word was to follow each question, for all of the question-answer pairs they viewed during the naming study. The experiment lasted approximately 30 to 45 minutes in total.

2.2.5 Data analysis

Likeliness ratings

Likeliness ratings in the main study were examined using a 2x3 ANOVA with Age (younger, older) as a between-subject variable and Context (matched, mismatched, neutral) as a within-subject variable. Data from two older participants were excluded from the likeliness-ratings analyses. Due to a technical fault one participant was unable to use the ratings scale to indicate a likeliness rating of greater than '3'. The other excluded participant provided a rating of '2' on all trials. Given that Mauchly's test of sphericity (for this task and all others) indicated the assumption of sphericity was violated for the Context variable, Greenhouse-Geisser corrected values are reported.

Picture naming

An accurate response in the picture naming task was either the intended target word or a closely related word (e.g., "painter" instead of "artist"). Other or no responses were scored as an inaccurate response. Picture naming accuracy was >75% for all participants (*M* older adults 96.04%, *SD* = 4.92; *M* younger adults 97.00%, *SD* = 3.37). As pre-registered, because accuracy was close to ceiling, it was not analysed further. Naming RTs were determined using Checkvocal (Protopapas, 2007). RTs <300ms or more than 2.5 *SD* above or below the mean per participant and per condition were removed, using the trimr package (Grange, 2015; removing 2.78% of correct responses). With the exception of the data for the older adults in the Mismatch condition, Shapiro-Wilk tests conducted using raw RTs confirmed the data were normally distributed (*p*s >.25).

RTs were analysed in SPSS using a 2x4 ANOVA to determine whether there was a main effect of Age, Context (i.e., matched, mismatched, neutral, without context), or an interaction between the two. If an effect of context was found, a pairwise comparison (Bonferroni corrected) was used to establish where the effect resided within the four levels. Given that we were specifically interested in the effect of each specific type of context and because we

wanted to examine them while also accounting for age-related slowing, we then also computed the Match, Mismatch, and Context effects based on z-scored RTs for each participant (z-scored separately per age group). The Match effect was the RT difference between the matched and neutral questions; the Mismatch effect was the difference between mismatched and neutral questions; and the Context effect was the difference between neutral questions and naming without context. In addition to a one-way ANOVA per effect (using a Bonferroni adjusted significance threshold of $p = 0.016$ to account for the three comparisons), we computed a Bayesian ANOVA using JASP version 0.17.3 (JASP Team, 2022), which examined evidence for/against an age-group difference on these contextual effects. For each contrast effect (Matched, Mismatched, Context), we compared a model including an age effect (between-groups difference) to a null model (no age-group difference). We report these results in the form of "BF_{01}", showing the evidence for the null hypotheses (no age group difference) over the alternative hypotheses (significant age group differences). Values below 1 indicate evidence for an age-group difference; values above 1 indicate evidence for no age-group difference.

Finally, we estimated the internal consistency of the Match, Mismatch, and Context effects using a permutation-based split-half approach (Parsons, 2020a) with 5000 random splits to check for within-subject variations in these effects.

Further exploratory analyses used linear mixed-effect analyses to examine the context effects while considering both participants and stimuli within one analysis. We also examined the potential role of a specific target word's likeliness scores as provided by the participants after completing the naming task. Finally, we examined the potential influence of word frequency, considering that older adults have shown increased difficulty retrieving low-frequency words. These analyses were conducted using R (4.4.1; lme4 package version 1.1.35) and started with the maximal random-effect structure including all within-participant and within-item slopes (following Barr et al., 2013). Where analyses did not converge, we removed slopes explaining the lowest amount of variance until convergence was reached. Two-level categorical variables were contrast-coded (Age group: younger adults = -0.5; older adults = 0.5). Simple coding was used for the four-level categorical variable Context. "Neutral" was used as the reference level to compare the other three levels to that baseline (contrary to dummy coding, the intercept corresponds to the mean of all cell means). The continuous variables' item frequency and item likeliness rating were z-scored.

Experienced workload (NASA-TLX)

Overall workload effects from the NASA-TLX were calculated by counting how often participants chose each experience as most important between two comparison options (e.g., how often they said they found "frustration" the most important compared to another experience in post-test evaluation). The raw score for each experience was computed and multiplied by the number of times it was chosen as most important. All weighted experiences were summed up and divided by the total number of comparisons participants had to choose from to compute the overall NASA score. We also calculated participants' mean workload score after each block (by calculating the average of their ratings on each Likert scale, without weighing categories). An ANOVA was conducted to determine whether there was a difference in *overall* workload experience between older and younger adults. A mixed ANOVA established whether there was a significant effect of Context (ratings provided after No Context, the first half of Context, the second half of Context, and for the overall task) and age group on experienced workload.

2.3 Results

2.3.1 Likeliness ratings

Table 2.4 Main study likeliness ratings

	Younger adults	**Older adults**
Likeliness rating		
Matched	6.69 (0.30)	6.72 (0.28)
Neutral	3.63 (0.64)	3.30 (0.73)
Mismatched	3.02 (0.56)	2.89 (0.53)

Note that contrary to pilot study 1, the main study used a 1-7 rating scale (1 = very unlikely, 7 = very likely).

Target words in the matched Context were rated as most likely and targets in the mismatched Context as least likely (see Table 2.4, $F(1.729, 159.107) = 2327.697$, $p < .001$, $\eta_p^2 = .962$). Pairwise comparisons showed significant differences between all Context combinations ($p <$

.001). There was no main effect of Age ($F_1(1,92) = 2.905$, $p = .092$, $\eta_p^2 = .031$), suggesting that overall ratings were similar for older and younger adults. However, there was a significant interaction between Age and Context ($F_1(1.729, 159.107) = 4.155$, $p = .022$, $\eta_p^2 = .043$). While ratings of the matched and mismatched sentences were similar for both age groups (Matched: $p = .718$; Mismatched: $p = .225$), neutral targets were rated slightly more likely by younger than older adults ($p = .025$, see Table 2.4). Crucially, however, each age group showed a significant difference in likeliness ratings between all three question contexts (Younger adults: $F(1.656, 77.820) = 1029.568$, $p < .001$, $\eta_p^2 = .956$; Older adults: $F(1.743, 78.441) = 1329.523$, $p < .001$, $\eta_p^2 = .967$, with all pairwise comparisons $p < .001$ in both age groups). Thus, for both age groups, as intended, matched targets were most likely, followed by neutral and mismatched targets.

2.3.2 Picture naming

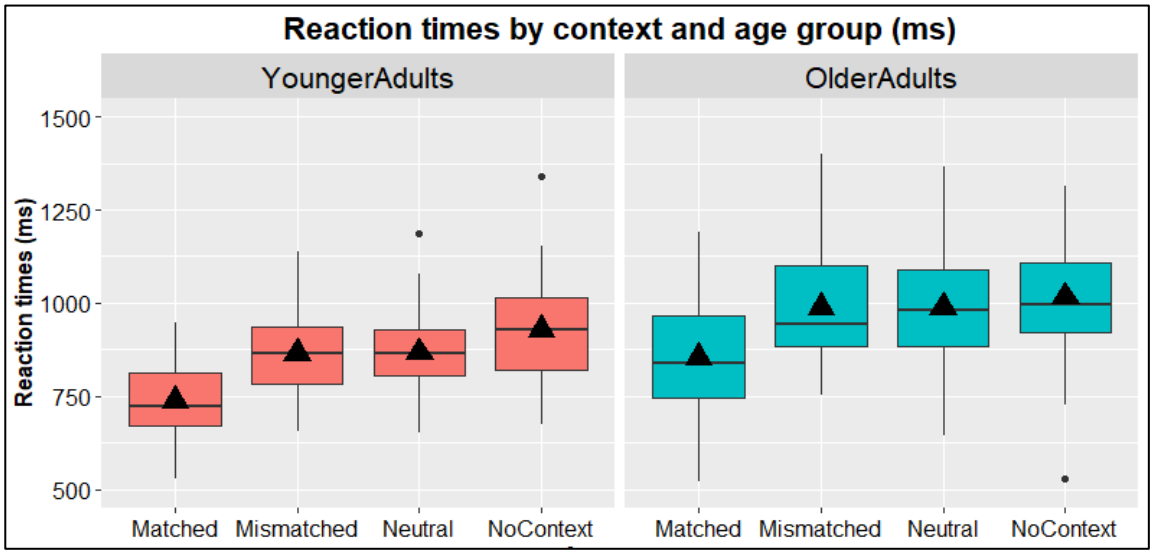

Figure 2.1 Box plots displaying untransformed mean RTs (ms) by age group (left panel: younger adults, right panel: older adults) and naming context. Box plot height denotes interquartile range; vertical lines below plots denote 25th percentile, while lines above denote 75th percentile. Black horizontal lines denote the median, triangles denote the mean. Black dots represent outliers.

We started the picture naming analysis with the untransformed RTs. As expected, there was a significant main effect of Age group on RTs, with older adults taking significantly longer to

name pictures compared to younger adults (Younger *M* = 848.70ms, *SD* = 100.76; Older *M* = 960.19, *SD* = 135.03; $F(1, 94) = 21.358$, $p < .001$, $\eta_p^2 = .185$). There was also a main effect of Context on picture naming times ($F_1(1.525, 143.337) = 110.600$, $p < .001$, $\eta_p^2 = .541$). RTs were fastest in the matched trials, followed by the mismatched and neutral trials, and were slowest in the no-context trials (see Table 2.5, Figure 2.1). Pairwise comparisons showed that there were significant differences between RTs in all naming contexts (*p*s <.001 matched versus mismatched, neutral, and no-context; $p = 0.005$ for mismatched versus no-context; $p = 0.008$ for neutral versus no-context), apart from trials in the neutral and mismatched contexts ($p > .999$). This showed that matched contexts facilitated RTs compared to neutral contexts (Match effect) and that context facilitated RTs in comparison to naming without context. However, no Mismatch effect was observed, suggesting the mismatched context did not negatively affect production. Importantly, there was no significant interaction between Age group and Context ($F_1(1.525, 143.337) = 1.266$, $p = .279$, $\eta_p^2 = .013$), suggesting that the context effects did not differ between the younger and older adults.

Table 2.5 Picture naming times

Context	Younger adults	Older adults
Without context	927.95 (141.75)	1016.28 (164.47)
Matched	738.31 (102.86)	852.66 (145.03)
Neutral	867.25 (106.57)	988.50 (139.05)
Mismatched	864.16 (105.66)	988.36 (150.16)

Means (and standard deviations) per age group and context

Contrast Analyses for Match, Mismatch, and Context effects

As the previous analysis showed that there was a significant effect of age group on RTs, we z-scored the data to account for age-related slowing. Then, for each participant, we computed their Match effect (match versus neutral RTs), Mismatch effect (mismatch versus neutral), and Context effect (neutral context versus no-context; see Figure 2.2).

Starting with the Match effect, both the one-way ANOVA ($F_1(1,94) = 0.233$, $p = .630$, $\eta_p^2 = 0.002$) and the Bayesian analysis ($BF_{01} = 4.20$, error = 0.02%) suggested the Match effect did not differ between age groups. The (Spearman-Brown corrected) split-half internal

consistency of the Match effect was r_{SB} = 0.60, 95%CI [0.46, 0.72], indicating moderate internal consistency of this effect within participants.

Similarly, both analyses suggested the Mismatch effect did not differ between age groups either (one-way ANOVA: $F_1(1,94)$= 0.116, p =.734, η_p^2 = 0.001; BF_{01} = 4.43, error = 0.02%). The (Spearman-Brown corrected) split-half internal consistency of the mismatch effect was r_{SB} = 0.19, 95%CI [-0.10,0.43], suggesting low internal consistency of this effect within participants.

Finally, the one-way ANOVA again showed no age-group difference in the Context effect ($F_1(1,94)$= 1.954, p =.165, η_p^2 = 0.020), although the Bayesian analysis only provided weak support for this null hypothesis (BF_{01} = 1.97, error = 0.02%). The (Spearman-Brown corrected) split-half internal consistency of the context effect was r_{SB} = 0.90, 95%CI [0.86,0.93], suggesting high internal consistency of this effect within participants.

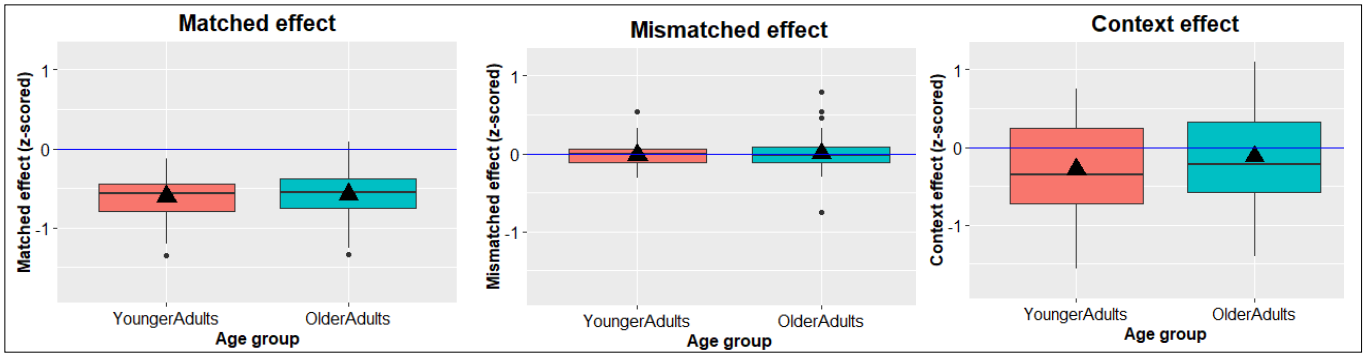

Figure 2.2 Box plots showing the facilitatory effects of matched contexts (left), no Mismatch effect in either group (middle), and facilitatory effect of neutral (context) relative to naming without context (right) in both younger and older adults. Box plot height denotes interquartile range, vertical lines below plots denote 25th percentile, while lines above denote 75th percentile. Black horizontal lines denote the median, triangles denote the mean. Black dots represent potential outliers.

While there is good internal reliability *within participants,* Figure 2.2 also shows high variability in the context effect across participants. Some participants exhibited context-related facilitation whilst others exhibited a context-related cost. We therefore conducted further exploratory analyses to examine if order (naming in context or without context first) could explain this variability. To that end, we computed a 2x2 ANOVA, with the Context effect as the dependent variable and Age and Naming order (completing context or no-context part

first) as independent variables. There was a significant main effect of naming order (F_1(1, 92) = 176.538, p < .001, η_p^2 = .657). Both younger and older adults showed a facilitatory effect of context only when the No-Context part was completed first ("context second group": *M* younger = -0.745, *SD* = 0.382; *M* older = -0.603, *SD* = 0.295) but not when the Context part was completed first ("context first group": *M* younger = 0.172, *SD* = 0.348; *M* older = 0.365, *SD* = 0.360). The effect of age was now significant (F_1(1, 92) = 5.584, p = .020, η_p^2 =.057), suggesting the context effect was slightly smaller for older adults, but this did not interact with naming order (F_1(1, 92) = 0.131, p = .718, η_p^2 = .001).

Exploratory analyses (pre-registered mixed-effect analyses exploring stimulus characteristics)

The analyses conducted so far used by-participant means. Given that language production is also influenced by the items used, we also ran a linear mixed-effect analysis. We report the analysis using untransformed RTs but also ran them using z-scored RTs, which showed the same findings (apart from the main effect of age group, as intended). The first mixed-effect analysis examined effects of Context and Age group, similar to the ANOVAs reported above, and converged with participant and item intercepts but no slopes. The Context findings were the same, with a Match effect (β = -134.128, *SE* = 4.103, *t* = -32.691, *p* < .001), no Mismatch effect (β = -2.836, *SE* = 4.127, *t* = -0.687, *p* = .492), and a Context vs No-Context effect (β = 42.732, *SE* = 4.118, *t* = 10.376, *p* < .001). The Match and Mismatch effects did not differ between age groups (Match x Age: β = -6.510, *SE* = 8.205, *t* = -0.793, *p* = .428; Mismatch x Age: β = 1.321, *SE* = 8.253, *t* = 0.160, *p* = .873). The Context vs No-Context effect was significantly smaller for older adults (β = -30.864, *SE* = 8.236, *t* = -3.747, *p* < .001).

Contrary to our predictions, there was no significant Mismatch effect. Likeliness ratings in the Mismatch condition were significantly lower than likeliness ratings in the other conditions, but within each condition our stimuli varied in their likeliness. We therefore also conducted an exploratory analysis to examine whether there was a direct relationship between a participant's likeliness rating for a given item and their RTs in response to that item. Across all conditions, higher likeliness ratings were associated with faster responses (β = -54.485, *SE* = 1.692, *t* = -32.198, *p* < .001). This was also the case when just considering the Mismatch trials (β = -12.514, *SE* = 3.755, *t* = -3.333, *p* < .001). None of the analyses, however, showed an interaction between likeliness and age group (all *ps* > .15).

A final exploratory analysis examined a potential role of word frequency, considering older adults have been found to have greater difficulty retrieving low-frequency words. No effects of frequency were observed on overall RTs (β = -5.249, *SE* = 6.743, *t* = -0.778, *p* = .439) and, importantly, frequency did not interact with age group (β = 0.294, *SE* = 2.940, *t* = 0.100, *p* = .920). Frequency did not interact with Context either (all *p*s > .25).

2.3.3 Experienced workload (NASA-TLX)

Subjective workload experiences showed no main effect of Age ($F_1(1, 94) = 0.052$, $p = .820$, $\eta_p^2 = .001$), suggesting that younger and older adults experienced workload similarly (see Table 2.6.). Furthermore, there was no significant effect of Context ($F_1(2.521, 236.968) = 2.378$, $p = .081$, $\eta_p^2 = .025$), suggesting that workload did not differ as a result of naming with or without context or throughout the context blocks. This suggests that workload was consistent throughout the naming task. There was also no significant interaction between Context and Age ($F_1(2.521, 236.968) = 0.224$, $p = .848$, $\eta_p^2 = .002$). A one-way ANOVA also showed that overall workload (weighted for the importance of each experience per participant) did not differ between younger and older adults either ($F_1(1, 94) = 0.055$, $p = .815$, $\eta_p^2 = .001$).

Table 2.6 NASA workload ratings

	Younger adults	**Older adults**
NASA		
Single	24.59 (14.65)	24.91 (14.61)
Context 1	27.29 (13.63)	26.70 (15.60)
Context 2	26.30 (15.77)	25.23 (15.69)
After Naming	25.02 (16.42)	23.70 (18.77)
Overall NASA	29.44 (17.92)	28.48 (21.91)

the first four rows show the means and standard deviations of older and younger adults' NASA workload ratings after each naming block. 'Overall NASA' reflects the weighted after-naming score for each age group.

2.4 Discussion

The work presented in this chapter examined the influence of different sentence contexts on word production in healthy younger and older adults. We compared matched contexts that predicted upcoming words, neutral contexts that did not predict a specific word, and mismatched contexts that predicted alternative words, as well as naming words without context. The likeliness ratings confirmed that targets were indeed most likely to follow questions in the matched contexts and least likely to follow questions in the mismatched contexts. In line with the Transmission Deficit Hypothesis, overall production was slower in older adults. This is in line with data that show that RTs are sensitive to age-related changes in language production from the age of fifty (Verhaegen & Poncelet, 2013). This could be related to older adults showing slower word production, although without specifically examining lexical-phonological retrieval difficulties in the current study, it might also reflect general processing speed. Participants in both age groups showed a (similar) significant Match effect, reflecting faster naming in the matched than neutral contexts. No Mismatch effect (difference between mismatch and neutral contexts) was found in either age group. Some effect of context was observed in both age groups, although this was only present when participants had to name pictures without context first.

2.4.1 Facilitatory effects of semantically matching contexts

Similar to Shao and Rommers' study with younger adults (2020), our data showed that target-word production was faster after matched than after neutral sentences. The Match effect is likely related to semantic priming effects, wherein words in the context prime related words, making their retrieval faster. Listeners might furthermore be predicting upcoming words and when these predictions are in line with the required response, production might be facilitated. This facilitation was found for both younger and healthy older adults to a similar degree. Previous research assessing context effects has mostly focused on comprehension and has returned mixed findings when comparing older and younger adults. Some of those studies have suggested similar benefits for older and younger adults (e.g., Kliegl et al., 2004) while others have suggested older adults benefit more (e.g., Pichora-Fuller et al., 1995) or less (e.g., Wlotko et al., 2012) from semantically related context than younger adults. However, many of these studies compared matched to incongruent or mismatched conditions, rather than to a

neutral baseline. This often makes it difficult to attribute any context effects and age group similarities or differences to benefits associated with *matched* semantic content specifically.

The observed matching context effects for younger and older adults can be explained through more automatic mechanisms such as semantically related words being primed by previously presented words, therefore facilitating word retrieval during production. In this specific task context, where participants had to alternate between comprehending speech and producing a response, this priming operated cross-modally. Linked to priming, speakers can use the speech presented to them to predict suitable upcoming responses to their conversation partner. The findings are also in line with the Transmission Deficit Hypothesis, which argues that older adults' slower word retrieval is related to slower connections between the lexical and phonological level, with the semantic level staying intact. The combination of overall slower naming but a preserved Match effect supports this hypothesis and suggest the preservation or proliferation of representations within the semantic system can facilitate connections between the semantic and lexical level.

Our study shows that older adults can indeed benefit from semantic information matching upcoming target-word production. In this specific case, they did not benefit more than younger adults (unlike e.g., Pichora-Fuller et al., 1995). However, such additional benefits for older adults might only arise when processing demands are high (e.g., when the task is presented in background noise, as done in Pichora-Fuller et al., 1995). Furthermore, more compensatory effects (exhibiting a larger Match effect in older adults) might be more likely to arise in low-frequency target words with poorer lexical-phonological connections (cf. James & MacKay, 2001). In contrast, our study used relatively high-frequency words, a familiarisation phase, and each word was produced multiple times to ensure the same words were used in all conditions. As a result, older adults may not have had to rely more heavily on semantic context than younger adults. However, the benefits observed are likely to require older adults to have intact semantic knowledge to facilitate priming of related words. Chapter 3 therefore examined semantic knowledge in both younger and older adults, to test the hypothesis that older adults indeed continue to benefit from a semantic knowledge and vocabulary size at least comparable to that of younger adults. We furthermore examined whether individual differences in the size of context effects during language production can be explained by individual differences in semantic knowledge.

2.4.2 Mismatched sentence contexts

Picture-word interference paradigms without context have shown that distractor words semantically related to target pictures (e.g., target: ball, distractor word: frisbee) often slow down target word retrieval compared to unrelated distractors (e.g., hammer) in both younger and older adults (Taylor & Burke, 2002). This suggests that competing active semantic information can interfere with the production of intended words. However, in our study, mismatched sentence contexts did not influence target-word production times. Although we expected a Mismatch effect with slower responses when the target word was unexpected, these findings align with a previous comprehension study using comparable sentences (Haigh et al., 2022), where no Mismatch effects were found either. Similarly, Luke and Christianson (2016) showed no negative impact of unpredictable information in an eye-tracking study. Previous research (Brothers & Kuperberg, 2021) has suggested people engage less in predicting upcoming words when the context is uninformative. However, considering the observed Match effect, it is unlikely that the absent Mismatch effect in our study is due to the context's overall low informative level resulting in people stopping predicting upcoming information entirely.

While our mismatched targets were *unlikely* answers to the preceding question, they were not *impossible*. Indeed, while likeliness ratings differed significantly between mismatched and neutral sentences, the differences were small. This is likely related to the mismatch trials not being unexpected enough, although the difference can also be increased by making the neutral trials more neutral/predictable. Previous comprehension studies have found the largest processing costs when sentence endings are semantic anomalies (e.g., Federmeier & Kutas, 1999). The mismatched contexts in the current study may not have increased competition between word candidates sufficiently to incur a cost, in particular when comparing these effects to a neutral baseline (as opposed to a comparison to expected, matched contexts). Recent research also assessing production indeed showed only a small cost when unpredictable words were used that were closely related to the expected target (e.g., "skull" instead of "brain") compared to a larger cost when the unpredictable word was not related to the expected target (Bannon et al., 2025). It is worth considering that within the mismatched trials in this chapter, sometimes the target word shared a distant semantic relationship with the predicted word (e.g., consider expected word: 'crown'- target: 'ring', expected word: 'banana'- target: 'potato', expected word: 'plane- target: 'ship') while others

did not share such as relationship. Some models of language production propose that the retrieval of all semantic relatives of encountered/expected words is facilitated due to priming effects (rather than creating interference), with more closer relatives creating a greater degree of facilitation due to greater featural overlap, while more distant relatives create a smaller degree of facilitation comparatively due to less featural overlap (Mahon et al., 2007). The inclusion of this type of stimuli in the current chapter may thus have reduced any potential interfering mismatching effects.

Our exploratory analyses further suggested that, also within the mismatch context, responses were slower for target words that were less likely to occur in the sentence context. This suggests that a selection of impossible rather than unlikely target endings would perhaps be more likely to show significant interference effects. Variability between items in their (un)likeliness also likely contributes to the low split-half reliability observed for the Mismatch effect. However, age group did not interact with the effect of likeliness in the continuous rating analyses either, suggesting the older adults in this study were truly not affected more by target-word likeliness.

2.4.3 Neutral sentence contexts and ageing

Picture naming was faster following neutral contexts compared to no context, suggesting sentence context facilitated production. Previous research comparing age-group differences during language production with and without context has suggested any age effects might be less likely to be observed during connected speech in context (Kavé & Goral, 2017). On the other hand, we expected that producing target words in neutral contexts (which involves speakers holding a sentence context in short-term working memory store) might have greater consequences for subsequent language production (relative to producing the word in isolation) if the former taxes working memory resources to a degree that creates a delay in subsequent production processes. Our finding that both older and younger adults appeared to benefit from context, compared to producing words in isolation, is in line with previous reviews of connected speech tasks (Kavé & Goral, 2017). However, comparison across such studies is made more difficult by the range of measures used (e.g., hesitations, naming times, circumlocutions). Furthermore, in connected speech, older adults can use compensatory mechanisms to mask word retrieval difficulties. The current chapter therefore compared language production in older and younger adults in the form of picture naming with and

without sentence context. This allowed us to compare the same measure (naming times) and removed any compensatory strategies that could be used during free speech.

In the work reported in this chapter, the facilitation of context appeared to be similar for both age groups, although some analyses suggested the effect was slightly smaller for older adults. Pictures were repeated throughout the task and participants were exposed to them beforehand through a picture familiarisation phase. It is possible that this facilitated older adults' overall retrieval (compared to not having seen the picture or word beforehand) and that older adults benefit more from context when words are harder to retrieve (e.g., without previous exposure or when using lower-frequency words). Older and younger adults also experienced the workload in No-Context and Context conditions comparably, suggesting there was no age-group difference in perceived effort involved in naming with or without context.

However, follow-up analyses suggested the facilitatory effect of context in terms of naming times might not be entirely driven by effects of context as such. While the order of naming with or without context first was counterbalanced across participants, only participants who completed the naming task without context first showed faster naming within context. Given that pictures were repeated, it is possible that participants benefited from naming those pictures without context first and therefore showed faster naming times in the sentence contexts. This could be because previous naming primed the lexical form (allowing for faster retrieval in the context task when completed second) and/or because participants expected the same pictures to appear again. A true effect of context facilitation on production should occur even if the context part is completed first. However, context effects varied between participants and it is possible that other differences (e.g., individual differences in terms of working memory) contributed to the variability observed in context effects. This is assessed in more detail in the next Chapter.

2.5 Conclusion

In conclusion, the findings reported in this chapter suggest that the contexts within which words are produced have significant consequences for word retrieval and production efficiency. Our study provides an important first step in comparing word production with and without context, while using the same measure (naming response times) to bridge the gap between existing literatures.

Chapter 3: Age-related changes in semantic and cognitive abilities and their role in context effects

This chapter is adapted from the following published article:
Hanif, N., Jefferies, E., & de Bruin, A. (2025). Naming speed during language production in younger and older adults: Examining the effects of sentence context. Quarterly Journal of Experimental Psychology, 0(0). https://doi.org/10.1177/17470218241309602

Abstract

Changes in language production abilities in old age have been attributed to a range of underlying changes in semantic and cognitive abilities including semantic knowledge, verbal fluency, semantic control, inhibition and verbal short-term working memory. In this chapter, we therefore sought to understand how these abilities change with age, and if the ways in which participants responded to the different contexts in Chapter 2 related to age or individual differences in these abilities. It was hypothesised that older adults' continued use of matched sentence contexts to facilitate their retrieval of predicted/primed words (termed the 'match effect') might be linked to an age-related preservation/ growth in semantic knowledge. We also expected that semantic and cognitive control might decline with age; and this might link to individual differences in the size or presence of an interference ("mismatch") effect. Finally, it was expected that short-term verbal working memory capacity would decline with age, and that individuals who performed most poorly on this task would benefit more from producing words in isolation (since this is less cognitive demanding) than producing them in context (termed the 'context effect'). In this chapter then, the participants who took part in in Chapter 2 were invited to take part in a second session wherein they completed a battery of tasks measuring these abilities. Older adults showed larger semantic knowledge but poorer inhibition and (on some measures) semantic control than younger adults. However, none of these measures predicted the sentence context effects observed in Chapter 2.

3.1 Introduction

The empirical work in this chapter aimed to further examine the potential mechanisms underlying the sentence context effects computed in Chapter 2 (i.e, the match, mismatch and

context effects). Specifically, we studied how the younger and older adults who took part there differed in terms of their cognitive and language abilities (such as semantic knowledge and control, inhibition, and working memory), and assessed how these abilities contributed to the computed context effects. Note that the work reported in the current chapter was pre-registered before completing data collection for Chapter 2, thus the hypotheses in this chapter do not take into consideration the Chapter 2 findings (e.g., the absence of a Mismatch effect). However, in this chapter we focus mostly on trying to explain *individual differences* in these context effects, which were indeed observed in Chapter 2, despite the absence of group-level differences.

3.1.1 Semantic knowledge

As introduced in Chapter 2, semantic knowledge has been found to increase with age (e.g., Carrol, 2023; Hoffman, 2018; Hoffman et al., 2018; Kavé & Halamish, 2015; Kavé & Yafé, 2014; Verhaeghen, 2003). For example, in a synonym selection task, Hoffman (2018) and Hoffman et al. (2018) asked younger and older adults to select the synonyms of probe words (e.g., *"which means the same as bombastic?"* answer: *pompous*, other answer options: *destructive, anxious, bickering*). In both studies, older adults provided significantly more correct answers than younger adults. Larger semantic knowledge scores might also relate to word retrieval/ production. On the one hand, faster language production has been observed for participants with a larger vocabulary (Shao et al., 2014), suggesting greater semantic knowledge is associated with faster word retrieval. In older adults, their larger vocabulary may act as a compensatory defence against age-related lexical access difficulties (Juncos-Rabadán et al., 2010). On the other hand, having access to more words could create additional interference and disrupt language production (Ramscar et al., 2014). For example, Hoffman and colleagues (2018) found a negative relationship between semantic knowledge and coherence in connected speech (but cf. Shafto et al., 2016).

The size of individuals' semantic knowledge stores might particularly relate to the Match effect measured in Chapter 2 (the degree of facilitation experienced by participants when producing words following predictive/ matched contexts compared to when producing them following neutral contexts). The Match effect is expected to depend on participants having access to semantic knowledge and connections, which allow for semantic priming and predictions about upcoming, semantically related words. In the current chapter, we therefore

used a synonym judgement task to assess semantic knowledge in the younger and older adults tested in Chapter 2, as well as checking for a potential relationship between semantic knowledge and context effects (in particular the Match effect). As discussed above, such a relationship could go in two directions. Greater semantic knowledge and a larger vocabulary could facilitate word retrieval in matched contexts if they increase priming and allow participants to more easily predict upcoming words. Alternatively, having access to more semantic knowledge can increase interference and disrupt production (Ramscar et al., 2014). This could then result in smaller Match and potentially larger Mismatch effects.

3.1.2 Fluency

In this chapter we also assessed semantic knowledge and word retrieval through verbal fluency tasks, which require participants to produce as many words as possible belonging to a specific category (semantic fluency) or beginning with a specific letter (letter fluency), within a specified time. Fluency has been linked to both word retrieval/ production efficiency and semantic knowledge. Counter-intuitively, older adults often perform more poorly than younger adults on semantic fluency tasks while age-group differences tend to be smaller or absent on letter fluency tasks (e.g., Gordon et al., 2018; Kavé & Knafo-Noam, 2015). Here, we were predominantly interested in verbal fluency in general, given its links to both word retrieval/ production efficiency and semantic knowledge across semantic and letter tasks (cf. Gordon et al., 2018). Verbal fluency has been associated with language production across younger and older adults (Higby et al., 2019). Greater semantic knowledge and word retrieval/ production efficiency, as measured through verbal fluency scores, could help participants to benefit more from semantic connections in matched sentence contexts. We therefore examined whether verbal fluency (across letter and semantic fluency tasks) influenced the Match effect.

3.1.3 Semantic and domain general control

As discussed previously, older adults have also shown declines in inhibition and semantic control. (e.g., Hasher & Zacks, 1988; Hoffman, 2018). These control mechanisms are believed to relate to language production, where speakers need to inhibit competitors in favour of producing required words (e.g., Faroqi-Shah et al., 2014). This could be particularly pertinent in scenarios wherein speakers are required to produce unexpected words (i.e., the

mismatched sentences in Chapter 2). Hoffman (2018) used a global and feature semantic association task to assess semantic control. For instance, in the feature association task, participants selected the feature associate of a probe word (e.g., the item related in colour or size). In the low control manipulation (congruent trials), probe and target shared a semantic relationship (e.g., "*which is the same colour as cloud* ?", target: '**snow**') and the distractors were not semantically related to the probe (e.g., '**egg**', '**step**', '**basket**'). In the high control manipulation (incongruent trials), the probe and target did not share a semantic relationship, but one of the distractors was semantically related to the probe (e.g., probe = '**salt**', colour associate = '**dove**', distractor = '**pepper**'). Older adults displayed poorer accuracy and longer RTs on the semantically incongruent trials, compared to younger adults (Hoffman et al., 2018, Hoffman, 2018).

Hoffman and colleagues also found semantic control to be related to coherence during language production, with participants who performed poorly on the semantic control task also producing less coherent speech. Here, we therefore examined potential relationships between semantic control and context effects (in particular, the Mismatch effect). We expected participants who exhibited poorer performance on measures of semantic control to exhibit a larger (negative) Mismatch effect (greater interference from competing semantic information).

3.1.4 Inhibitory control

Domain general inhibitory control has also been found to decline with age. For example, older adults have been found to experience greater interference costs than younger adults from incongruent stimuli in colour Stroop tasks, where participants are asked to produce a response based on the text colour of presented words while ignoring word meaning (e.g., when the word '**red**' is presented in green text; Spieler & Faust, 1996; West & Alain, 2000). Similar age effects have been found on other inhibition tasks too (e.g., Hoffman, 2018) although they might depend on the task used (e.g., de Bruin & Della Sala, 2018; Rey-Mermet et al., 2018).

Similar to semantic specific control processes, domain general inhibitory control processes may also influence language production by facilitating the suppression of competing words. This could be especially relevant for the Mismatch effect, where participants have to produce an unexpected word while controlling interference from an expected word. Indeed, poorer inhibition skills can predict the level of coherence in

individuals' speech (Arbuckle & Gold, 1993; Yin & Peng, 2016). In contrast, however, Higby et al. (2019) did not find a relationship between inhibition and picture or object naming response times. In addition to semantic control, the current chapter therefore also assessed inhibition to examine whether poorer inhibition skills are associated with a larger Mismatch effect.

3.1.5 Short-term memory capacity

In line with resource theories, short-term memory capacity has been found to decline with age (Van der Linden et al., 1994; Waters & Caplan, 2003). For example, the mean number of items recalled in span tasks (including forward and backward digit, and letter and word span tasks) is lower in older adults (see Bopp & Verhaeghen, 2005, for a review). Short-term memory has also been associated with language processing (e.g., Feier & Gerstman, 1980; Kemper et al., 1989; Walsh & Baldwin, 1977). In the current chapter, we therefore explored how age-related short-term memory declines (measured through the digit span task) may contribute to the Context effects (compared to no context). When producing words in context, speakers must hold the context within their working memory whilst planning and producing their verbal response. Producing words without context might not place the same demands on working memory resources. Thus, poorer working memory capacity might modulate how much participants can benefit from producing words in context and therefore relate to the context effect in Chapter 2.

3.1.6 Lifestyle and social network

Finally, we included participants' education, lifestyle, and social network in the analysis in this chapter. Education may serve as a protective factor that may reduce word-retrieval difficulties associated with ageing (Gordon & Kindred, 2011). Higher education is also argued to help preserve inhibitory control mechanisms in old age (Paolieri et al., 2018), which might be linked to how much interference participants encountered when producing words within mismatching contexts in Chapter 2. Particular lifestyle activities such as reading, gardening and walking have been suggested to have positive effects on cognition in older adults, and to slow the rate of cognitive decline (Lee et al., 2025; Scarmeas et al., 2003); while others such as watching tv have been associated with negative effects, such as declines in verbal memory (Fancourt & Steptoe, 2019). However, these links are tentative, and a causal relationship has not been established between these activities and specific aspects of cognition that might

contribute to the context effects studied in Chapter 2. We furthermore assessed the frequency and nature of social interactions, as previous research suggests that reduced social interactions are linked to poorer language ability in older adults, as measured by naming tasks (Burke & Shafto, 2004; Farrell et al., 2014; Keller-Cohen et al., 2006). Social interactions are believed to promote language ability through the 'use it or lose it mechanism' whereby the regular activation and retrieval of words preserves their underlying connections, thus keeping them 'alive' (Keller-Cohen et al., 2006; Salthouse, 1991). More frequent interactions with a variety of different people may therefore be associated with the preservation of semantic networks (leading to overall reductions in word retrieval times/ more effective use of predictive (matched) contexts to retrieve words). Based on the described relationships between educational, lifestyle, and social factors and language, in the current chapter we were interested in exploring whether and how these factors can contribute to the context effects in Chapter 2.

3.1.7 Rationale

In this chapter, we assessed potential differences between younger and older adults on measures of language and cognitive functioning (including semantic knowledge, fluency, semantic control, inhibition, and verbal short-term memory capacity), as well as in terms of lifestyle and social interactions. We also studied how these variables related to language production in the different contexts assessed in Chapter 2. For the match effect, we were particularly interested in semantic knowledge and fluency. For the mismatch context, although neither the younger nor older adult group exhibited a mismatch effect in Chapter 2, we were interested in whether individual differences were linked to semantic control and/or inhibitory control. Finally, we assessed whether the magnitude of the context effect (neutral context versus naming in isolation) was related to individuals' short-term memory capacity.

Taken together, the empirical work reported in the previous and current Chapter allowed us to examine potential age-related changes in word retrieval during word production in sentence context, as well as the underlying mechanisms contributing to word retrieval/ production in context. Specifically, considering the observed age-group differences in previous studies, our studies also aimed to better understand the potential impact of (age-related changes in) semantic knowledge and control on language production.

3.2 Methods

This study was pre-registered on the Open Science Framework: https://osf.io/wf8cm. The data are available on https://osf.io/8qexr/, together with Chapter 2.

3.2.1 Participants

Ethical approval was obtained from the Department of Psychology at the University of York. All participants who took part in Chapter 2 were invited to take part in a second session, with 45 older and 38 younger adults taking part. Compared to the full set of participants in Chapter 2, the participant profile was comparable in terms of age (*M* younger adults = 24.61 years, *SD* = 4.79; *M* older adults = 69.09, *SD* = 4.12) and education (*M* younger adults = 16.57 years, *SD* = 2.99; *M* older adults = 15.60, *SD* = 3.36). The mean interval between completing the tasks in Chapter 2 and the current study was 22.86 days (range = 6-45 days, *SD* = 7.65).

3.2.2 Materials/Tasks

Semantic knowledge

Semantic knowledge was assessed through a synonym judgement task (adapted from Wu & Hoffman, 2022). Participants viewed 67 word pairs and decided if the two words shared a similar meaning by pressing a keyboard button (**S** = related; **D** = different). Half of the word pairs were unrelated (e.g., '**formidable**', '**obdurate**') and the other half were related ('**recondite**', '**abstruse**'). There was no time limit per trial and the next trial started automatically when participants completed the previous trial. For each participant, we calculated an ISDT score (an index based on signal detection theory) to correct for guessing and response style (Huibregtse et al., 2002). This score takes into consideration participants' hit rate (proportion of correct "related" responses), as well as their false alarm rate (proportion of "related" responses to different pairs of words). This score ranges from 0-1 (a score closer to 1 indicates better performance, whereas a score closer to 0 indicates lower performance).

Verbal fluency

We measured both letter and semantic fluency. Participants completed three letter fluency trials (producing as many words as they could beginning with '**F**', '**A**', or '**S**') and three semantic

fluency trials (‘**animals**’, ‘**fruits**’, and ‘**items of clothing’**). Each trial was 60s long. For each participant we computed a composite fluency score (the average of number of words produced across the letter and semantic trials). Fluency data from one older and one younger participant were excluded from the analyses because the recordings were of poor quality or empty.

Semantic Control

Semantic control was measured through two semantic association tasks (used in Hoffman, 2018). In the global association task, participants selected the word associated with a probe, from a set of possible answers. Within the low demand condition (50% of trials), there was a strong semantic relationship between the probe and target (e.g., probe*:* ‘**town**’*,* target: ‘**city**’). In the high demand condition, there was a weak semantic relationship between the probe and target (e.g., probe: ‘**iron**’**,** target: ‘**ring**’).

In the feature association task, participants selected the word matched with a probe word on a particular feature (e.g., colour or size, whilst ignoring distractor words). Half of the trials required participants to select the word most closely related in size, and the other half to select the word most closely related in colour to the probe. Half were congruent trials (target and probe shared a semantic relationship). For example, participants would see “*which is the most similar in size to door*?” with ‘**window**’ being the target and none of the distractors *(*‘**bottle**’, ‘**report**’, ‘**factory**’*)* sharing a semantic relationship with the probe. In the incongruent feature trials, the target and probe were not semantically related but one of the distractors was related to the probe. For example, participants would see “*which is most similar in size to* **ashtray**?”, with target: ‘**diary**’ and distractor ‘**cigarette**’. There was no time limit (following Hoffman, 2018) and the next trial started automatically when they completed the previous trial.

In both tasks, half of the trials included four answer options, and the other half two; in the analyses we collapsed across number of options. In the main analysis, we calculated the accuracy cost within each task for each participant. This was the z-scored accuracy difference between the low and high control conditions in the global association task (global cost) and the low and high control conditions in the feature association task (feature cost). Additional analyses were conducted using RTs, for which we removed RTs two standard deviations above

or below participants' conditional mean (following Hoffman, 2018, and considering this task did not have a time limit) as well as incorrect responses.

Inhibitory control

Inhibitory control was measured through a verbal and non-verbal Stroop task. In the verbal version, participants provided a keyboard response based on the colour of a written word. In congruent trials (n=28), word meaning and colour were the same (e.g., "**red**" presented in red), while they differed in incongruent trials (n=28, e.g., the word "**blue**" presented in red). Neutral trials used a non-colour word (n=28, e.g., "**flower**" presented in blue). The non-verbal version was a digit-based task, where participants decided which of two presented numbers was larger physically in terms of text size whilst ignoring numerical size. In the congruent trials (n=28), the number which was physically bigger was also numerically bigger than the other number. In the incongruent trials (n=28), the numerically smaller number was presented in a larger font. In neutral trials (n=28), the same number was presented twice.

In each trial, participants saw a fixation cross for 500ms, followed by the stimulus. The next trial was presented as soon as a response was given or, if no response was provided, after 3000ms. A verbal and a non-verbal Stroop interference cost was computed for each participant as the mean RT difference between the neutral and incongruent trials in each task. By computing the difference between neutral and incongruent trials, we specifically looked at interference, leaving out the influence of facilitation on congruent trials. Prior to calculating Stroop cost, we first removed incorrect responses and RTs 2.5 *SD*s above or below the mean per condition and per participant. Participants' composite Stroop cost was calculated by computing the mean cost across the two tasks. The colour Stroop data from one older participant was not saved successfully and was therefore excluded from analysis.

Short-term memory span

Short-term memory was assessed using a digit span task. Participants viewed sequences of two to eight digits, with the sequence size increasing after every two consecutive trials (16 trials in total). Within each trial, participants viewed a fixation cross for 250ms, followed by the individual presentation of each digit. Digits were presented in the centre of the screen for 800ms before the screen automatically proceeded to the next digit. At the end of each trial, a text box appeared, and participants were asked to type all the digits they could remember

from that trial. We calculated the proportion of correct sequences recalled by each participant. All participants saw all sequences, even if they made mistakes earlier on in the task (with shorter sequences).

Lifestyle and social network scores

Demographic details were derived from two questionnaires (administered in Chapter 2 as well as this one). Four participants did not provide the total number of years of formal education they had received. A lifestyle questionnaire (adapted from Scarmeas et al., 2003) asked participants to state for a range of activities (e.g., reading or traveling, 18 items) on a scale from 1 (never) to 3 (often) how often they did them during the year preceding the Covid-19 pandemic (the study was conducted during 2021, when many Covid-19 social distancing restrictions were still in place). Participants' total score on the questionnaire was computed as their 'lifestyle score' (min = 18, max score = 54). In addition, we assessed social network size by asking participants to estimate the number of people they had had regular contact with in the past 6 months (including face-to-face, by phone or mail, or on the Internet). This was assessed across categories including close friends, family, neighbours, co-workers, school/child relations, people who provide a service, and others. We added together the total number of people in each category to compute each participant's social network size (adapted from Bruine de Bruin & Parker, 2020). We removed 'social network size' from the analysis for participants who reported having been in regular contact with more than 1000 people in the past 6 months (*n*=3).

3.2.3 Procedure

The experiment was conducted using Gorilla.sc (Anwyl-Irvine et al., 2020). Participants read a study information sheet and completed a consent form to confirm that they met the study criteria and agreed to participate. Participants also completed a sound check to ensure their microphone was working (required for the fluency task). After this, participants completed the tasks in the following order: synonym judgement, semantic control tasks, Stroop (verbal), Stroop (digit), digit span, letter and semantic fluency, and lifestyle/social questionnaire. The study took approximately 45-60 minutes.

3.2.4 Data analysis

First, we examined whether the age groups differed in their performance on these tasks, using independent t-tests. Next, we conducted three regression analyses assessing if the measures described above explained the match, mismatch, and context effects observed in Chapter 2. We z-scored the data for each continuous predictor variable (across age groups, given that part of our analysis aimed to examine individual differences in relation to language/cognitive abilities across age). Pearson's correlation coefficients between the predictors (all below .5) as well as VIF statistics (VIF values <2.5 for all predictors) suggested there were no multi-collinearity issues.

We conducted hierarchical multiple linear regression analyses with the Match, Mismatch, and Context effects as outcomes. In each model, demographic variables were inputted first (age, lifestyle score, gender, social network size, years of formal education); this was followed by the cognitive and language variables (synonym judgement ISDT score, fluency composite score, semantic control global cost, semantic control feature cost, Stroop composite interference cost, and digit span accuracy); and finally we also included interactions between the cognitive/language variables and age. Although all participants completed all tasks, there were some technical issues with some data files and we therefore removed missing data pairwise in the analyses. We computed split-half reliability estimates for the cognitive variables entered into the regression models. These were generally moderate (Spearman-Brown scores ranging between .43 and .74).

3.3 Results

3.3.1 Age group comparisons

Below, we first present the analyses comparing the age groups. In terms of semantic knowledge and verbal fluency, older adults outperformed the younger adults on the synonym judgement task (*M* older adults = .66, *SD* = .13, *M* younger adults = .47, *SD* = .16; $t(81) = 6.169$, $p < .001$, $d = 1.359$; see Supplementary Figure 1) and the verbal fluency task (*M* older adults = 18.37, *SD* = 3.35, *M* younger adults= 16.95, *SD* = 2.86; $t(79) = 2.037$, $p = .045$, $d = .454$, see Supplementary Figure 2).

In terms of semantic control (see Supplementary Figure 3), our pre-registration focused on accuracy scores. For both tasks, we computed a semantic control score by taking the

difference between strong and weak trials (global association) and congruent and incongruent trials (feature association). In the global association task, older adults' semantic control cost (*M* = 4.35%, *SD* = 8.61) was, surprisingly, significantly *lower* than the younger adults' cost (*M* = 11.40%, *SD* = 6.76; *t*(81) = -4.093, *p* <.001, *d* = -0.902). In the feature association task, the numerical pattern went in the same direction but was not significant (older adults *M* = 12.78%, *SD* = 19.67; younger adults *M* = 17.11%, *SD* = 23.35; *t*(81) = -0.917, *p* = .362, *d* = -0.202). The RT data showed no significant cost difference in the global association task (older adults *M* = 836.24ms, *SD* = 478.02; younger adults *M* = 699.74ms, *SD* = 487.26; *t*(81) = 1.285, *p* = .203, *d* = .283). In the feature association task, the RT cost was significantly higher in the older (*M* = 1277.61ms, *SD* = 815.00) than in the younger age group (*M* = 703.84ms, *SD* = 735.63; *t*(80) = 3.325, *p* = .001, *d* = 0.736). Similar RT results were obtained when comparing the z-scored RTs, considering older adults responded more slowly overall. Given the potential speed-accuracy trade off (older adults showing larger RT costs with smaller accuracy costs), we also computed inverse efficiency scores (RT/percentage correct). These did not differ between age groups for the global task (*t*(81) = -.267, *p* = .790, *d* = -0.059) or for the feature task (*t*(80) = -.497, *p* = .620, *d* = -0.110).

The Stroop analysis, as pre-registered, focused on the RTs only. This analysis excluded incorrect responses. Accuracy in the non-verbal Stroop task was 82.07% (*SD* = 4.66) for older adults and 80.03% (*SD* = 9.19) for younger adults. In the verbal Stroop task, accuracy was 74.73% (*SD* = 13.97) for older adults and 81.21% (*SD* = 4.79) for younger adults. The Stroop RT interference cost was significantly larger in the older (*M* Stroop = 82.41ms, *SD* = 80.29) than younger adults (*M* = 39.54ms, *SD* = 85.18; *t*(80) = 2.344, *p* = .022, *d* = 0.519; see Supplementary Figure 4). This significant difference remained when analysing the z-scored RTs (considering overall slower responses in older adults).

Short-term memory capacity (measured through the digit span task) did not differ between older adults (*M* = 65.56%, *SD* = 14.06) and younger adults (*M* = 65.41%, *SD* = 17.30; *t*(81) = 0.041, *p* = .967, *d* = 0.009; see Supplementary Figure 5).

Finally, social network size was slightly larger in the younger participants (*M* = 33.03, *SD* = 23.09) compared to the older participants (*M* = 27.73, *SD* = 17.25) but this difference was not significant (*t*(78) = -1.174, *p* = .244, *d* = -.264). Older adults scored significantly higher on the lifestyle questionnaire (*M* = 37.64, *SD* = 3.93) in comparison to the younger participants (M = 34.66, SD = 3.84; *t*(81) = 3.483, p <.001, d =.767).

3.3.2 Hierarchical Regression

Hierarchical linear regressions were computed next to measure the contribution of each predictor on each context effect from Chapter 2 (Match, Mismatch, and Context effect). Tables 3.1- 3.3 show the relationships between the education, lifestyle and social variables, synonym judgement, verbal fluency, the two semantic control (global and feature association), inhibition, and digit span tasks with the Match, Mismatch, and Context effects.

Match effect

None of the included cognitive variables were significant individual predictors of the Match effect (see Table 3.1.). The contribution of demographic variables together (Model 1) was close to significance, with social network reaching significance. This suggests that people with a larger social network showed a smaller Match effect. Models 2 and 3, which included the cognitive and language variables, and age interactions with those cognitive and language variables, did not contribute significantly beyond the model only including demographic variables.

Table 3.1 Hierarchical regression (match effect)

model 1				**model 2**				**model 3**			
	B	***SE***	***P***		***B***	***SE***	***P***		***B***	***SE***	***P***
Intercept	-0.54	0.04	<.001	*Intercept*	-0.55	0.04	<.001	*Intercept*	-0.52	0.06	<.001
Age	-0.02	0.03	.632	*Age*	< -0.001	0.04	.999	*Age*	0.02	0.05	.666
Education	-0.04	0.03	.164	*Education*	-0.04	0.03	.219	*Education*	-0.03	0.04	.332
Lifestyle	0.06	0.03	.059	*Lifestyle*	0.07	0.04	.057	*Lifestyle*	0.06	0.04	.111
Social net	-0.07	0.03	.035	*Social net*	-0.07	0.03	.034	*Social net*	-0.07	0.04	.068
Gender	-0.10	0.06	.106	*Gender*	-0.08	0.07	.226	*Gender*	-0.09	0.07	.175
				Synonym judgement	-0.001	0.04	.981	*Synonym judgement*	-0.02	0.05	.716
				Fluency	-0.04	0.04	.292	*Fluency*	-0.02	0.04	.556
				Global Cost	-0.01	0.04	.751	*Global Cost*	-0.03	0.04	.534
				Feature Cost	0.01	0.04	.860	*Feature Cost*	-0.01	0.04	.893
				Inhibition	-0.01	0.04	.864	*Inhibition*	-0.02	0.04	.546
				Digit span	0.01	0.03	.677	*Digit span*	0.003	0.03	.918
								Age Synonym judgement*	-0.07	0.05	.148
								Age Fluency*	-0.03	0.04	.459
								Age Global Cost*	0.07	0.05	.143
								Age Feature Cost*	-0.01	0.04	.756
								Age Inhibition*	-0.02	0.04	.496
								Age Digit span*	-0.02	0.03	.627
Model 1: (F_1(5,70) = 2.252, p = .059) Total variance explained: 13.9%				Model 2: (*FChange*(6,64) = .328, p = .920) Variance explained relative to model 1: 2.6%				Model 3: (*FChange*(6,58) = .850, p = .537) Variance explained relative to model 2: 6.8%			

Contributions of all predictors to the Match effect. Global cost = accuracy difference between strong and weak trials within the global association task, Feature cost = accuracy difference between the congruent and incongruent trials within the feature association task.

Mismatch effect

The mismatch effect was not explained significantly by any of the individual predictors. None of the three models reached significance either (see Table 3.2)

Table 3.2 Hierarchical regression (mismatch effect)

model 1				**model 2**				**model 3**			
	B	***SE***	***P***		***B***	***SE***	***P***		***B***	***SE***	***P***
Intercept	0.03	0.03	.281	*Intercept*	0.03	0.03	.393	*Intercept*	0.06	0.04	.102
Age	0.02	0.02	.495	*Age*	0.03	0.03	.319	*Age*	0.03	0.03	.413
Education	-0.02	0.02	.432	*Education*	-0.02	0.02	.349	*Education*	-0.03	0.02	.211
Lifestyle	0.01	0.02	.626	*Lifestyle*	0.01	0.02	.798	*Lifestyle*	0.003	0.03	.904
Social net	-0.001	0.02	.964	*Social net*	-0.002	0.02	.919	*Social net*	0.01	0.02	.599
Gender	-0.08	0.04	.059	*Gender*	-0.07	0.04	.122	*Gender*	-0.08	0.05	.078
				Synonym judgement	0.02	0.03	.418	*Synonym judgement*	0.01	0.03	.662
				Fluency	-0.01	0.02	.657	*Fluency*	0.002	0.03	.940
				Global Cost	-0.05	0.03	.103	*Global Cost*	-0.04	0.03	.136
				Feature Cost	-0.003	0.02	.915	*Feature Cost*	-0.01	0.03	.603
				Inhibition	-0.01	0.02	.726	*Inhibition*	-0.01	0.03	.653
				Digit span	-0.02	0.02	.485	*Digit span*	-0.02	0.02	.403
								Age Synonym judgement*	-0.02	0.03	.478
								Age Fluency*	-0.03	0.02	.238
								Age Global Cost*	-0.04	0.03	.299
								Age Feature Cost*	-0.03	0.03	.276
								Age Inhibition*	0.01	0.02	.765
								Age Digit span*	0.001	0.02	.947
Model 1 Stats: ($F_1(5,70) = 1.077$, p= .381 Total variance explained :7.1%				Model 2 Stats: (*FChange*(6,64) = .866, p= .525 Variance explained relative to model 1: 7.0%				Model 3 Stats: (*FChange*(6,58) = .790, = .582) Variance explained relative to model 2: 6.5%			

Contributions of all predictor variables to the Mismatch effect. Global cost = accuracy difference between strong and weak trials within the global association task, Feature cost = accuracy difference between the congruent and incongruent trials within the feature association task.

Context effect

Finally, none of the included variables were significant predictors of the Context effect (see Table 3.3). None of the three overall models reached significance either.

Table 3.3 Hierarchical regression (context effect)

model 1				**model 2**				**model 3**			
	B	***SE***	***p***		***B***	***SE***	***P***		***B***	***SE***	***p***
Intercept	-0.22	0.10	.024	*intercept*	-0.22	0.10	.029	*Intercept*	-0.13	0.13	.327
Age	0.12	0.08	.120	*Age*	0.05	0.10	.623	*Age*	0.01	0.10	.953
Education	0.06	0.07	.422	*Education*	0.05	0.07	.519	*Education*	0.02	0.08	.821
Lifestyle	-0.10	0.08	.180	*Lifestyle*	-0.07	0.08	.372	*Lifestyle*	-0.05	0.09	.599
Social net	0.10	0.07	.183	*Social net*	0.10	0.07	.194	*Social net*	0.12	0.08	.132
Gender	0.12	0.14	.380	*Gender*	0.12	0.15	.401	*Gender*	0.12	0.15	.439
				Synonym judgement	0.05	0.10	.602	*Synonym judgement*	0.03	0.11	.807
				Fluency	-0.05	0.08	.548	*Fluency*	-0.03	0.08	.729
				Global Cost	0.04	0.10	.697	*Global Cost*	0.06	0.10	.548
				Feature Cost	0.02	0.08	.785	*Feature Cost*	0.02	0.09	.807
				Inhibition	0.09	0.08	.248	*Inhibition*	0.10	0.08	.224
				Digit span	-0.04	0.07	.567	*Digit span*	-0.07	0.08	.372
								Age Synonym judgement*	-0.03	0.11	.769
								Age Fluency*	-0.01	0.08	.909
								Age Global Cost*	-0.17	0.11	.136
								Age Feature Cost*	0.00 3	0.09	.974
								Age Inhibition*	0.01	0.08	.949
								Age Digit span*	-0.05	0.07	.531
Model 1 Stats: (F_1(5,70) = .967, p = .444) Total variance explained: 6.5%				Model 2 Stats: (*FChange*(6,64) = .553, p = .766) Variance explained relative to model 1: 4.6%				Model 3 Stats: *(FChange*(6,58) = .670, p = .674) Variance explained relative to model 2: 5.8%			

Hierarchical regression table showing the contributions of the predictors to the Context effect. Global cost = accuracy difference between strong and weak trials within the global association task, Feature cost = accuracy difference between the congruent and incongruent trials within the feature association task.

3.4 Discussion

In this chapter, we assessed how the older and younger adults tested in Chapter 2 differed on various cognitive abilities including semantic knowledge and control, inhibition, and short-term working memory capacity. Semantic knowledge and fluency scores were larger for older than younger adults. However, older adults performed more poorly in terms of inhibition costs and (some aspects of) semantic control, although they did outperform younger adults in terms

of semantic control accuracy. The language and cognitive variables assessed did not relate to language-production match, mismatch, or context effects tested in Chapter 2.

3.4.1 Semantic knowledge

Corroborating previous research (e.g., Carrol, 2023; Hoffman, 2018; Hoffman et al., 2018; Kavé & Halamish, 2015; Kavé & Yafé, 2014; Verhaeghen, 2003), older adults performed significantly better on the synonym judgement task than the younger adults, suggesting that semantic knowledge was higher in the older than younger adults. Furthermore, composite verbal fluency was also higher in the older adult group. In line with previous research (e.g., Gordon et al., 2018), exploratory analyses presented in the Appendices showed this benefit for older adults was driven by the letter fluency trials rather than the semantic fluency trials. This aligns with a frequently, although seemingly paradoxically, observed pattern in the literature reflecting older adults are more likely to experience difficulties on the semantic than letter fluency task (cf. Gordon et al., 2018). Although letter fluency is often associated with executive control and would therefore be expected to be influenced more strongly by age, previous research has suggested letter fluency relies more heavily on vocabulary knowledge. The finding that our older adults outperformed the younger adults on this fluency task specifically supports, in line with the synonym judgement task and the literature, the interpretation that older adults continue to benefit from their (larger) vocabulary knowledge. In contrast, semantic fluency has been found to be more heavily influenced by word retrieval speed (Gordon et al., 2018). Indeed, our supplementary analyses showed no significant age-group difference here. If anything, older adults performed a little worse than younger adults on this task.

3.4.2 Relationship with the Match effect (Chapter 2)

Chapter 2 showed that both younger and older adults' language production is facilitated by semantically matching contexts. This suggests older adults may continue to use their semantic knowledge to facilitate word retrieval during production. Although individual differences in participants' semantic knowledge and fluency reported in the current chapter were not directly related to the size of the Match effect, it is possible that a certain degree of semantic knowledge is a prerequisite for speakers to benefit from matching contexts. With our older adults on average outperforming the younger adults on semantic knowledge and fluency

tests, this prerequisite seemed to have been met by the older adults as a group and almost all older adults at the individual level. Indeed, only four of the older adults' semantic knowledge scores fell below the mean score for the younger adults. However, in the absence of a direct relationship between semantic knowledge and the Match effect, the exact contribution of semantic knowledge in older adults requires further research.

The finding that older adults showed the same Match (and no Mismatch) effect as younger adults while showing more semantic knowledge argues against previous research suggesting increased (semantic) knowledge in older adults could potentially create more interference (e.g., Ramscar et al., 2014, cf. Hoffman et al., 2018, showing adults with greater semantic knowledge were less coherent in connected speech). If larger semantic knowledge is associated with greater interference, our older adults should have shown a smaller Match (and potentially larger Mismatch) effect than younger adults at the group level.

It is worth further exploring the role of lifestyle variables in relation to semantic connections in sentence contexts. Social network size predicted the Match effect during language production. People with a smaller social network showed greater facilitation from matched contexts. This relationship was small and not significant in all models and does therefore require further research. It is, however, in line with previous research suggesting a relationship between social interactions and language difficulties (Farrell et al., 2014).

3.4.3 Semantic and domain general control

To measure semantic control, we used global association and feature association tasks (Hoffman, 2018). Both showed costs (poorer accuracy and longer RTs) in the conditions associated with higher control demands. In terms of accuracy, both measures showed higher costs in younger than older adults, although this difference was not statistically significant in the feature association task. These accuracy effects were contrary to our predictions regarding older adults experiencing difficulties with semantic control (and the findings observed in Hoffman, 2018). Some previous research has reported age-related increases in motivation and engagement with lab-based tasks (Frank et al. 2015; Jackson & Balota, 2012). This could explain why older adults were less hindered by the more challenging conditions in the semantic control tasks, especially given that there was no time constraint within these tasks. On the other hand, older adults did exhibit a greater RT cost, although only on the feature association measure. The feature association task was the task showing the largest accuracy

and RT costs across age groups, suggesting older adults showed larger RT costs only on the more demanding control task. The combination of accuracy and RT findings suggests the older adults needed more time during the semantic-control task to suppress irrelevant features, but were able to achieve higher accuracy by doing this. Older adults also showed a larger interference Stroop cost. This was especially the case in the verbal Stroop task, where older adults also showed lower accuracy than younger adults (suggesting this larger RT cost was not due to a speed-accuracy trade off). These findings suggest older adults showed poorer semantic (on some measures) and inhibitory control in terms of response times, lending support to the inhibition deficit hypothesis (cf. also Hoffman, 2018).

3.4.4 Relationship with the Mismatch effect (Chapter 2)

Contrary to our expectations, in Chapter 2 neither age group showed an interference cost during mismatching contexts (that were designed to prime a word other than the target). Furthermore, neither semantic nor inhibition costs (measured in the current chapter) were significant predictors of the Mismatch effect in the main analysis. Although some previous studies have shown a relationship between semantic and inhibitory control and certain aspects of language production (e.g., Hoffman et al., 2018), our findings align with previous research showing no direct relationship between inhibition and picture naming times in younger and older adults (Higby et al., 2021). This could suggest that measures related to the types of words older adults use (e.g., speech coherence, Hoffman et al., 2018) could be more closely related to one's ability to suppress interfering information than speed of word retrieval during production (as examined in Higby et al., 2021). We did also expect such a relationship (with semantic and inhibitory control) to arise in our mismatch contexts, which specifically required the production of an unexpected rather than expected target word. With this Mismatch effect not arising at the group level in either age group, however, it is very likely that a stronger mismatching sentence context is necessary for any role of semantic or inhibitory control abilities to emerge.

In this chapter, we did find that older adults' responses were slower during high-control trials in both semantic and inhibitory control tasks, corroborating previous findings that these cognitive abilities can decline in old age (e.g., Hoffman, 2018; Spieler et al., 1996). This suggests that the absence of a Mismatch effect in the group of older adults was not the consequence of recruiting a sample of older adults with particularly high or fully preserved

semantic or inhibitory control. Rather, it suggests that interference in the mismatched sentences might have been too weak to lead to a noticeable impact on language production. Future research will need to study the potential relationship between age-related changes in inhibitory or semantic control and stronger context violations during language production (as also done in Chapter 5). Furthermore, such research might want to include older adults with more difficulties in terms of their semantic and inhibitory control. While age-group differences were observed in the current chapter, they were not present on all semantic control tasks and measures, and older adults might have slowed down their responses to achieve a higher accuracy level. As is common in these types of studies, the older adults (like the younger adults) had a relatively high level of education and a relatively active lifestyle. Research including a wider range of older adults from various backgrounds, including lower socio-economic status, would be more representative of the general population and might be more likely to capture age-related changes in terms of interference effects during cognitive and language-production tasks. Including a larger sample size would also be beneficial, as the current sample size might have limited the power to detect effects related to individual differences in Chapter 3. Furthermore, the older adults were quite young (mean age <70 years). Future research might also want to include older adults with a higher age, including more participants over 75 years old.

3.4.5 Working memory

Working memory capacity did not differ between the two age groups. It is worth noting that forward digit span tasks measuring storage capacity might not be sufficiently sensitive to age-related declines in working memory, relative to tasks that focus on both storage and information manipulation such as the backward digit span task (e.g., Babcock & Salthouse, 1990; Bopp & Verhaeghen, 2005). Furthermore, to adjust the task to an online environment, our participants were able to continue onto longer sequences even if they made errors in shorter trials, which may have influenced performance on this task relative to traditional task versions not allowing this.

3.4.6 Relationship with the Context effect (Chapter 2)

The context effect computed in Chapter 2 was not related to any of the language or cognitive abilities tested, including digit-span performance. This suggests that keeping the question in

mind in a context as compared to naming pictures in isolation was not modulated by working memory capacity as assessed through this task. Sentences used in Chapter 2 were relatively short and simple. It is possible that more complex sentences, or producing words in connected speech and interactions with others, do tax short-term memory capacity more strongly (e.g., Kemper, 1986).

3.4.7 Lifestyle and social network

Finally, older and younger adults showed some differences in terms of their lifestyle although this did not predict the context effects studied in Chapter 2. Overall older adults reported higher engagement in the lifestyle activities than younger adults did. Participating in some of the tested activities (including gardening, walking, attending lectures, or playing card games) has been linked to better cognitive health in old age, and lower rates of cognitive decline. In addition to this, a large proportion of the older participants were educated to at least degree level and reported a similar social network size to the younger adults. Thus, although not directly predictive of our language variables of interest, their education, maintained social network and participation in the lifestyle activities may have supported the older adults in maintaining a similar ability in language retrieval across the contexts to the younger participants. Future research may focus on comparing language retrieval in different contexts in older adults reporting varying levels of engagement in these lifestyle activities, as well as with different social network sizes, and levels of education to try and better understand their impact.

3.5 Conclusion

Within the context of cognitive ageing, our findings from the current chapter highlight the important role of preserved semantic networks within older adults who can continue to benefit from context when retrieving words to respond to questions asked by their conversation partner. Further exploration of how age-related declines in other areas including semantic and cognitive control and verbal short-term working memory capacity present within language production contexts is needed. Chapter 4 therefore further addresses the role of sentence complexity and the potential relationship with working memory. Chapter 5 further examines mismatched sentence contexts and the potential role of semantic control.

Chapter 4: Matched sentence contexts: the influence of context length and semantic prime position on word production

Abstract

Semantic information contained within predictive sentence contexts might facilitate language production processes in younger and older adults (Chapter 2). The current study sought to explore how specific characteristics of such contexts (length and semantic prime position) affect the production of predicted words. 40 younger and 40 older adults completed a sentence completion task (producing one word to complete a verbally presented sentence), followed by two working memory span tasks. We studied the effects of sentence length and the position of the semantic prime (beginning or end of the sentence) on language production. While sentence length effects were not evident in RTs, longer sentences were associated with lower accuracy (i.e., less frequent selection of the primed target word). Furthermore, higher working memory scores were associated with more frequent target word selection, especially in response to longer sentences. With regards to semantic prime position, both younger and older adults' performance within the sentence completion task benefitted from earlier priming more than from later priming. This suggests that listeners benefit from hearing relevant information earlier to prepare their own responses, lending support to a vein of studies showing that verbal response formulation within conversation begins as soon as there is enough information to do so. Furthermore, this chapter again suggests that older adults continue to benefit from semantic information in sentence contexts facilitating their own responses.

4.1 Introduction

Ageing is associated with cognitive changes, including changes in language processes that are pivotal for everyday conversation (Burke & Shafto, 2008). One such process is word retrieval during production, which involves translating mental concepts into words, to produce as verbal output (Friedmann et al., 2013). While older adults know as many or even more words than younger adults (cf. also Chapter 3), selection and retrieval of specific words from amongst close semantic competitors in the lexicon might become more difficult with advancing age (Hoffman, 2018). Conversation could pose further challenges. For example, conversational partners must simultaneously process the speech produced by their partner while formulating

appropriate responses. Older adults have been found to exhibit greater difficulty in processing longer sentence constructions (e.g., Kynette & Kemper, 1986; Kemper, 1986; Feier & Gerstman, 1980), which may impact how they process their conversation partner's sentence and plan their subsequent word production. As shown in Chapter 2, however, both younger and older adults' word retrieval efficiency can benefit from information in preceding sentences matching the upcoming target word. The current chapter further examined how and when younger and older adults can benefit from this matching semantic information. Specifically, previous studies with younger adults have suggested that the magnitude of this semantic facilitation may depend on where in an utterance a semantic prime is presented (Bögels et al., 2015; Corps & Pickering, 2024). The current chapter explored the position of the semantic prime further with an ageing population. It furthermore examined a potential influence of sentence length, considering older adults' reported difficulties processing longer and more complex sentences.

4.1.1 Word production difficulties in older adults

As also reviewed in the previous chapters, there is substantial evidence for age-related declines in word retrieval/ production efficiency (Burke & Shafto, 2004; Lovelace & Twohig, 1990). For example, in confrontational naming studies, older adults are less accurate and slower to produce picture names than younger adults (Barresi et al., 2000; Connor et al., 2004; Goral et al., 2007; MacKay et al., 2002; cf. also Chapter 2 for overall slower naming in older adults).

The Transmission Deficit Hypothesis posits that ageing weakens the connections between lexical and phonological levels of representation. As a result, while the semantic and lexical forms of a word may be active, its phonological constituents may be difficult to access, resulting in production difficulties (Burke et al., 1991). This weakening of connections might especially affect low frequency words/concepts which are not retrieved/ accessed often, making their underlying connectivity (particularly to their phonological constituents) vulnerable to age-related erosion. In addition, low frequency words sometimes have fewer phonological neighbours which may further prevent connection strengthening (Gordon & Kurczek, 2014). Indeed, some evidence suggests that such words are more difficult for older adults to retrieve (Burke et al., 2004; Cohen & Faulkner, 1984; Lovelace & Twohig, 1990). For example, Burke et al. (1991) examined 'tip-of-the-tongue' (TOT) states: instances where a

speaker can describe a word's meaning, and name related words, but cannot recall the specific word they are trying to produce. Older adults experience these TOT states more often than younger adults, but especially when having to retrieve low frequency words (proper nouns such as "**Casablanca**," and low frequency object names such as "**algae**"). While the previous chapters focused on relatively high-frequency words, the current chapter therefore aimed to examine production of a wider range of words, including more low-frequency words. This allowed us to better examine potential age differences for words that might be more sensitive to changes with age.

4.1.2 Working memory and sentence length

In addition to changes within the language system, age-related changes in cognitive processing more broadly can also affect language processes (Burke & Shafto, 2008). Word retrieval/ production difficulties in old age can be attributed to declines in cognitive resources. Generalised slowing models propose that cognitive functioning becomes slower with age, which can affect processing speed (Birren, 1965; Cerella, 1985; Madden, 2001; Myerson et al., 1990; Salthouse, 1985, 1996, 2000), including language-related processes. In addition to changes in inhibition and semantic control (as reviewed in the previous chapters), working memory has also been found to decline with age (Park et al., 2002; Van der Linden et al., 1998; Van der Linden et al., 1994; Waters & Caplan, 2003). This can impact language processing too. For instance, older adults have shown increased difficulties in comprehending complex syntactic structures (Feier & Gerstman, 1980; Lash, 2010; Walsh & Baldwin, 1977), and producing (Kemper et al., 1989), recalling (Kemper, 1987; Norman, et al., 1991), and understanding (Emery, 1985) multiclause sentences that are believed to strain working memory resources. Furthermore, older adults experience difficulties in comprehending sentences with a larger distance (number of words; Gibson, 1998, 2000) between meaningful units. This includes sentences containing relative clauses: clauses attached to an antecedent using a relative pronoun such as 'which' or 'whom' (e.g., "*He helped the man whom the wolf attacked*."). For example, in a self-paced reading task, Liu and Wang (2019) showed that both younger and older adults' comprehension accuracy declined when this distance within sentences increased, and this affected older adults more than younger adults. Components of working memory, including storage and integration, are required in language comprehension, for holding recently processed syntactic information while adding incoming words to it

(Linares & Pelegrina, 2023). Older adults' increased difficulties with processing of longer and more complex sentences have therefore been explained with regards to working memory declines that occur in old age.

Although the focus has been on language comprehension, working memory resources could also influence production, in particular when considering the interaction between comprehension and production in a conversation, which places demands on a speaker to simultaneously process and plan/produce speech (Meyer, 2023). Within the context of conversation, cognitive resources (such as working memory) may be required to process sentences produced by the conversation partner, to plan one's own response, and to then produce the corresponding response. These demands may increase when speakers are required to process more complex speech (e.g., utterances that are longer or comprise relative clauses). These complexities might have consequences for the speed of one's own word retrieval during production if much capacity is required for the conversation partner's speech processing, thus leaving insufficient or lower resources for subsequent language production in response to the conversation partner's speech. The current chapter therefore examined the potential influence of sentence length, in relation to its potential influence on cognitive resources, during language production in older adults, as further explained in section 4.1.4.

4.1.3 Semantic priming/ prediction through sentential contexts

Although processing other people's speech while preparing your own utterances can be demanding, it can also include information that can facilitate a speaker's responses. Specifically, words that relate to the speaker's own target production can facilitate word retrieval in both younger and older adults (see Chapter 2; Shao & Rommers, 2020). In Chapter 2 and Shao and Rommers (2020), participants named pictures (e.g., of a '**present'**) more quickly when they were preceded by a (matched) question (e.g., "*What did she wrap before Christmas*?") than when they were preceded by a neutral question such as "*What did grandfather perceive*?". These facilitation effects in matching sentence contexts were comparable for younger and older adults (Chapter 2). These findings can be explained through semantic priming. The matched context comprises words (i.e., "wrap", "Christmas") that are connected to the target ('**present**') on a semantic level. Therefore, activation of these words spreads to related concepts, allowing those words to be accessed and retrieved more easily. In relation to priming, speakers and listeners might use sentence contexts to predict upcoming

words, either as an automatic component of language processing or as an optional mechanism (cf. Pickering & Gambi, 2018). During comprehension, it has been proposed this prediction mechanism is generated through use of the production system (e.g., Lelonkiewicz et al., 2021), suggesting prediction during comprehension and production are closely related.

The use of semantic information in sentence contexts to facilitate language processes appears to be fairly robust, even in old age. For example, older adults' picture naming times are also faster following the presentation of semantically related words, (e.g., Balota et al., 1999; Faust, Balota et al., 2004) and older adults can sometimes even benefit more from semantic priming than younger adults (Laver & Burke, 1993). This facilitation also extends to word retrieval in sentence contexts as shown through similar Match effects for younger and older adults in Chapter 2. These findings suggest older adults can continue to use their intact semantic knowledge to make word retrieval more efficient during language production (Burke et al., 1991).

4.1.4 Semantic prime position

If both younger and older adults' word retrieval/ production during production can benefit from semantic information provided by the conversation partner, the question arises *when* speakers can benefit most from this information. In other words, does it matter where in the context the semantically predictive information is placed? Previous research with younger adults suggests that early presentation of critical information within their conversational partners' speech has the largest influence on subsequent production (Barthel et al., 2016; Bögels et al., 2018; Bögels et al.,2015). For example, Bögels et al. (2015) used sentences where the critical information was presented earlier or later in the sentence (e.g., "*Which character, also called **007** appears in the famous movies*" (earlier) vs "*Which character from the famous movie is also called **007***" (later)). In these sentences, the cue "007" is the critical information needed to formulate the response '**James Bond**'. Participants' responses were initiated sooner when critical information was presented earlier compared to when it was presented later in the utterance. This may result from the critical information having more time to prime the target word and/or for the listener to form predictions about upcoming words. When it is placed too close to the end of the utterance, there may not be sufficient time for activation to spread to the target word and aid in word retrieval. Although these studies focused on younger adults, this effect may be particularly pronounced in older adults because generalised

slowing in old age might delay priming effects. For example, Howard et al. (1986) manipulated the interval between semantic primes and the target word in a lexical decision task. Younger adults displayed semantic facilitation both when the prime was presented earlier (1000ms before lexical decision) and later (150ms before lexical decision). However, older adults only benefitted from the semantic prime when it was placed earlier and not when it was placed later, suggesting that in the ageing lexicon more time might be needed for semantic priming processes to occur. Therefore, older adults' production in particular might benefit more from semantic information occurring earlier in the conversation partner's sentence.

Alternatively, some evidence suggests that speakers may not begin planning their own speech until the end of their conversational partners' turn. For example, Sjerps and Meyer (2015) used a dual-task paradigm wherein participants completed a finger tapping task whilst they listened to picture descriptions. Eye-tracking data combined with declines in participants' finger tapping performance suggested that participants only started planning their speech response shortly before they were required to speak. These findings suggest that speakers might preserve their cognitive resources during speech comprehension, and only allocate these resources to speech production processes when their turn in the conversation is approaching. Although this has not been studied directly, this might apply even more strongly to older adults if they experience increased needs for preservation of cognitive resources during sentence processing (i.e., if speech processing takes up more resources and therefore does not leave sufficient resources for early production planning). The current chapter therefore manipulated the place of semantic information in the preceding sentence. This way, we studied whether older adults would benefit more from earlier presentation (potentially allowing more time for semantic priming and/or prediction to develop) or from presentation later in the sentence (potentially closer to the planning of their own production).

The influence of semantic primes, and their position in a sentence, might also relate to sentence length. The strain on cognitive resources when processing long complex sentences may have consequences for the presence (or lack) of facilitation through semantic priming/prediction. For example, when working memory was taxed (through a high-load dot memory task), semantic priming effects disappeared in younger adults (Heyman et al., 2014). These findings suggest beneficial effects of semantic information may be reduced in more demanding contexts. This might be even more pronounced in older adults if they are more affected by sentence complexity influencing working memory resources. Furthermore,

sentence length and complexity might interact with prime location effects, as these effects might become more pronounced when early semantic primes are further apart from the speaker's production in longer, more complex sentences (compared to shorter sentences).

4.1.5 Current study

The current study therefore aimed to examine how sentence length and semantic prime position influence word production in younger and older adults. It used a sentence completion task presenting participants with incomplete sentences, after which they verbally produced the word they believed should complete the sentence.

The first question aimed to investigate how processing speech (sentences) of different lengths affects speed of subsequent word retrieval during production in younger and healthy older adult groups. Longer sentences place greater demands on cognitive resources (working memory) than shorter sentences and we therefore expected slower sentence completion times for longer than shorter sentences in younger and older adults (main effect Sentence Length). We expected this effect to be largest for older adults (interaction Age group x Sentence Length), possibly as a result of the availability of limited cognitive resources in this age group (Feier & Gerstman, 1980; Liu & Wang, 2019). To further examine the role of working memory, we also measured working memory capacity. We expected these scores to be lower for older adults (Beese et al., 2017; Bopp & Verhaeghen, 2005; Salthouse, 1994). If better working memory is related to more efficient processing of longer sentences (potentially especially in older adults), we expected working memory performance to be a significant predictor of the sentence-length effect.

The second question investigated if and how word retrieval/ production efficiency is affected by the location of the semantic prime (early or late in the preceding sentence). We hypothesised two possible directions of this effect (main effect of Prime Position). On the one hand, previous research with younger adults (Bögels et al., 2015; Corps & Pickering, 2024) suggests that earlier presentation of 'critical' information in the conversation partners' speech allows speakers to begin formulating their upcoming utterances sooner and allows them to respond more quickly to questions. In line with this finding, semantic primes positioned at the beginning of our sentences may lead to faster word completion times in younger and older adults. On the other hand, we may observe the strongest priming effects (fastest production

times) when there is a shorter interval between the prime and the to-be-produced target word (later prime position) if speakers start planning their own utterance towards the end of the conversation partner's speech. In this case, the target word would be freshly activated when the interval between prime and target is short (Rossell et al., 2003) while an early prime may result in a decay of target-word activation by the end of the sentence. Regardless of the direction, such effects of Prime Position might be more pronounced in older adults if they benefit more from more time to prepare their response after an early prime and/or from sufficient time for semantic priming to build up (direction 1) or if they benefit more from less target-word activation decay after a late prime (direction 2).

Any effect of Prime Position might be most pronounced in longer sentences (Interaction Prime Position x Sentence Length). In short sentences, there is little temporal difference between early and late primes. This can reduce differences in the magnitude of the prime effect as well as differences in planning time. In the longer sentences, the distance between the prime in the beginning versus end of sentences is larger, and as such effects of prime position might be more pronounced. This, in turn, might also mean that any age-group differences are most likely to occur in the longer sentences (Interaction Age Group x Prime Position x Sentence Length).

4.2 Methods

This study was pre-registered on the Open Science Framework: https://osf.io/dnk9g/. The data files can be accessed via the following link:
https://osf.io/dnk9g/?view_only=62dba7f9931a45b3bdb1eff988132d8c

4.2.1 Participants

Ethical approval was obtained from the Department of Psychology at the University of York. Eighty English-speaking monolinguals completed the study. The groups of younger and older adults were matched on sex ratio, number of years of formal education received, and the number of participants within each age group who had completed at least an undergraduate degree. Details of participants completing the picture naming study are provided in Table 4.1. Older adults (aged 65-80) were recruited through prolific.co (n = 40), while younger adults (aged 18-35) were recruited through SONA (the university's internal participant recruitment system; n = 16) and prolific.co (n = 24). Participants recruited through Prolific received

monetary compensation, while those who were recruited through SONA received course credit for their participation

Our sample size was based on a previous study exploring sentence complexity effects in healthy ageing (Lash, 2010). Their study included 23 younger and 25 older adults, and reported a significant main effect of sentence complexity with a large effect size (partial $\eta 2$ = .43). However, given that previous literature has not studied age-group differences in prime position effects, we could not conduct further power analyses. We therefore opted for a larger sample size than previous studies comparing sentence complexity effects, to examine prime position effects and potential relationships with working memory too. Furthermore, this sample size was in line with the sample sizes used in the other chapters.

Participants had to meet the following criteria: they were not fluent in any languages apart from English, did not use a hearing aid, and had normal or corrected-to-normal vision. They also should not have used medication affecting their concentration in the last three months and did not have a neurodegenerative/cognitive impairment or a language/reading disability (self-reported). Given that the study was conducted online, we were not able to use an assessment of cognitive functioning such as the ACE-III. However, we used Prolific's screening criteria to only invite participants without a history of head injury, cognitive impairment, or dementia. Participants also confirmed that they met the inclusion criteria detailed above in a background questionnaire which they completed while taking part in the study.

In addition to the 80 participants included, full datasets from nine younger adults and fourteen older adults were removed. Five younger participants were removed due to them not meeting the study's inclusion criteria. A further three were removed as they did not pass one or both of the attention checks included in the study; and one participant was removed for scoring less than 50% accuracy on the sentence completion task. This accuracy cut-off was pre-registered and lower than what we used in the other studies, to account for participants not having to name specific pictures. Four older adults were removed due to failing one or both attention checks (one of these participants also did not meet the study's full inclusion criteria). Six were removed due to scoring less than 50% accuracy on the sentence completion task; and a further three were removed as their audio files from the sentence completion task were empty recordings. Finally, one older participant was removed as they did not follow the

instructions for the sentence completion task. We replaced all removed participants to ensure that the pre-registered sample size was met.

Table 4.1 Participant details

Participants	*N*	Age	Social Network	Sex			Formal Education	Graduates
				M	F	Non-binary		
Younger	*40*	24.88 (6.35)	34.65 (53.64)	15	24	1	15.23 (4.35)	47.37%
Older	*40*	68.15 (3.30)	33.46 (30.83)	15	25		14.4 (3.30)	47.06%

Demographic details of younger and older participants. Number of participants in each age group; Age (mean number of years, and standard deviation); Social network size: mean number of people (and standard deviation) participants reported having regular contact with in the past 6 months including: close friends, family members, neighbours, co-workers, school/childhood relations, people who provide a service, and others; Sex (total number of participants belonging to each category); Formal Education (mean number of years of formal education received, and standard deviation)*;* Graduates (percentage of participants who had completed at least an undergraduate degree). The two age groups did not differ significantly in terms of social network size (p = 0.90) or years of formal education (p = 0.34).

4.2.2 Design

Participants completed a sentence completion task wherein they produced verbal responses to complete pre-recorded sentences missing their final word. The task incorporated a mixed design with Sentence Length (Long/Short, see Table 4.2 for examples) and Semantic Prime Position (Early/Late) as within-subject variables and Age group as the between-subject variable. The main dependent variable was response onset times (ms), defined as the onset of verbal responses relative to when a cross appeared on a blank screen immediately after the recordings finished playing on each trial. We also analysed accuracy and examined the type of non-target responses made. Contrary to Chapter 2, we asked participants to complete a sentence with the word they thought fitted best, without asking them to name specific

pictures. This approach was chosen to include low-frequency words, which have been argued to be more influenced by ageing than the high-frequency words that are needed to create depictable stimuli for a picture naming task. Unlike Chapter 2, we therefore also did not include a baseline (neutral) condition in the sentence completion task to which we could compare RTs following the experimental (predictive) sentences. It was not possible to include a neutral condition in this study owing to the fact that we did not use pictures, and therefore we could not determine what a target word would be in neutral sentences. Therefore, while we included sentences of a type comparable to the Match sentences in Chapter 2, we could not compute a Matching effect as such.

4.2.3 Materials

One hundred and one sets of sentence-target word pairs were created. Low-frequency target words (*range* log frequency = 0.16 – 2.55, *M* = 0.72, *SD* = 0.35) were paired with four different incomplete sentences. As intended, the log frequencies of these words were lower than the mean frequency of the target words in Chapter 2 (Chapter 2: Stimuli set 1: log frequency *M* = 1.58, *SD* = 0.55; set 2: *M* = 1.46, *SD* = 0.60). The four sentences belonged to the following categories: long prime beginning, long prime end, short prime beginning, short prime end (see Table 4.2 for examples). In general, we kept the number of active and passive sentence constructions equal between the four conditions. The sentence content was kept similar across these four conditions too. Participants only saw one sentence from each set since all stimuli within a set were greatly similar, and thus this could have created priming effects.

Overall sentence length (mean number of words) differed significantly between the long and short sentences (long *M* = 19.91 words, *SD* = 4.32, *range* = 12-31, *M* short sentences = 8.34 words, *SD* = 2.18, *range* = 4-15, *p* = <.001). Sentence length did not differ significantly between the two types of short sentences (prime beginning *M* = 8.34 words, *SD* = 2.20*, range* = 4-15; prime end *M* = 8.34 words, *SD* = 2.18*, range* = 5-15, *p* > .999); nor between the two types of long sentences (prime beginning *M* = 20.00 words, *SD* = 4.26*, range* = 12-31; prime end *M* = 19.82 words, *SD* = 4.40*, range* = 12-31, *p* = 0.770).

We constructed the sentences such that in the “early priming” sentences, there was a larger distance between the prime and the end of the sentence (i.e., the point at which the participants started naming the target word) compared to the “late priming” sentences. This was computed separately for the short sentences (early prime *M* distance (number of words)

from sentence's ending = 5.57 words, *SD* = 2.01; late *M* = 2.49 words, *SD* = 1.35, *p* = <.001) and for the long sentences (long early *M* = 17.24 words, *SD* = 4.24; late *M* = 2.50 words, *SD* = 1.37, *p* <.001).

The sentences were recorded to reflect natural speech as much as possible by a female English speaker. They were pre-processed using Praat (Boersma & Weenink, 2022) to add 50ms to the beginning and end of each recording and to scale all recordings to 60dB. Background noise was also reduced using Audacity® version 3.0.0.

Table 4.2 Example stimuli

Length	Condition Prime placement	Example
Short	Beginning	*"The **birthday cake** was covered with twenty …"*
	End	*"They covered the **birthday cake** with twenty …"*
Long	Beginning	*"The **birthday cake**, which the guest personally felt lacked flavour, was covered with twenty …"*
	End	*"They covered, though the guest personally felt it lacked flavour, the **birthday cake** with twenty …"*

An example set of stimuli used in the sentence completion task. The target word was "candles" (prime: '**birthday cake'** shown in bold).

Pilot studies

We ran two pilot studies to make sure the sentences strongly predicted our intended target word (or a similar word) and to make sure the sentences were evaluated as being grammatically correct. Pilot study 1 was completed by healthy younger (*n* = 5, age (years) *range* = 18 – 19, *M* = 18.40, *SD* = 0.55) and older adults (*n* = 5, age (years) *range* = 67 – 76, *M* = 70.60, *SD* = 4.16) not taking part in the main study. They completed a sentence completion task similar to the main task used in the experiment, with a subset of stimuli. For each intended target word, participants heard the four possible sentences belonging to the four combinations of length and prime position conditions (64 sentences in total). They also heard 16 long and 16 short neutral sentences that were designed to not prime a specific word.

Both younger and older adults performed the sentence completion task to an adequate standard (younger adults produced the target word or a comparable response in 83.13% of trials (*SD* = 11.61) while older adults did this in 63.44% of trials (*SD* = 15.92). Unlike the main task, the pilot presented each sentence context within the set of four, which potentially allowed participants to repeat the same responses throughout. However, accuracy rates in both younger and older adults were preserved even when just considering responses to the first sentence of each quartet (Younger *M* = 71.77%, *SD* = 8.41; Older *M* = 60.50%, *SD* = 7.55). We removed constructions with low accuracy in the pilot for the main study.

We also checked if participants' responses were faster in the primed sentences than in the neutral sentences. The main study did not include neutral sentences and we therefore wanted to make sure in the pilot that participants did in fact benefit from seeing a prime. This was indeed the case, with faster responses to the experimental prime sentences than to the neutral sentences.

In the second pilot with older adults (*n* = 3, age (years) *range* = 65 – 75, *M* = 69.67, *SD* = 5.03), and younger adults (*n* = 4, age (years) *range* = 18 – 20, *M* = 19.00, *SD* = 0.82) we checked the full set of stimuli. Participants completed a sentence-completion task before providing ratings of how likely each target ending was to complete each sentence on a scale of 1-7, and indicated whether each sentence-ending pair was grammatically correct or not. We used this pilot to remove sentences with low average likeliness ratings (as we only wanted to include highly predictable sentences) and/or low grammaticality ratings before the main study. Working with the final stimulus set, participants in the main study also provided likeliness ratings for the sentences they completed in the main task, as well as for 15 mismatching filler sentences and 15 neutral filler sentences only included in this likeliness rating task to avoid participants adjusting their likeliness scores in response to only seeing highly likely sentences. As expected, participants in both age groups in the main study rated the experimental sentences as being highly likely (scale 1-7: Younger *M* = 6.76, *SD* = 0.31; Older *M* = 6.78, *SD* = 0.35). Although this was not relevant for the current study and therefore not analysed further, these scores were higher than in the mismatched filler (Younger *M* = 1.85, *SD* = 0.75; Older *M* = 1.28, *SD*= 0.42), and the neutral filler trials (Younger *M* = 2.79, *SD* = 0.99; Older *M* = 2.38, *SD* = 0.80). This confirmed that our sentence materials were indeed highly predictive of our chosen target endings.

4.2.4 Procedure

The experiment was conducted on Gorilla.sc (Anwyl-Irvine et al., 2020). Participants first read the information sheet and provided informed consent. Next, they completed a sound check to ensure that they could record audio files and to adjust their device's volume to hear the sentences.

In each trial of the sentence completion task, participants were instructed to listen to the presented sentence, which was missing its final word. They were asked to verbally produce the word they believed completed the sentence. During each trial, participants first viewed a fixation cross (500ms), followed by a blank screen while they heard the pre-recorded sentence. This was followed by another fixation cross (which was presented for 4000ms). While this fixation cross remained on screen, participants' verbal response was recorded. After 4000ms, the screen automatically progressed on to the next trial. It was emphasised to participants to only start producing their response when the fixation cross appeared on screen, in order to ensure that they had finished listening to the sentence and for their naming onset times to be accurately determined. Despite this, some of the participants still consistently begun producing their responses too quickly ($\leq$ 300ms, as preregistered, these trials were removed). Furthermore, participants who started producing their responses too quickly on $\geq$50% of trials were removed from the study (see *Participants* section). Participants who began naming too early but whose total number of trials did not drop below 50% after removal of these invalid trials were still included in the study.

A sentence completion task (instead of a picture naming task similar to that used in Chapter 2) was adopted as it was deemed more appropriate for eliciting the production of low frequency words (i.e., low frequency words are often less imageable, and thus more difficult to depict through pictures).

Next, participants completed the likeliness ratings as described in the Materials section. Embedded within the likeliness ratings task were two attention checks (e.g., "now select 5"), one at the beginning and one at the end, to make sure participants were paying attention (those who failed this were excluded from the study, see "Participants" section). Participants then also completed two tasks measuring working memory: The Operation Span Task (OSPAN, Turner & Engle, 1989) and the Reading Span Task (RSPAN, Daneman & Carpenter, 1980). These tasks were selected because they somewhat mimic the working memory demands that are present in conversation, such as simultaneous storage of incoming information whilst carrying

out other cognitive operations (similar to formulating verbal responses whilst keeping in mind the speech produced by one's conversational partner). In the OSPAN task, participants viewed blocks of two to six trials. Within each trial, they were instructed to judge the correctness of a maths equation (e.g., (3 x 4) - 3 = 9) by pressing the 'L' key if the presented equation was correct and 'R' if it was incorrect. After this, participants briefly saw a single digit number on the screen before the subsequent trial. At the end of each block, participants were prompted to verbally recall the digits presented at the end of each trial in the block. In total, participants completed 60 trials.

In the RSPAN task, participants viewed a mixture of sensical sentences (e.g., "*During winter you can get a room at the beach for a very low rate*") and non-sensical sentences (e.g., "*Andy was stopped by the policeman because he crossed the yellow heaven*") within blocks of two to six trials. On each trial, participants were instructed to read the presented sentence out loud before deciding if it made sense (by presenting the 'L' key if the sentence made sense, or 'R' if it did not). Participants then briefly saw a word (e.g., '**ring'**). At the end of each block, they were prompted to verbally recall all of the words presented subsequently to each trial in the block. In total, participants completed 60 trials. Half of the participants in each age group completed the RSPAN task first, whilst the other half completed the OSPAN task first.

Finally, participants also completed a background questionnaire to provide demographic details and to ensure they met the inclusion criteria (see Participants section). They also completed a social network questionnaire (adapted from Bruine de Bruin & Parker, 2020). To gauge social network size, participants were asked to provide the number of people they had frequently been in contact with over the past 6 months including: friends, family, neighbours, people providing a service and others. Finally, participants were asked some questions regarding their motivation in the study (adapted from Jun et al., 2017) to get a better understanding of why older and younger adults took part in the study, and potential differences between them. Participants were asked to rate their level of motivation during the sentence completion task on a scale of 1-9 (1=very low, 9=very high). Participants were also asked the following question: "To what extent are you participating in this experiment for the following reasons?" followed by five Likert scale items. (1) "I want to help science"; (2) "I want to learn how I respond to cognitive tasks"; (3) "I want to compare myself to others"; (4) "I am bored"; (5) "For fun"; (6) "For course credit/ monetary compensation"; (7) "I am curious about science and/or language research" Finally, participants competed the 4-item Intrinsic

Motivation Inventory (IMI, Ryan & Deci, 2000) to gauge how the participants felt about the sentence completion task. This involved rating how fun, boring, interesting and enjoyable they found the task.

4.2.5 Data analysis

Sentence Completion task

An accurate response in the sentence completion task was either the intended/expected target word, or any other response that was a grammatical and viable ending of the presented sentence. For instance, instead of the target "**cheese**" we would also accept words like "**onions**" and "**nuts**". Answers scored as inaccurate were those that were not viable endings of the sentence (e.g., "**pasta**" or "**flour**" when the sentence referred to grating food). We also did not accept answers that were repetitions of words already included in the sentence. Trials containing inaudible responses were also scored as inaccurate.

Overall accuracy (including both exact target word responses and alternative accepted answers) in the sentence completion task was >50% for all participants (*M* older adults 88.96%, *SD* = 7.75; *M* younger adults 88.27%, *SD* = 7.06). When including only exact target word responses, accuracy was still above 50% in both age groups (*M* older adults 72.70%, *SD* = 11.55; *M* younger adults 72.67%, *SD* = 9.25). This was considerably lower than in Chapters 2 and 3 (as expected) since the sentence completion task was open to participants to producing any word (and thus the likelihood of producing an incorrect response was higher in this task), whilst in the picture naming-task the word produced was determined by the presented picture.

Naming RTs were determined using Checkvocal (Protopapas, 2007). RTs <300ms or more than 2.5 *SD* above or below the mean per participant and per condition were removed, using the trimr package (Grange, 2015), removing 10.52% of accurate trials. Five hundred and forty three trials (7.58% of trials) were removed due to participants producing a verbal response too early (<300ms); while two hundred and ten trials (2.93% of trials) were removed due to RTs being 2.5 *SD* above or below the mean per participant and per condition. Shapiro-Wilk tests conducted using raw RTs confirmed the data relating to the longer sentences were normally distributed (*p*s >.06). Data relating to the shorter (prime end) sentences within the older group; as well as the shorter (prime beginning) sentences in both age groups were not

normally distributed (*p*s <.05). Visual inspection of the Q-Q plots showed that these data followed a relatively straight line, which suggested a reasonable correspondence to the theoretical normal distribution.

A 2x2x2 mixed ANOVA was used to analyse the RT data. Through this, we sought to establish whether there was a main effect of Age, Sentence Length and/or semantic Prime Position, as well as possible interactions between these variables. Although an initial check of the raw RTs did not actually reveal a significant overall RT difference between older and younger adults ($F_1(1,78) = 2.370$, $p = .128$, $\eta_p^2 = .029$; $\beta = 0.082$, SE = 0.053, t = 1.570, $p = .120$), we had pre-registered that we would run the analysis on z-scored RTs to account for any overall slowing in older adults. The z-scored RTs are therefore still used in all subsequent analyses. As pre-registered, given the null results observed, we also ran Bayesian ANOVAs to further investigate the main effects of sentence length and prime position. We compared models including Prime Position to a null model (to examine evidence for/against a Prime Position effect) as well as including Sentence Length to a null model (to examine evidence for/against a Sentence Length effect). We will report these results in the form of "BF_{01}", showing the evidence for the null hypotheses. Values below 1 indicate evidence for an effect while values above 1 indicate evidence for no effect. As pre-registered, we also ran a linear mixed effects (LME) analysis, which is reported next to the ANOVA in the Results, for RTs, and a generalised LME for accuracy. In addition to Sentence Length, Prime Position, and Age, this included Working Memory. For each participant we computed the mean proportion of correctly recalled items in each of the two span tasks. Next, we calculated the mean of these two scores to determine each participant's composite working memory score. For ten participants (younger *n* = 4, older *n = 6*), we were not able to compute a WM score due to technical issues (no audible sound recording and/or recordings being too short). We first started the LME analysis using R (4.4.1; lme4 package version 1.1.35) with a full model, including participant and item as random effects with intercepts and slopes. Upon non-convergence, correlations between intercepts/slopes were removed, followed by slopes that explained the lowest amount of variance until convergence was reached. For the RT analysis, the model converged with participant and item intercepts, all participant slopes, and all item slopes apart from Age x Sentence Length. Accuracy was only analysed through the GLME and not through an ANOVA, as the GLME is more suitable for this type of binary DV analysis. The accuracy model converged with participant and item intercept, all participant slopes, and item

slopes for Prime Position, Sentence Length, Age, Prime Position x Sentence Length, and Age x Prime Position x Sentence Length.

4.3 Results

4.3.1 Accuracy sentence completion task

Before presenting the main analyses looking at the RT data, we examined the accuracy data as participants made relatively many mistakes. First, we conducted a generalised linear mixed-effect analysis on accuracy in general. As discussed above, accurate responses included target words and other acceptable responses. Older and younger adults did not differ significantly in terms of overall accuracy (β = 0.276, *SE* = 0.193, *z* = 1.425, *p* = .154, see Table 4.3.). There was a significant effect of prime position (β = -0.426, *SE* = 0.165, *z* = -2.585, *p* = .010), reflecting higher accuracy (*M* = 90.67%, *SD* = 7.64) when the prime was presented at the beginning compared to the end of a sentence (*M* = 86.57, *SD* = 8.55). There was also an effect of sentence length (β = -0.503, *SE* = 0.109, *z* = -4.610, *p* < .001), with higher accuracy in shorter sentences (*M* = 90.62%, *SD* = 7.00) than in longer sentences (*M* = 86.6%, *SD* = 9.25). Accuracy was related to working memory (β = 0.268, *SE* = 0.093, *z* = 2.884, *p* = .004), with participants with a higher working memory score being more accurate in the production task too. This interacted with sentence length (β = 0.176, *SE* = 0.087, *z* = 2.015, *p* = .044). The relationship with working memory was more pronounced in long sentences (*p* < .001) than in short sentences (*p* = .080). There were no further interactions, including no interactions with age (all *p*s > .19).

We also counted non-target responses that still fitted the sentence as intended, given that there was no specific picture to be named, and examined the type of non-target responses older and younger adults made. These are described in Table 4.3. In both age groups, the majority of non-target responses were still semantically related to the sentence context, and this percentage was comparable for the two age groups. The non-target responses were on average higher in frequency than the expected targets (experimental stimuli log frequency *range* = 0.16 – 2.55, *M* = 0.72, *SD* = 0.35), but this non-target word frequency was similar across age groups (older log frequency *range* = 0.13 – 3.03, *M* = 1.19, *SD* = 0.62; younger log frequency *range* = 0.13 – 3.17, *M* = 1.19, *SD* = 0.61).

Table 4.3 Accuracy scores (sentence completion task)

Age group	Overall Accuracy (%)	% exact target responses	% of non-target-semantically related	% of non-target-semantically unrelated
Younger	88.27%(7.06)	72.67% (9.25)	62.78% (16.62)	20.31% (10.06)
Older	88.96% (7.75)	72.70% (11.55)	62.34% (11.35)	17.53% (9.12)

Accuracy data for the younger and older adults. The first column shows the mean overall accuracy. The second column shows the percentage of target responses per age group. The final two columns show the percentage of non-target responses that were semantically related to the target versus unrelated. Note that the latter two columns do not add up to 100% as "non-target responses" includes no responses.

4.3.2 Reaction times sentence completion task

Table 4.4 Reaction time data (sentence completion task)

Condition	Younger adults	Older adults
Short sentences		
Prime beginning	874.75 (272.29)	964.45 (257.35)
Prime end	906.76 (206.59)	978.89 (279.81)
Long sentences		
Prime beginning	869.70 (257.73)	937.18 (266.69)
Prime end	902.05 (262.79)	981.05 (242.08)

Mean RTs and (standard deviations) per age group and per condition

Sentence length

It was hypothesised that the longer sentences would be associated with longer naming times. Neither analysis showed a significant main effect of sentence length. (ANOVA: $F_1(1,78) = .309$, $p = .580$, $\eta_p^2 = .004$; LME: $\beta = -0.022$, $SE = 0.017$, $t = -1.304$, $p = .196$). Bayesian analyses also

strongly supported the null hypothesis with regards to there not being a length effect (BF_{01} = 7.76, error = 0.11%). Interestingly (but not significantly), responses were somewhat faster following *longer* sentences (*M* = 923.21ms, *SD* = 240.45) than shorter sentences (*M* = 934.11ms, *SD* = 234.33). The (absence of a) sentence length effect did not differ between younger and older adults ($F_1(1,78) = .057$, $p = .811$, $\eta_p^2 = .001$; LME: β = -0.036, *SE* = 0.028, *t* = -1.291, *p* = .202; see Figure 4.1.).

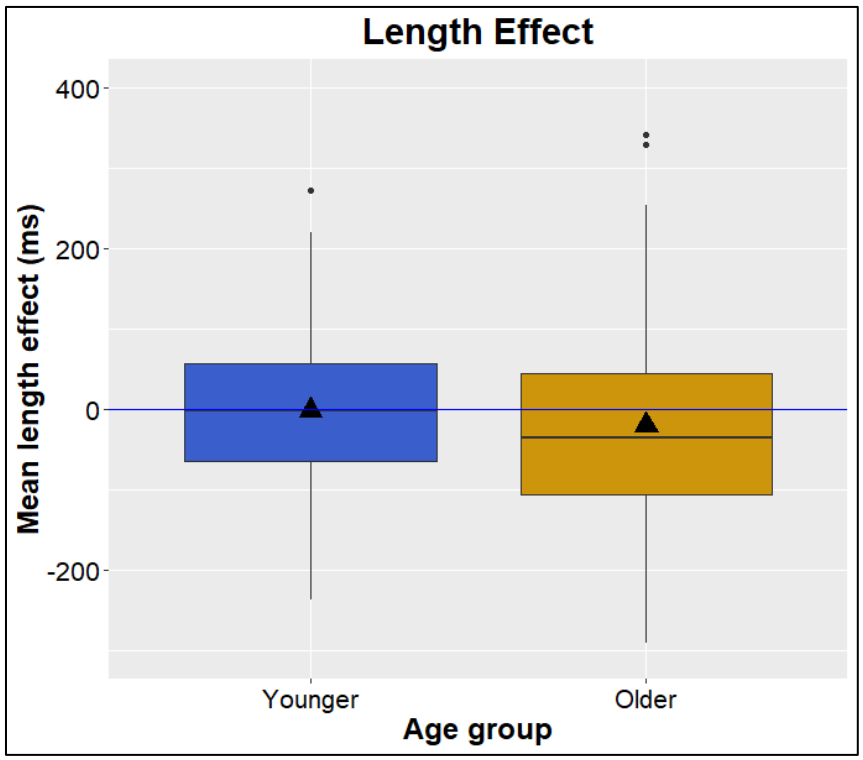

Figure 4.1 Box plots displaying mean length effect by age group (difference in mean RTs between longer and shorter sentences). Values below 0 denote mean RTs that were longer in the short sentences condition (left panel: younger adults, right panel: older adults). Box plot height denotes interquartile range; vertical lines below plots denote 25th percentile, while lines above denote 75th percentile. Black horizontal lines denote the median, triangles denote the mean. Black dots represent outliers.

Predictive role of working memory in sentence length effect

Mean composite working memory scores were numerically higher in the younger adult group (*M* =76.83%, *SD* = 15.74) than the older adult group (*M* = 61.85%, *SD* = .26.01). However, this age group difference in composite working memory score was not statistically significant (LME: β = 0.042, *SE* = 0.064, *t* = 0.652, *p* = .516). Working memory was not correlated with Sentence Length (LME: β = -0.017, *SE* = 0.015, *t* =-1.139, *p* = .260). Finally, there was no significant interaction between Age, Sentence Length, and Working Memory scores (LME: β = 0.049, *SE* = 0.029, *t* =1.691, *p* = .096). No other interactions were observed with working memory scores either (*p*s > .08).

Semantic prime position

Numerically, RTs were shorter when semantic primes were placed at the beginning of sentences (*M* = 912.00ms, *SD* = 247.30) compared to at the end of the sentence (*M* = 944.64, *SD* = 230.45), in both age groups (see Table 4.4). The main effect of prime position did not reach statistical significance in the ANOVA analysis (F_1(1,78) = 2.991, *p* = .088, η_p^2 = .037), however, it was significant in the linear mixed effect analysis (β = 0.059, *SE* = 0.020, *t* = 2.996, *p* = .004). The facilitatory effect of earlier prime placement was similar across the two age groups (Older *M* = -32.58ms, *SD* = 161.94, Younger *M* = -32.71, *SD* = 148.92; F_1(1,78) = .011, *p* = .917, η_p^2 < .001; β = -0.024, *SE* = 0.033, *t* = -0.722, *p* = .473, see Figure 4.2). The Bayesian ANOVA moderately supported the null hypothesis with regards to the effect of semantic prime position (BF_{01} = 4.65, error = 0.07%). We suspected the difference between the ANOVA and LME analyses might be related to items being included in the LME but not in the ANOVA analysis. We therefore also ran an F2 ANOVA (by-item means, not pre-registered), which indeed also showed a significant effect of Prime Position (F_2(1,196) = 9.001, *p* =.003, η_p^2 =.044). The Bayesian by-item analysis showed weak support for the alternative hypothesis (BF_{01} = 0.737, error = 0.016%).

*S*entence Length did not interact with semantic Prime Position (F_1(1,78) = .240, *p* = .626, η_p^2= .003; β = 0.024, *SE* = 0.027, *t* = 0.880, *p* = .382), nor was there a three-way interaction with Age (F_1(1,78) = .229, *p* = .634, η_p^2 = .003; β = 0.026, *SE* = 0.056, *t* = 0.465, *p* = .644).

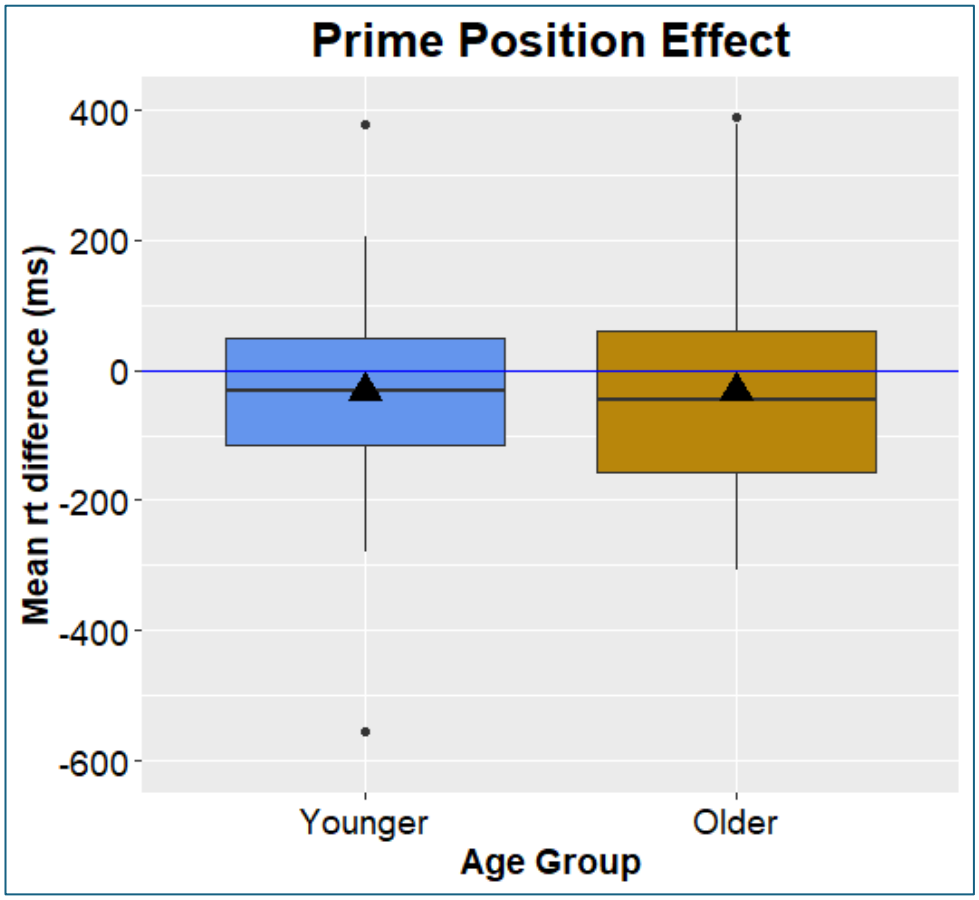

Figure 4.2 Box plots displaying mean Prime position effect by age group (difference in mean RTs between Prime beginning and Prime end sentences). Values below 0 denote mean RTs that were longer in the Prime end condition (left panel: younger adults, right panel: older adults). Box plot height denotes interquartile range; vertical lines below plots denote 25th percentile, while lines above denote 75th percentile. Black horizontal lines denote the median, triangles denote the mean. Black dots represent outliers.

4.3.3 Motivation

Participants completed three motivation-related questionnaires. Firstly, we asked participants about different types of motivation for taking part in the study. We carried out a 2 x 7 mixed ANOVA to examine effects of Age group and the seven motivation questions participants were asked. Participants rated some motivations more highly than others (F_1(4.105, 320.211) = 57.471, p = <.001, η_p^2 = .424), as reported in Table 4.5. Taking part for credit/money, to help science, to learn about one's own responses, and out of curiosity scored high. There was no significant effect of age, suggesting that overall the two groups' reasons for taking part were

similar (F_1(1, 78) = 2.258, p = .137, η_p^2 = .028). There was however a significant interaction between motive and age group (F_1(4.105, 320.211) = 5.490, p <.001, η_p^2 = .066). This appeared to be driven by younger adults rating particular motives more highly (these included: being bored, for course credit/money) while they provided a lower rating to the motive 'to learn how I respond to cognitive tasks'.

Older adults reported having a higher level of overall motivation in the sentence completion task in the current study (scale 1-9; M = 8.38, SD = .71) than the younger adults (M = 7.55, SD = 1.24; t(61.84) = 3.660, p = <.001, d = .818). Overall motivation level and mean RTs in the sentence completion task were not significantly correlated with each other [r(80) = -.061, p = .589], suggesting that while older adults might have been more motivated, this did not influence how fast they responded.

When providing task ratings, older adults also reported feeling more positively about the sentence completion task overall (see table 4.6.). For example, they found the task significantly more fun, interesting and enjoyable, and less boring than the younger adults did (F_1(1, 78)= 6.421, p = .013, η_p^2 = .076). There was also a significant main effect of task rating, showing that across groups, some categories were rated more highly overall (the lowest ratings were for the category 'boring', whilst the highest ratings were for the category 'fun'; F_1(2.539,198.017)= 14.477, p = <.001, η_p^2 = .157). Age group and category did not interact however, (F_1(2.539,198.017)= 0.892, p = .432, η_p^2 = .011) suggesting that younger and older adults rated the categories in a similar way.

Overall, the data from the motivation questionnaire suggest that participants in both age groups had similar reasons for taking part in the study. Some of these motives were stronger in the younger adult group (as demonstrated by their higher ratings of those motives). The data also showed that participants in both groups were highly motivated during the sentence completion task, although comparatively, older adults were more motivated and rated the task more positively (i.e., enjoyable) than the younger adults.

Table 4.5 Younger and older participants' motives to take part in the study

Motive	**Age group**	
	Younger	*Older*
Help science	7.25 (2.04)	7.55 (1.99)
Learn how I respond to cognitive tasks	6.25 (2.77)	7.85 (2.25)
Compare myself to others	3.52 (2.55)	4.57 (2.76)
Bored	3.60 (2.59)	2.30 (2.44)
For fun	5.45 (2.36)	6.90 (2.15)
For course credit/money	8.23 (2.33)	7.20 (2.08)
Curiosity	7.00 (2.05)	7.65 (2.24)

Younger and older adults' mean ratings and standard deviations of different motives to take part in the current study. Participants rated each motive using a slider that ranged from 0 -10. 0= Strongly Disagree, 10=Strongly Agree.

Table 4.6 Sentence completion ratings

Task rating	**Age group**	
	Younger	*Older*
Fun	5.20 (1.34)	5.70 (1.45)
Boring	2.53 (1.36)	1.83 (1.22)
Interesting	4.65 (1.53)	5.53 (1.54)
Enjoyable	4.90 (1.36)	5.65 (1.39)

Younger and older adults' reports of how they felt about the sentence completion task. Participants rated on a scale of 1-7 how much they agreed that the task was fun, boring, interesting and enjoyable. 1= strongly disagree, 7= strongly agree. Category 'Boring' was reverse coded before statistical analysis.

4.4 Discussion

Previous research suggests that both younger and older adults utilise semantic information to aid the retrieval of upcoming related words within their own speech (Chapter 2). The aim of

this chapter was to explore how particular characteristics of the sentences produced by our conversational partners (i.e., the position of the aforementioned semantic cues, as well as sentence length) can affect the efficiency of word retrieval during our own speech. We also examined the potential role of working memory. Younger and older adults completed a task wherein they produced single words to complete verbally presented sentences. Surprisingly, processing longer sentences was not associated with slower word production. However, both younger and older adults appeared to benefit somewhat from encountering semantic primes earlier rather than later on within the preceding sentence.

4.4.1 Sentence length

We expected processing longer utterances during conversation to negatively affect subsequent word production, as longer sentences were expected to tax valuable cognitive resources. This expectation was rooted in previous research showing that in language comprehension tasks, use of complex sentences is associated with poorer performance accuracy in both younger and older participants (Liu & Wang, 2019). Furthermore, older but not younger adults' word production efficiency is slower for syntactically complex sentences in comparison to shorter/simpler constructions (Kynette & Kemper, 1986; Kemper, 1986; Feier & Gerstman, 1980). Our accuracy analysis indeed showed poorer accuracy after longer compared to shorter sentences, suggesting participants were better able to formulate an accurate response to shorter sentences. This did not differ between older and younger adults. However, in terms of response times (our main focus), word retrieval during word production was not negatively affected by the longer sentences in either age group. If anything, completing the shorter sentences took slightly longer than completing the longer sentences, and this was especially apparent in the older adults (although this was not significant). One explanation for this pattern of results could be that while shorter sentences more frequently elicited correct target responses, suggesting they were easier to complete, these shorter sentences did not offer any benefits in terms of the time needed to plan those target responses.

Accuracy effects furthermore showed a relationship with working memory. For longer sentences in particular, higher working memory scores were associated with more frequent target-word production. This could suggest participants with a higher working memory score were better able to keep the long sentence in mind and retrieve an appropriate response.

However, working memory did not relate to RTs and did not differ between age groups. Previous studies have focused on the comprehension of more complex sentences (Gibson, 1998, 2000) or on the production of syntactically complex utterances (e.g., Kynette & Kemper, 1986; Kemper, 1986; Feier & Gerstman, 1980), as opposed to the production of single words as measured in the current study. The production task used in our study may not have been demanding enough to elicit a relationship between RTs and working memory, or to show additional working memory benefits for older adults in terms of accuracy. Furthermore, the working memory scores were larger for older adults, but not significantly so. This could suggest the tasks were not demanding enough and/or the older adults tested in our study continued to experience a relatively high level of cognitive functioning. Future studies might consider the effect of complex sentence processing on word production by studying the production of utterances rather than single words, as that may provide more insight to the cost of speech processing on language production in a more naturalistic context.

4.4.2 Semantic prime position

With respect to the position of semantic primes, research with younger adults has shown that earlier placement of critical information within their conversational partners' speech is associated with earlier onset of utterance planning processes (Barthelet al., 2016; Bögels et al., 2018; Bögels et al., 2015; Corps & Pickering, 2024), which is displayed in faster word production times (Bögels et al., 2015). Faster word production following sentences when this information is presented earlier may be due to there being sufficient time to prime the target word and/or for the listener to form predictions about upcoming words. The current study supports previous findings with younger adults. Participants in both the younger and older age groups were somewhat faster to produce words following sentences wherein the prime occurred earlier. Furthermore, accuracy was higher when the prime was presented at the start. However, it should be noted that the RT effect only reached significance when items were included in the analysis (LME and F_2 analysis), suggesting there was considerable item variability to account for. Furthermore, the interaction between Sentence Length and semantic Prime Position did not reach statistical significance, although earlier priming should have been more influential in longer sentences due to the greater distance between the prime presentation and target production.

These findings, however, speak against the other proposed direction that later occurring semantic primes might be beneficial if speakers do not start planning their own speech until the end of their conversational partner's turn (e.g., Sjerps & Meyer, 2015). This suggests that speakers indeed start preparing their responses while still listening to the conversation partner, rather than waiting until the start of their own turn. Furthermore, the absence of age-group differences suggests that older adults too can benefit from primes occurring early in the preceding sentence context and that any facilitation occurring thanks to early primes does not decay too quickly for older adults to benefit.

It is furthermore worth pointing out that the age groups showed no differences in the sentence completion task despite the use of low-frequency words. Contrary to the other chapters, we here used low-frequency items as research suggests that lower frequency items are more challenging for older adults to retrieve. However, older and younger adults did not differ significantly, surprisingly including in terms of their overall RTs. One reason for the lack of an observed age effect could be because participants were required to produce these low-frequency targets following sentences containing words that primed them. As such, while the words themselves were of lower frequency, the semantic relationships with the prime words might have been sufficient to overcome any retrieval difficulties. In this respect, it is also important to note that in this study, we did not incorporate a neutral baseline condition (as this was not possible in this sentence-completion task). As there was no neutral baseline condition, it is not possible to compare the degree of facilitation leveraged by each age group from the semantic primes.

4.4.3 Motivation

Data from the motivation questionnaire showed that older adults reported higher motivation and enjoyment of the study. Furthermore, their reasons for taking part in the study differed in some respects from the younger adults. For example, younger adults provided the highest mean ratings for the motive 'for course credit or monetary compensation' while older adults provided the highest mean ratings for the motive 'to learn how I respond in cognitive tasks.' (even though no feedback was given during the task). These findings are in line with previous suggestions that younger adults are motivated to take part in scientific research to obtain physical rewards (such as course credit or money); whereas older adults are more motivated to take part for reasons such as learning about their cognitive health (Ryan & Campbell, 2021).

Overall, both the younger and older adults reported enjoying the sentence-completion task but older adults reported higher task motivation and enjoyment. This is in line with previous studies showing that older adults reported greater levels of this type of motivation in comparison to younger adults (Frank et al., 2015). It is important to understand task-motivation since this has been linked to performance (with higher levels of task motivation being associated with better task performance; Brosowsky et al., 2023). In our study, a *higher* level of motivation in the older adults compared to the younger adults could be one explanation for the observed similarities between the two age groups with regards to their language production abilities. Older adults with intact cognitive abilities may be more strongly motivated to seek out research studies such as this one than those who have reduced cognitive functioning. This may explain the unexpected lack of a significant age group difference in production onset times in the current study, and the preserved use of semantically predictive contexts in old age (also observed in Chapter 2). However, motivation level did not correlate with overall naming RTs in the current study, suggesting that motivation level may not be a driving factor underlying language-production. Thus, even though task motivation levels differed between age groups, they were unlikely to explain any language-related (lack of) differences between age groups.

4.4.4 Limitations

In contrast to a picture naming task where the production of a specific target word is elicited, the paradigm used in the current study did not allow for such control. Instead, participants varied in their responses and we scored accuracy leniently with responses being counted as correct if they completed the sentence meaningfully and grammatically. This has implications for the aims of the current study as we were interested in investigating the production of low-frequency words, while participants could choose a higher frequency word to complete the sentence instead. Indeed, further investigation of this data showed that the alternative words selected by participants were on average of higher frequency than our expected target words. It is worth pointing out though, that the younger and older adult groups did not differ in terms of average frequency of the words they chose when they did not use the target, ruling out the option that older adults might have selected higher frequency words due to age-related difficulties in accessing and retrieving lower frequency words (Burke et al., 1991). Furthermore, accuracy and the type of alternative words provided did not differ between age

groups. Both older and younger adults typically used a semantically related alternative when they did not use the target word, suggesting older adults did not use compensatory mechanisms to avoid retrieval of difficult target responses.

Another pressing limitation, which also relates to the paradigm we used in the current study, is that participants' production onset times were not recorded relative to the presentation of a target picture that they were asked to name (as was the case in Chapters 2 and 5). Instead, in the current study, production onset times were recorded relative to a fixation cross (participants were instructed to wait until the sentence recording had finished playing, and a fixation cross appeared on the screen). This was done to ensure they had heard the full sentence before speaking and because their response could not be recorded while the audio was still ongoing. Participants who consistently produced their answers too early were replaced; however, many trials still had to be removed from the study for participants who were included (i.e., on trials where participants who were included in the final sample started producing their response too early). Removing too many trials could reduce the study's power, making it more difficult to detect significant effects. Furthermore, even if they did not produce the word out loud too early, participants likely prepared their word production before the end of the sentence. This could mean our measure of onset times relative to a fixation cross at the end of the sentence was not always a reliable measure of word retrieval during production. It could potentially also explain why sentence length effects only emerged in the accuracy data but not the RT data. These findings deserve further exploration, also considering the types of measures used (i.e., accuracy) that might be more sensitive to the effects of the manipulations reported in this study.

4.5 Conclusions

The study contributes to a growing body of research concerning prediction mechanisms within conversational contexts. We elaborate on the findings from Chapter 2, showing that characteristics of matched (predictive) contexts, namely the position of semantic primes, and (to some extent) context length can impact subsequent word retrieval and production in both younger and older adults. Similar to previous chapters, no differences were observed between younger and older adults, suggesting older adults too benefited from early response planning in predictive sentence contexts.

Chapter 5: Mismatched sentence contexts: the role of target likeliness and semantic relationships during word production

Abstract

The previous chapters in this thesis have demonstrated that the sentence contexts within which younger and older speakers are required to produce words can significantly impact how efficiently they produce those words. Sentence contexts that are strongly matched with target words can ease word retrieval (Chapter 2). However, speakers sometimes need to produce contextually unexpected words. Competition between such words and contextually expected words might cause interference and slow down language production. Additionally, older adults' poorer semantic control (e.g., Hoffman, 2018) suggests older adults could experience greater difficulty managing competition between semantic representations during language production. In the current chapter, healthy younger (*n* = 41) and older adults (*n* = 41) named pictures to complete verbally presented sentences. Target words were either a) unexpected but possible endings of the sentence ("*What did the bride throw*?" Response: '**Jelly'**); b) impossible, semantically related endings ("*What did the bride throw*?" Response: '**Groom**'); or c) impossible, semantically unrelated endings ("*What did the bride throw*?" Response: '**Receptionist**'). Both younger and older adults were slower to name these three types of unexpected target words compared to a neutral baseline, showing producing contextually incongruent words creates interference. Focussing on the effect of semantic relatedness, this interference was greatest when the impossible target word was semantically related to the context, suggesting overlap in semantic features created additional competition. However, despite the observed age-related declines within a separate semantic control task, these language-production interference effects did not differ significantly between younger and older adults, suggesting potential age-related changes in semantic control did not influence language production in this task.

5.1 Introduction

During conversation, speakers continually predict upcoming words. Furthermore, Chapter 2 showed that the words that speakers predict/expect within a speech context are retrieved more quickly than unexpected words. However, within naturalistic speech, speakers cannot

always accurately predict upcoming words, as the words that are statistically most likely to occur within a specific context are oftentimes not the ones a speaker is required to produce (e.g., Luke and Christianson, 2016). For instance, the most likely completion of "*I got my finest china out when I knew you were coming, would you like me to pour you some*" is '**tea**'. However, speakers could need to produce an alternative, unexpected word like '**beer**' or '**custard**'. In these cases, interference is expected based on the ensuing competition between more strongly predicted words, and the word that speakers are required to produce (Kim et al., 2023). However, in Chapter 2 we did not find support for an interference cost stemming from mismatched contexts, in older nor younger adults.

As discussed there, and in more detail below, an important factor to consider (and potentially the reason why no Mismatch effect was observed in Chapter 2) is the degree of unpredictability or unexpectedness of a target word occurring in a particular context. In the current study we therefore worked with different levels of "unexpectedness" to examine whether interference effects are more likely to be observed (or larger) when target words are not only unlikely but also (near) impossible. This also allowed us to examine whether age-group differences depend on the degree of mismatch between sentence contexts and target words (i.e., whether older adults experience greater interference in comparison to younger adults based on the degree of mismatch). Another key question we address in this study is whether the semantic relationships between sentence contexts and target words influence interference, as well as investigating differences in how older and younger adults respond to those relationships.

To that end, participants named pictures which were a) unlikely but possible endings of presented sentences, (similar to the mismatched sentences used in Chapter 2); b) impossible endings, semantically related to the sentence context; or c) impossible endings, semantically unrelated to the sentence context.

5.1.1 Matched vs mismatched sentence contexts

Both younger and older adults' production and processing of words are sensitive to the relationship between those words and the preceding sentence context (as shown in Chapter 2 as well as Altmann & Kamide, 1999; Balota et al., 1985; Ehrlich & Rayner, 1981; Kamide et al., 2003; Koornneef & Van Berkum, 2006; Ness & Meltzer-Asscher, 2018; Rayner & Well, 1996; Shao & Rommers, 2020; Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben,

1985). Younger and older adults are faster to produce words that are matched with the context. This facilitation can occur through priming, whereby producing/processing semantic cues (e.g., '***priest***', '***read***') triggers spreading activation to connected nodes in the semantic system (e.g., to concepts such as '**bible**'). This pre-activation of associated concepts allows for them to be retrieved more efficiently if/when needed (MacKay, 1987). Another way in which context facilitates production is through sentential prediction. This refers to speakers utilising the combination of words occurring in a context to formulate predictions about which words are most probable/ likely to follow (Broderick et al., 2021).

However, one notable aspect of everyday conversation is that the word(s) a speaker needs to produce can be mismatched with the context (i.e., the word that is needed is not the word that is most strongly primed/predicted by the context). For instance, '**bouquet'** is the most likely completion of the sentence: "*To her bridesmaids, the bride threw the*…", but a multitude of potential (less likely) endings to this sentence exist (i.e., anything that can be thrown such as '**cake**', '**cat**', '**pencils**', etc). Even endings that might seem extremely unlikely/impossible (such as a bride throwing the groom) could occur in the flexible way language is often used to, for example, talk about hypothetical situations. The degree of "unexpectedness" can greatly vary, with examples such as "cake" being more likely to be thrown than a "groom".

5.1.2 Interference from mismatching sentence contexts

ERP studies

Evidence from comprehension studies has revealed that the degree of mismatch (i.e., degree of unexpectedness) between target words and preceding sentence contexts indeed can affect neural processing. Many studies have reported an inverse correlation between cloze probability (the probability of a word being used as a completion/continuation of a sentence) and the N400 amplitude (a negative wave form peaking around 400ms post stimulus onset, Kutas & Hillyard, 1984, Thornhill & van Petten, 2012). For example, Federmeier and Kutas (1999b) and Federmeier et al. (2002) found that during both reading and listening, N400 amplitudes were higher for unexpected sentence endings (e.g., "*they wanted to make the hotel look more like a tropical resort, so along the driveway, they planted rows* of" **pines/ tulips**") than they were for expected endings (i.e., '**palms**'). Furthermore, these changes in N400 amplitudes appear to be influenced by how likely the target word is. For example, in a

sentence-comprehension study, the smallest N400 amplitudes were, as expected, observed when participants were processing highly expected sentence continuations (e.g., "*Sam knew the rain was coming because he could smell it in the* **air**"; Thornhill & van Petten, 2012). N400 amplitudes were slightly greater for intermediately expected continuations that were near-synonyms of the expected target word (e.g., "*Sam knew the rain was coming because he could smell it in the* **wind**"). However, the largest amplitudes were observed for words that were dissimilar to the expected target word (e.g., "*Sam knew the rain was coming because he could smell it in the* **morning**"). This suggests that the degree of unexpectedness of a target word can indeed modulate how easily it is processed in a context. A limitation of this vein of studies, however, is that the (facilitatory or interfering) effects of matched and mismatched sentence-word relationships are difficult to tease apart since N400 amplitudes are directly compared between these experimental conditions. The use of a neutral baseline would allow for interference effects to be precisely quantified without including the facilitation gained from more expected, matched context-word relationships. Furthermore, these studies have focused on processing (through ERP data) and do not directly speak to any potential effects on word retrieval during language production.

Behavioural studies and the role of target expectedness

While N400 studies examining comprehension consistently report significant effects of mismatched sentence contexts, behavioural data do not always lend support to these findings. In our behavioural study that investigated word production (reported in Chapter 2) no interference effects (delays in RTs) were observed when participants produced unexpected words to answer highly predictive questions relative to when they produced those words following neutral questions. Corroborating these findings, in a language comprehension study employing a similar paradigm to Chapter 2, also including a neutral baseline, no interference effect in terms of reading times was observed in either younger or older adults (Haigh et al., 2022). One implication of this is that perhaps the sentence-target pairs used in these studies were not unexpected/mismatched enough to strain cognitive/semantic control processes to a degree that resulted in an interference cost. In line with this interpretation, the data reported in Chapter 3 suggested that neither semantic nor domain general control processes were significant contributors of the mismatch effect reported in Chapter 2.

Indeed, in addition to the methods used (word naming times versus EEG measures), results can also vary across studies due to the type of language materials used. Specifically, one factor that appears to play a role in behavioural studies too is how expected a target word is to occur within the preceding context. Behavioural studies that do report interference costs have often used highly incongruent stimuli (i.e., sentences with extremely unlikely/impossible endings). For example, in Yoon et al. (2015), both younger and older adults exhibited poorer performance in a plausibility judgement task for implausible constructions (i.e., "*because the ceiling light is on, the room is* **dark**") than plausible constructions (e.g., "*because the ceiling light is off, the room is* **dark**"). In Chapter 2 (no interference from mismatched stimuli), mismatched contexts were unlikely but possible question-answer pairs (e.g., "*what did the man read to his daughter before bed? – "***bible**"). These types of stimuli were chosen to ensure that the scenarios were surprising but still largely naturalistic. However, it is possible that they did not create sufficient interference to be detected through behavioural measures, especially when compared to a neutral baseline rather than directly to the matching sentence context. This interpretation highlighting the role of degree of unexpectedness was confirmed by exploratory analyses in Chapter 2 showing that RTs were slowest for targets that received the lowest likeliness ratings, both within the Mismatch condition and across conditions.

In a related study, and one of the few other studies examining production, Bannon et al. (2025) showed that participants' picture naming times were affected by the degree of mismatch between a presented sentence context and a target picture. In their study, participants were significantly slower to name pictures following mismatched sentences (e.g., "*the children all sang happy birthday before they cut the* **pie**", expected answer = '**cake**') than following neutral sentences (e.g., "*the artist painted a picture of a single* **apple**"). Crucially, they also included sentences with a larger mismatch with their final target word (e.g., "*the frontal lobe is an important part of the* **cake**"). As expected, this condition was associated with the largest cost compared to the neutral condition. The results of this study corroborate suggestions of a graded interference effect whereby interference within language production depends on how likely or expected a word is within the sentence context (with least likely/impossible words causing the greatest amount of interference). However, one pressing limitation of Bannon et al. (2025) is that participants were asked to press the 'space bar' when they were producing the word and then were asked to type the word afterwards. This could have allowed participants more flexibility to process and interpret words before pressing the

space bar and it leaves open the question how the actual onset of speech production is influenced by word likeliness.

5.1.3 Semantic relatedness

In addition to a word's likeliness/expectedness, its semantic relationship with the context and/or the predicted ending might also affect how easily it can be retrieved. For instance, a strong semantic relationship between unexpected target words and predicted words could have a facilitatory effect (i.e., faster retrieval of the required word compared to a target word not related to the predicted word; Federmeier, 2007, Federmeier & Kutas, 1999). Some evidence for this account is derived from EEG studies showing that N400 amplitudes are indeed smaller for these types of unexpected but semantically 'related' sentence completions than unexpected, unrelated completions, despite both endings being matched on cloze probability (e.g., Federmeier, 2007). For example, Federmeier and Kutas (1999a) measured younger adults' ERPs whilst they read sentences that were completed by a) the most expected ending (e.g., "*The day before the wedding, the kitchen was just covered with frosting. Annette's sister was responsible for making the* **cake**", b) an anomalous ending belonging to the same semantic category as the expected ending (e.g., '**cookies**'), or c) an anomalous ending unrelated to the expected ending (e.g., '**toast**')**.** N400s were smallest for the expected endings, followed by endings belonging to the same semantic category; and were largest for the endings belonging to a different semantic category. Similar findings to these were also reported in a number of other studies (including in Federmeier, 199b, Federmeier et al., 2002, Federmeier, 2007, and Federmeier & Kutas, 1999). In a related study, Broderick (2021) showed dissociable neural correlates for lexical surprisal and semantic dissimilarity. Based on predictive processing at the level of lexical representations, neither "**salt**" nor "**socks**" would be an expected ending of "I take my coffee with cream and …". However, based on predictive processing at the level of semantic representations, "**salt**" (which shares more semantic features with the expected "**sugar**") would be a more expected ending than "socks". Broderick et al's N400 effects showed that younger adults use both lexical and semantic predictive processing, including pre-activating semantic features of upcoming words in speech. This could be expected to facilitate processing of unexpected but semantically related words at the behavioural level too.

Alternatively, a strong semantic relationship between expected and required words could also be argued to have a detrimental impact on language production processes. For example, according to models of word production (e.g., Bloem & La Heij, 2003; Bloem, et al., 2004; La Heij et al., 2006; Levelt, 1992; Levelt et al., 1999; Roelofs, 2018), the co-activation of associated semantic representations creates a competitive environment wherein target words must compete with related words for selection. This in turn could cause delays in target retrieval (cf. Damian & Bowers, 2003; Dell'Acqua et al., 2010; La Heij et al., 2003; Melinger & Abdel Rahman, 2013; Roelofs, 1992; Schriefers et al., 1990; Vieth et al., 2015). An interfering effect of semantic relatedness is most evident in studies employing the picture-word interference paradigm. In these studies, in the presence of semantically related distractor words, naming targets are produced more slowly (e.g., target: '**ball**', distractor: '**frisbee**') than when distractors are not semantic relatives of the target (e.g., '**hammer**'; Taylor & Burke, 2002 but also see Glaser and Düngelhoff, 1984, Hantsch et al., 2009; La Heij, 1988 & Vermeij, 1987; Lupker, 1979; Schriefers et al., 1990). Furthermore, target words with more, strong semantic neighbours are retrieved more slowly and less accurately in picture naming studies than targets that have fewer strong semantic neighbours (e.g., Fieder et al., 2019; Mirman, 2011), but cf. Hameau et al., 2019).

Thus, a stronger semantic relationship between an unexpected word and the expected target could be argued to facilitate production in context due to the shared semantic features or could be argued to interfere with production due to more similar expected target words increasing competition further. In Chapter 2, we did not control for the target words' semantic relatedness to the predicted word or to the sentence context. We used both question-answer pairs where the target was unrelated (e.g., "*what did the alpinist climb*?" '**ladder**', unrelated to expected answer '**mountain'**) and related ("*what did she eat on her birthday?* '**bread**', related to expected answer '**cake**'). Considering the role semantic relationships could play, in addition to examining the role of likeliness/unexpectedness, we therefore also examined the role of semantic relationships between sentences and unexpected target words.

5.1.4 Age-related changes in cognitive control and language processing

Turning to effects of age, older adults might experience greater interference costs from producing mismatched information during verbal utterances. As discussed in previous chapters, this could be related to weaker inhibitory or semantic control in older adults (but cf.

Chapter 3). The inhibition deficit hypothesis proposes that ageing is characterised by weakened inhibitory control; a skill that might be central in conversations to suppress unwanted/irrelevant information (Hasher & Zacks, 1988). In line with this, older adults have been found to perform more poorly on measures of cognitive control, such as the Stroop task (for example, see Chapter 3). Semantic control has also been found to decline with age. For example, in Hoffman (2018), older adults performed more poorly than younger adults on a feature association task wherein participants had to suppress distractors semantically related to a probe. In line with these findings, if required to produce unexpected/unlikely words that create larger interference, older adults would be expected to be more strongly affected than younger adults. In Chapters 2 and 3, we observed some evidence for older adults showing poorer inhibitory and semantic control, but no age-group differences in terms of interference costs during language production. Crucially, as discussed above, this was in the absence of a mismatch effect across age groups in the first place. As the first and second research questions in the current chapter, we therefore manipulated the strength of mismatching/interfering information to examine a) whether interference effects can be observed during production, and b) whether such effects depend on the degree of target unexpectedness. We specifically aimed to examine whether any age-group differences were more likely to arise in contexts (i.e., those with highly unlikely target endings) that elicit most interference.

Similar to younger adults exhibiting graded ERP responses based on a target word's expectedness within the context it appears in (e.g., Federmeier & Kutas, 1999b, and Federmeier et al., 2002), the N400 responses of older adults can exhibit these graded effects based on target expectancy too. For example, Federmeier et al. (2002) found that older adults too exhibited smaller N400s when listening to sentences paired with their expected endings (e.g., "*They wanted to make the hotel look more like a tropical resort. So, along the driveway they planted rows of*" '**palms**') than when they were paired with less expected endings, (e.g., '**pines**'/'**tulips**'). However, this effect was smaller and somewhat delayed in comparison to the younger adult group. Furthermore, both age groups exhibited smaller N400 responses for more likely within-category violations (i.e., '**pines'**) than for less likely between-category violations (e.g., '**tulips'**). Interestingly though, this facilitatory effect for more likely (but unexpected) words was comparatively smaller in the older adult group, and was only observed when sentential predictability/ constraint was lower (making the semantically related word more expected/likely). Arguably, these age-group differences could be due to

the likeliness of the final word and/or the within- versus-between semantic category violations.

As a third research question, we therefore also examined the role of semantic relatedness between words. As discussed above, semantic relatedness could either facilitate production of unexpected words due to shared semantic features with the expected target priming those unexpected words, or could cause further disruption due to increased semantic competition between the expected word and unexpected target. Both priming and semantic competition have been argued to change with age. As already discussed above, older adults have shown decreases in inhibitory and/or semantic control when managing competition. For instance, performance on semantic picture-word interference tasks is often poorer in older adults than in younger adults (e.g., Taylor and Burke, 2002), suggesting that older adults have more difficulty in inhibiting semantic distractors during cognitive tasks (but cf. Lorenz et al., 2018).

Semantic priming effects are usually preserved (Balota et al., 1999; Faust et al., 2004), if not exacerbated (see Laver & Burke, 1993 for a review), in old age. These findings are supported by Federmeier et al. (2003) who investigated the neural basis of semantic priming effects within sentential contexts. They found that the neural correlate of semantic association was robust in the older adult group. This strongly suggests older adults can continue to use semantic information when processing sentences and producing words, as also shown in the Match effect in Chapter 2. However, other studies have suggested that older adults might not use semantic similarity as much as younger adults, possibly because of lower semantic pre-activation and prediction of upcoming words. For example, in Broderick et al. (2021), the neural effect of 'semantic dissimilarity' in older adults was significantly reduced in comparison to the younger adult group. Both younger and older adults showed predictive processes at the level of lexical representations. Younger adults also processed anomalous words that were semantically related to expected/predicted words (e.g., "*I take my coffee with cream and* **salt**", semantically related to the expected word '**sugar**') differently than anomalous words that were unrelated to the expected word (e.g., "*I take my coffee with cream and* **socks**"). In contrast, older adults exhibited a significantly smaller 'semantic (dis)similarity effect'. This suggests that unlike younger adults, older adults do not use sentence context to prime the semantic features of expected upcoming words. This is supported by Cameli and Philips (2000) showing that while older adults benefited from

semantic priming in a sentence context, several of the semantic-priming related comparisons showed either no effect or a smaller effect in older compared to younger adults.

Thus, based on the current literature, an effect of semantic relatedness between unexpected endings and the expected target/sentence would be expected. Such effect could be facilitatory or interfering and is expected to be modulated by age, with older adults potentially predicting and pre-activating semantic relationships with upcoming words less and/or having greater difficulty managing interference between unexpected words semantically competing with expected targets.

5.1.5 Current study

In Chapter 2, we did not observe an interference effect in word production. Further analyses and the literature suggest interference effects are more likely to occur when target words are more unexpected (e.g. Bannon et al., 2025). To that end, in the current study we again asked participants to name pictures after listening to incomplete sentences. The naming targets belonged to one of four conditions A) possible but unlikely endings of the preceding sentence context; B) impossible endings semantically related to the sentence context; C) impossible endings semantically unrelated to the sentence context; and D) one of many possible (neither strongly likely nor unlikely) endings following a neutral context (neutral baseline condition) (see Table 5.2 for examples).

In this chapter we addressed three research questions. The first question addressed whether the mismatch conditions outlined above create interference in younger and older adults' language production (i.e., longer target naming latencies relative to neutral contexts). While it was unclear if naming in Condition A (Unlikely but possible endings of the preceding sentence) would elicit an interference cost, owing to the existing mixed findings in this area (e.g., Bannon et al., 2025; Haigh et al., 2022; Chapter 2), it was expected that the two 'impossible' conditions would be associated with a (larger) interference cost (based on findings such as those of Bannon et al., 2025 and Yoon et al., 2015) . The closely related second research question of this chapter addressed the role of 'target expectedness' on interference. Specifically, due to the reasons stated above, we expected words in Condition C (impossible, semantically unrelated endings) to be retrieved more slowly than words in Condition A (unlikely possible endings of the context, which were not semantically related to the sentence context or to the expected ending). We chose to compare these two conditions to investigate

the effect of 'target expectedness' whilst controlling for the potential (facilitatory or inhibitory) effects of semantic relatedness on production. With regards to the effect of age, if older adults experience greater difficulty in suppressing interfering information encountered during contextual processing than younger adults do (e.g., Hoffman, 2018), we expected that they may be more prone to exhibiting 'target expectancy' effects (i.e., stronger interference within Condition C than Condition A).

Our final research question addressed whether 'semantic relatedness' facilitates or hinders word production processes. As mentioned previously, we expected the strongest interference effects within the most strongly mismatched (impossible) stimuli conditions (Conditions B and C). However, these two conditions differed in how semantically *related* or *unrelated* the target ending was to the sentence context. With regards to this research question, two potential outcomes were hypothesised. On the one hand, a strong semantic relationship between the sentence context and target ending could facilitate the target word's retrieval (due to the context priming the target word), leading to less interference (and thus faster naming times) in the semantically related than in the semantically unrelated contexts (cf. also Federmeier 1999a, 1999b; Broderick et al., 2021). Although this direction has been observed most commonly in ERP studies, alternatively and as pre-registered, a direct semantic association between words in the sentence context and a highly unlikely target word may also cause greater *interference* (therefore longer naming times) in the semantically related compared to unrelated condition. This direction would be expected if a semantic relationship between the sentence context and unexpected target words creates competition (e.g., Levelt, 1992; Roelofs, 2018).

Moving on to the effect of age, regardless of the direction of semantic related versus unrelated sentences, we expected such effects to differ between younger and older adults. On the one hand, if older adults experience stronger semantic priming than younger adults (Balota et al., 1999; Faust et al., 2004; Laver & Burke, 1993), we expected them to either show a larger facilitating effect of semantically related context or a larger interference effect of semantically related context. If semantic relatedness facilitates retrieval of unexpected target words (in the former case), older adults might benefit more from semantic priming. Alternatively, if semantic relatedness increases competition (in the latter case), older adults might experience greater competition and might find it more difficult to manage this interference due to diminished inhibitory and/or semantic control. Alternatively, older adults

might not be as strongly affected by semantic relationships, particularly if they do not utilise semantic context in the same way younger adults do (e.g., Broderick et al., 2021) to prime upcoming words and their associates. This would result in either smaller facilitation or smaller interference of related compared to unrelated semantic stimuli.

Finally, to address these questions further, we also examined participants' semantic control (difference in RT performance on the congruent and incongruent trials on the feature association task, Hoffman, 2018). We examined potential relationships between semantic control and the 'target expectedness effect' in the production task (Research Q2: the difference between production RTs on the possible (unlikely) and impossible (unrelated) trials); as well as with the 'semantic relatedness effect' (Research Q3: the difference between production RTs on the impossible (related) and impossible (unrelated) trials). This allowed us to examine the potential role of semantic control in response to language production varying in degree of likeliness and semantic relationships.

5.2 Methods

This study was pre-registered on the Open Science Framework: https://osf.io/7e9h3/ The data are stored here https://osf.io/7e9h3/?view_only=b2dfe719e78d45e88843df7f917efec2.

5.2.1 Participants

Ethical approval was obtained from the Department of Psychology at the University of York and participants provided informed consent at the start of the study. The final sample included 82 native English-speaking monolinguals. Forty-one older adults (aged 65-76, *M* = 68.8, *SD* = 3.5) were recruited through prolific.co (*n* = 34) and through our departmental database (*n* = 7). Forty-one younger adults (aged 18-30, *M* = 20.1, *SD* = 2.7) were recruited through SONA (the university's internal participant recruitment system; *n* = 36) and Prolific (*n* = 5). Participants received either monetary compensation, Amazon vouchers, or course credit for their participation. The two age groups were matched on the number of years of formal education they had undergone (see Table 5.1). The percentage of participants who had completed a bachelor's degree was higher for the older adult group (46.3%) in comparison to the younger adult group (12.2%). This between-groups difference is owed to the fact that the younger adults were undergraduate students who were still completing their degree. This, taken with the fact that the two groups *were* matched on the mean number of years of formal

education they had undergone at the time of completing the study, provides assurance that the two groups were indeed educated to a similar level. Gender distribution (the number of male and female participants) was not equal in either age group, with a larger number of female participants making up the full sample population. This was especially true for the younger adults (See Table 5.1). Although efforts were made to recruit more male participants (i.e., through using the relevant filters during recruitment), we still did not receive many sign-ups from this demographic group. This might in part reflect differences in males' and females' interest in psychology/psychological research. This is especially true of the younger population wherein there is an underrepresentation of male students enrolled on A-level and undergraduate Psychology courses (Barrow et al., 2016).

Several additional participants completed the study but were replaced to meet our sample size aim. We removed two younger participants from the analysis as we could not ascertain that they met our full inclusion criteria (due to them providing incomplete information when completing the background questionnaire). One additional younger participant was not included as their recordings were not audible. Three older participants were removed as they completed an incorrect version of the experiment and three additional older adults were not included as their recordings were empty or not audible. Finally, one older participant was not included as they scored less than 70% accuracy on the naming task and two were not included because they provided incomplete information when completing the background questionnaire.

Our final sample size of 82 participants (after exclusion) was based on a GPower analysis. We conducted a power analysis using a medium effect size (f = 0.3) for an interaction between Age group and Condition. This suggested our sample size yielded over 95% power to detect a medium-sized effect.

All included participants furthermore confirmed meeting the following eligibility requirements: they did not use a hearing aid, had (corrected-to-) normal vision, had not been using medication affecting their concentration in the past three months, did not have a language/reading disability, and had not been diagnosed with a neurodegenerative disease or cognitive impairment. Given that the study was conducted online, we were not able to use an assessment of cognitive functioning such as the ACE-III. In addition to asking participants to confirm each eligibility point, where possible, we also used existing screening criteria to only invite participants without a history of head injury, cognitive impairment, or dementia.

Table 5.1 Participant details

Participants	N	Age (Years)	Gender distribution		Formal Education	Higher education (%)
			Males	Females		
Younger	41	20.1 (2.7)	5	36	14.9 (2.1)	12.2
Older	41	68.8 (3.5)	12	29	14.4 (3.1)	46.3*

Demographic details of all participants who took part. Age (mean number of years and standard deviation), Gender distribution (total number of males and females in each age group), Formal education (mean number of years and standard deviation), Higher education (percentage of participants in each age group who had completed at least an undergraduate degree before taking part in the study). *Note that completed higher education was lower in the younger adults as they were still in the process of completing their UG.

5.2.2 Design

Participants completed a picture naming task with the within-participant independent variable Condition. This had four levels (see Table 5.2 for examples): Unlikely, Impossible (semantically related to the sentence context), Impossible (semantically unrelated to the sentence context), Neutral. Age group was a between-subject variable. The dependent variable was picture naming times (ms), defined as onset of naming relative to picture presentation.

Table 5.2 Example stimuli

Condition	Example
A. Unlikely	*The patient was rushed to hospital in a* **boat**
B. Impossible, Semantically Related	*The patient was rushed to hospital in a* **nurse**
C. Impossible, Semantically Unrelated	*The patient was rushed to hospital in a* **wheel**
D. Neutral	*During the contest, he lost his* **briefcase**

The four types of experimental sentences used in the study, including example sentences. The incomplete sentence always preceded the presentation of a target picture, which participants had to name.

5.2.3 Materials

All target pictures for the naming task were presented in greyscale and were sourced from the Multipic database (Dunabeitia et al., 2018) or from Google images. Pictures were preceded by a spoken sentence missing its final word. We opted for this type of sentence completion paradigm (similar to what we used in Chapter 4, but now with target pictures), as opposed to the question-answer task we used in Chapter 2 since this was a little more naturalistic, perhaps better capturing the processes involved when we finish sentences. Each sentence was recorded by a female English speaker, reflecting natural speech as much as possible. They were pre-processed using Praat (Boersma & Weenink, 2022) to add 50ms to the beginning and end of each recording and to scale all recordings to 60dB. Background noise was also reduced using Audacity® version 3.0.0.

As a starting point, we created 90 sentences wherein the final word was highly predictable from the sentence context (e.g., "*the patient was rushed to hospital in an* **ambulance**"). For the experimental stimuli, we paired the sentence context with three different endings (production targets) that each corresponded to one of our (mismatching) experimental conditions (See Table 5.2). We measured semantic overlap between target words in each condition and sentence context (i.e., content words in the sentence) as well as the overlap between target words and the expected/ predicted endings. This relationship was computed using word2vec, which provides a numerical rating of the semantic overlap between words (with higher ratings indicating a strong relationship between two words, whilst lower ratings indicate a weaker relationship between them). The target words used in

condition A) were viable but unlikely endings of the sentences and were not strong semantic associates of (words in) the sentence context, or of the predicted/expected ending of the sentence (i.e., target word to semantic cues in the preceding sentence: word2vec Mean rating = 0.12; target word to predicted/ expected ending of the sentence: word2vec Mean rating = 0.16). Condition B comprised words that were an impossible ending of the preceding sentence but were semantically associated to words in the context (word2vec Mean rating = 0.35) and to the expected ending (word2vec Mean rating = 0.22). Finally, condition c) was also composed of words that were an impossible ending of the context, but were *not* strongly semantically related to the sentence context (word2vec Mean rating = 0.10), or expected ending (word2vec Mean rating = 0.14).

To address the effect of 'target expectedness' (Research question 2), we wanted to compare conditions A and C, without semantic overlap with context or ending acting as a confound. We therefore ensured the word2vec ratings were not significantly different between these two conditions. This was confirmed by a t-test for the word2vec ratings between target and context ($p = 0.08$) and targets and expected endings ($p = 0.20$). In contrast, we ensured that these ratings *were* significantly different between conditions B and C to allow for an analysis of the 'semantic relatedness' effect (Research question 3) (*p*s for target-context and target-expected ending both <.001).

Additionally, we created a set of 30 neutral sentences that did not strongly predict one single ending, and these were paired with one of many possible endings (semantic overlap with sentence context word2vec = 0.07). These were included as a baseline against which we measured the degree of interference caused by each mismatched condition. Finally, we created a set of 30 matched sentences (that predicted the target ending) to use as filler sentences in the naming study. These were included to make sure participants' predictions were still correct on a set of the trials.

It is important to highlight that in Chapter 2 participants named the same target words within every one of the experimental conditions (i.e., participants named each picture four times, once each following a 'matched', 'mismatched', and 'neutral' question and once in the 'no context' trials). This design was possible in that study since the *questions* differed in each trial (to capture the effect of *context* on word production), and consequently there was no risk of participants predicting/guessing the target based on what target they named following a previous question. However, in the current study, the *target words*' relationship to the

preceding context was manipulated and thus the contexts (sentences) remained the same whilst the *endings* differed based on their relationship to the sentence. Since repeating the questions within participants would lead to participants predicting previously used words, we split the stimuli into three separate lists. This meant all participants heard each context only once, with the target word following that sentence counterbalanced across participants.

To control for the potential confounding effects of target word properties, we matched these on the following characteristics: word frequency, imageability, and length (number of syllables and phonemes) between Conditions A and C for Research Q2, and Conditions B and C for Research Q3. We also matched the targets from each experimental condition to the neutral targets, as well as matching the subject and verb frequency in the experimental sentences to the neutral sentences. The full list of stimuli, with further details about the target and sentence characteristics and matching, is provided in the Appendices.

To make sure the stimuli functioned as intended (i.e., the pictures we had selected accurately represented the target words, and the target words were more likely/expected in Condition A (Unlikely) constructions than in either of the two 'impossible' constructions (Conditions B and C)), we ran a pilot study. The pilot study was completed by ten younger (*M* Age = 18.7 years, range = 18-20) and ten older (*M* Age = 71.8 years, range = 65-79) adults. The first part of this study was a written-picture naming task to make sure all pictures could be recognised and named easily. We replaced pictures where this was not the case. Next, participants completed a likeliness ratings task. The purpose of this was to ensure that 'target expectedness' was highest for Matched (filler) sentences followed by Condition D (Neutral), then Condition A (Unlikely) and were lowest for the two sets of impossible sentences (Conditions B and C). Participants viewed each sentence-target pair and were asked to rate on a scale of 1-7 how *likely* the presented target was to follow the preceding sentence. Table 5.3 summarises the likeliness ratings for each condition by age group. If the mean rating was too high/ low in one or both age groups (based on which condition it belonged to), changes were implemented (i.e., the target word was swapped for a more suitable one). As a large number of changes were made to the stimuli based on this pilot study, the statistical analyses run on those pilot data are not reported here. As intended, however, the likeliness ratings were higher in the unlikely than impossible conditions, with no significant difference between the two impossible conditions.

The main study also included a likeliness task on the final stimuli. This confirmed that target words were indeed more expected/likely in unlikely contexts than in impossible contexts (see "Results" for an analysis of these ratings).

Table 5.3 Pilot 1 Likeliness ratings

		Younger adults	Older adults
Condition	**Likeliness rating (1-7)**		
A	Unlikely	3.04 (0.79)	2.13 (0.64)
B	Impossible (semantically related)	1.77 (0.45)	1.17 (0.24)
C	Impossible (semantically unrelated)	1.58 (0.43)	1.21 (0.27)
D	Neutral	4.91 (0.65	4.31 (0.57)
	Matched	6.76 (0.21)	6.73 (0.21)

Pilot 1 likeliness data: mean likeliness ratings and (standard deviations) for the question-answer pairs obtained from younger and older adults.

5.2.4 Procedure

The experiment was conducted using Gorilla.sc (Anwyl-Irvine et al., 2020). Participants first read the information sheet and provided informed consent. They then completed a background questionnaire (see "Participants"). Next, they completed a sound check to ensure that they could record audio files through their browser and to adjust their device's volume so that they were able to clearly hear the sentences. For the naming task, participants named pictures to complete sentences they heard. Participants were allocated to one of three stimulus lists. Each list comprised the same neutral (n = 30) and matched (n = 30) constructions. Additionally, in each list, there was a set of trials corresponding to each of the three experimental conditions (A: Unlikely, n = 30; B) Impossible, semantically related, n = 30; C) Impossible, semantically unrelated, n = 30). The presentation order of the stimuli was pseudo-randomised so that there were no more than three consecutive trials of the same type of sentence. While participants all heard the same sentences, in the experimental trials, they named a different target picture to complete the sentence (based on the condition which that sentence belonged to in that list). This ensured participants only heard each individual

sentence once, with the type of experimental condition each sentence belonged to counterbalanced across participants. Lists were equally distributed similarly across the younger and older adult groups.

Participants first completed a picture familiarisation task in which they saw the target pictures and words, asking them to read the word aloud and use it during the task. This phase was included to make sure all participants recognised the pictures when naming them in the study. In the naming task, participants were instructed to name the pictures as quickly and accurately as possible. Participants first saw three practice trials. On all trials, participants first viewed a fixation cross (500ms) followed by a blank screen while they heard a pre-recorded question. This was followed by another fixation cross (presented for 500ms). Next, the target picture was presented. The picture remained on screen for 2500ms, regardless of when a response was given.

Next, participants completed a task measuring semantic control (feature association task adapted from Hoffman, 2018, also used in Chapter 2). Participants selected the word matched with a probe word on a particular feature (e.g., colour or size, whilst ignoring distractor words). Half of the trials required participants to select the word most closely related in size, and the other half to select the word most closely related in colour to the probe. Half were congruent trials (target and probe shared a semantic relationship) while the other half were incongruent trials. In these trials, the target and probe were not semantically related but one of the distractors was related to the probe.

Finally, participants completed a likeliness-rating task in which they rated on a scale of 1-7 how likely the target word was to complete each sentence, for all sentence constructions they viewed during the naming study. The experiment lasted approximately 40 minutes in total.

5.2.5 Data analysis

Likeliness ratings

Likeliness ratings were examined using a 2x4 ANOVA with Age (younger, older) as a between-subject variable and Condition (A- Unlikely, B-Impossible, related, C-Impossible, unrelated, D-Neutral) as a within-subject variable. Data from the matched condition trials were not included in any of the analyses since they were only incorporated in the study as 'filler'

sentences, so that participants did not start expecting to produce only words that were unlikely/impossible completions of the preceding sentence.

Picture naming

An accurate response in the picture naming task was either the intended target word or a word that was fitting to the picture as well as the condition. For example, we accepted '**cup**' where the target was '**saucers**' (in the Impossible, semantically related condition) to complete the sentence "*she sweetened the tea with the*..." This was since the picture did include cups (an arrow was pointing to the saucers to highlight the naming target); and because '**cup**' is an unviable ending of this sentence whilst still being semantically related to the context (i.e., '**tea**')**.**

Other or no responses were scored as an inaccurate response. Picture naming accuracy was >82% for all participants (*M* older adults 91.28%, *SD* = 2.59; *M* younger adults 93.85%, *SD* = 3.18). As pre-registered, because accuracy was below ceiling (<95%), we ran some further accuracy analyses as well (see below). Naming RTs were determined using Checkvocal (Protopapas, 2007). RTs <300ms or more than 2.5 *SD* above or below the mean per participant and per condition were removed, using the trimr package (Grange, 2015; removing 2.78% of correct responses). Shapiro-Wilk tests conducted using raw RTs confirmed the data were normally distributed (*p*s >.25). RTs were z-scored to account for age related slowing.

RTs were analysed using a 2x4 ANOVA (SPSS version 29.0.2.0 (20)) to determine whether there was a main effect of Age or Condition, or an interaction between the two. (Bonferroni corrected) pairwise comparisons were used to compare data from each experimental level with data in Condition D (the Neutral,baseline) condition, to establish the presence or absence of interference effects stemming from the three experimental manipulations (Research Q1). As pre-registered, pairwise comparisons (Bonferroni corrected) were also run to compare conditions for Research Q2 (effect of target expectedness/likeliness): we compared Unlikely and Impossible (unrelated) conditions (Conditions A and C); and Research Q3 (effect of semantic relatedness): we compared Impossible related and unrelated conditions (Conditions B and C).

We also computed Bayesian ANOVAs using JASP version 0.17.3 (JASP Team, 2022), which examined evidence for/against age-group differences on the two effects of interest.

We computed the difference in naming times between each of the experimental conditions and the neutral baseline per participant. Using these data, we computed three Bayesian ANOVAs in which we compared a model including an age effect (between-groups difference) to a null model (no age-group difference on the interference effect) for each condition. This allowed us to better understand if there were any age-group differences on any of the interference effects. We also conducted two additional Bayesian analyses for Research Questions 2 and 3. We report these results in the form of "BF_{01}", showing the evidence for the null hypotheses (no age group difference) over the alternative hypotheses (significant age group differences). Values below 1 indicate evidence for an age-group difference; values above 1 indicate evidence for no age-group difference.

Finally, we estimated the internal consistency of each of the interference effects using a permutation-based split-half approach (Parsons, 2020a) with 5000 random splits to check for within-subject variations in these effects.

Semantic control

For each participant, we computed a semantic control score (the difference in their mean RTs within the congruent and incongruent trials in the feature association task). Firstly, we computed an independent samples t-test to check for a between-groups difference in semantic control score. We then ran another analysis to check if semantic control scores were correlated with our effects of interest: '*target expectancy/ likeliness*' = difference between mean RTs in Unlikely and Impossible (unrelated) naming trials, and '*semantic relatedness*' = difference between mean RTs in Impossible related and unrelated trials.

Linear mixed effects analyses

Further exploratory analyses used linear mixed-effect analyses to examine the relationship between sentence-target likeliness ratings (as provided by participants after completing the study) and naming RTs. These mixed-effects analyses also allowed us to include both participants and items in the analysis. Likeliness ratings for all three experimental conditions of interest (Unlikely, Impossible- related, and Impossible- unrelated) were treated as a continuous variable (without the inclusion of 'condition'). The two-level categorical variable age group was contrast-coded (younger adults = -0.5; older adults = 0.5). These analyses were conducted using [4.4.1; lme4 package version 1.1.35] with the maximal random-effect

structure including all within-participant and within-item slopes (following Barr et al., 2013). The analyses did converge, so removal of slopes explaining the lowest amount of variance was not required. We carried out this analysis twice; once with the untransformed RTs and then with the z-scored data.

Since participants' accuracy scores were not at ceiling, as per our pre-registration, we analysed these data using generalised linear mixed effect analyses. We conducted two analyses; in the first one, we compared accuracy scores within each of the three experimental conditions (Unlikely, Impossible-related, and Impossible unrelated) to the Neutral baseline condition (simple coded with Neutral as the baseline, using the grand mean as the intercept). In the second analysis, we compared accuracy scores in the Unlikely and Impossible-related conditions to the Impossible-unrelated condition (again using simple coding, but now with the impossible-unrelated condition as baseline). The two-level categorical variable age group was contrast-coded (younger adults = -0.5; older adults = 0.5). We began with a full model including all fixed and random effects. The final converging model included participant and word intercepts but no slopes.

5.3 Results

5.3.1 Likeliness ratings

Condition had a significant effect on likeliness ratings ($F_1(1.675, 134.037) = 1005.607$, $p < .001$, $\eta_p^2 = .926$; Greenhouse-Geisser outcomes are reported as Mauchly's Test of Sphericity was significant). As expected, participant likeliness ratings were highest for Condition D (Neutral sentence-ending pairs), followed by Condition A (Unlikely pairs). They were lowest for the impossible conditions (B and C, see Table 5.4). Pairwise comparisons showed significant differences between all condition combinations ($p < .001$). As can be seen in Table 5.4, the Impossible (semantically related) constructions were also rated as slightly more likely than the Impossible (semantically unrelated) constructions; however, importantly, ratings for both conditions were still very close to floor. There was no main effect of Age ($F_1(1,80) = 1.444$, $p = .233$, $\eta_p^2 = .018$) suggesting that overall ratings were similar for older and younger adults. However, there was a significant interaction between Age and Condition ($F_1(1.675, 134.037) = 3.298$, $p = .048$, $\eta_p^2 = .040$). While ratings of the Unlikely and Neutral constructions were similar for both age groups (unlikely: $p = .157$; neutral: $p = .416$), both types of Impossible

constructions were rated as slightly more likely by younger than older adults (related: p = .002, unrelated: p = .011, see Table 5.4). Crucially, however, each age group showed a significant difference in likeliness ratings between all four sentence types (Younger adults: $F_1(1.665, 66.595) = 453.359$, $p < .001$, $\eta_p^2 = .919$; Older adults: $F_1(1.648, 65.923) = 555.540$, $p < .001$, $\eta_p^2 = .933$, with all pairwise comparisons $p < .001$ in both age groups).

Table 5.4 Likeliness ratings

		Younger adults	**Older adults**
Condition	**Likeliness rating (1-7)**		
A	Unlikely	2.56 (0.62)	2.35 (0.67)
B	Impossible (semantically related)	1.58 (0.42)	1.34 (0.24)
C	Impossible (semantically unrelated)	1.33 (0.29)	1.19 (0.16)
D	Neutral	4.60 (0.84)	4.75 (0.83)
	Matched	6.83 (0.20)	6.77 (0.29)

Likeliness ratings obtained from participants completing the full study (note: the matched ratings are included in the table for completeness but were not included in the analysis as they were only included as filler items, and were not matched with the other items on target frequency, imageability, etc).

It is furthermore worth noting that in the current study, participants rated the Unlikely type sentence constructions as a little less 'likely' than they did in Chapter 2 (Chapter 2 mismatch likeliness ratings, older M = 2.89, SD = 0.53, younger M = 3.02, SD = 0.56). In contrast, the Neutral constructions were rated as a little less likely in Chapter 2 than here (Chapter 2: older M = 3.30, SD = 0.73, younger M = 3.63, SD = 0.64).

5.3.2 Picture naming

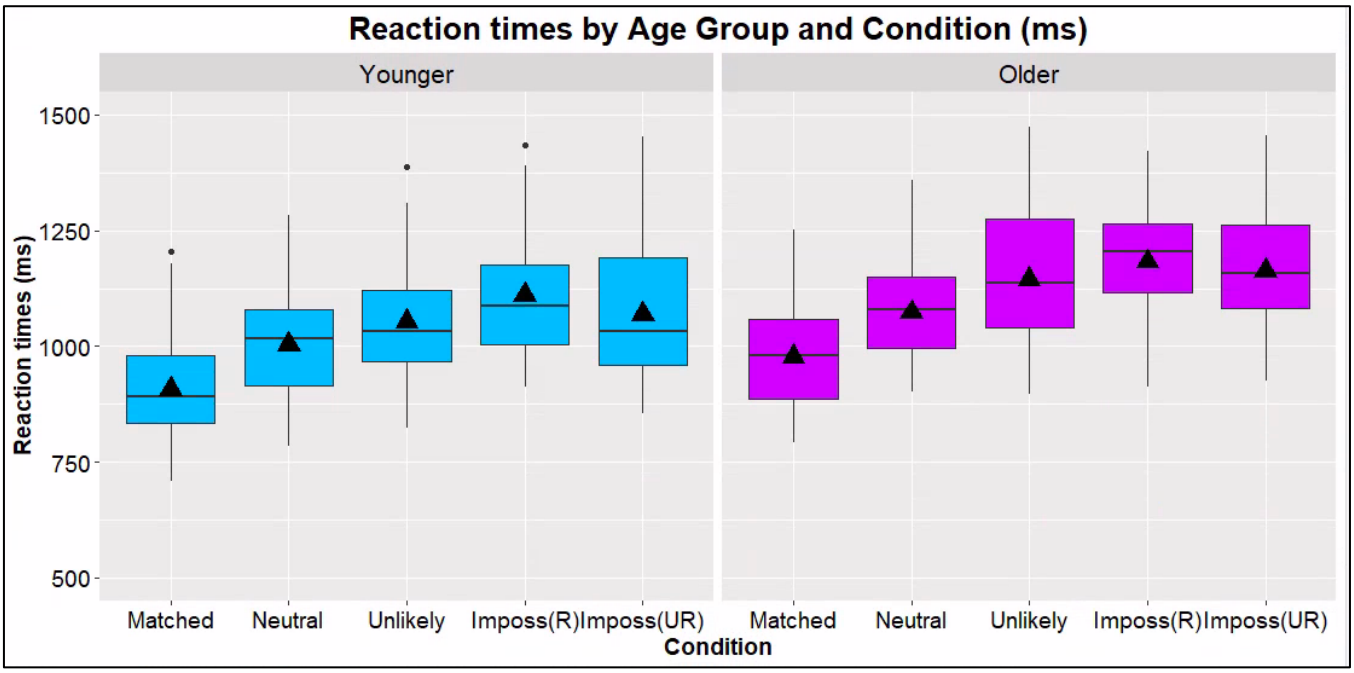

Figure 5.1 Box plots displaying untransformed mean RTs (ms) by age group (left panel: younger adults, right panel: older adults) and naming context. Box plot height denotes interquartile range; vertical lines below plots denote 25th percentile, while lines above denote 75th percentile. Black horizontal lines denote the median, triangles denote the mean. Black dots represent outliers. Imposs (UR) refers to the impossible unrelated condition while Imposs (R) refers to the impossible related condition.

Naming times

We started the picture naming analysis with the untransformed naming RTs (see Table 5.2). Figure 5.1 includes the Matched trials for completeness, but the analysis only included the three experimental mismatch conditions and the Neutral condition. As expected, there was a significant main effect of Age group on RTs, with older adults taking significantly longer to name pictures compared to younger adults (Younger M = 1056.84ms, SD = 125.65; Older M = 1143.78, SD = 104.34; $F_1(1, 80) = 10.081$, $p = .002$, $\eta_p^2 = .112$).

As pre-registered, and because of the age-group effect, we z-scored the data and then conducted the mixed ANOVA again. As intended, there was now no significant difference between RTs in the younger and older group ($F_1(1, 80) < .001$, $p = .988$, $\eta_p^2 < .001$). There was however, a significant main effect of Condition ($F_1(3, 240) = 38.310$, $p < .001$, $\eta_p^2 = .324$) There was no significant interaction between Condition and Age group ($F_1(3, 240) = .796$, $p = .497$, $\eta_p^2 = .010$).

Bonferroni-corrected pairwise comparisons showed significant differences (*ps*< .001) between the Neutral condition and all three mismatched conditions (A-C), confirming a significant mismatch effect in all conditions (Research Q1). In line with the absence of an interaction between Condition and Age group, Bayesian evidence supported that none of these mismatch effects differed between age groups (Mismatch effect Condition A, Unlikely: $BF_{01} = 2.740$, error = 0.016%; condition B, Impossible-related: $BF_{01} = 4.335$, error = 0.018%; condition C, impossible-unrelated: $BF_{01} = 2.360$, error = 0.015%; see Figure 5.2).

Table 5.5 Picture naming times and Interference effects

Condition	**Younger adults**	**Older adults**
Neutral	1004.57 (123.48)	1073.05 (101.89)
Unlikely	1054.04 (137.90)	1145.62 (146.18)
Impossible (related)	1111.44 (137.99)	1182.42 (121.86)
Impossible (unrelated)	1068.74 (144.26)	1163.04 (116.57)
Interference effects (relative to neutral)		
Unlikely	49.46 (80.10)	72.57 (100.42)
Impossible (related)	106.87 (98.55)	109.37 (102.85)
Impossible (unrelated)	64.17 (72.32)	89.99 (99.44)

Top: mean picture naming times (ms) (and standard deviations) per Age group and Condition, bottom: mean interference (ms) caused by each experimental condition. The interference effect is the RT difference relative to neutral baseline condition, with positive scores reflecting interference costs in the form of slower RTs compared to the Neutral condition (and standard deviations).

Our second research question focused on the effect of target expectedness/ likeliness, comparing conditions A (Unlikely) and C (Impossible, semantically unrelated). There was no significant difference between those two conditions (*p* = .788). Furthermore, this (absence

of a) difference was similar between the two age groups (as confirmed through a one-way ANOVA ($F_1(1,80) = .009$, $p = .923$, $\eta_p^2 < 0.001$, $BF_{01} = 4.332$, error = 0.018%).

Our third research question examined a potential influence of semantic relatedness with the impossible target endings (comparing Conditions B (Impossible, semantically related) and C (Impossible, semantically unrelated)). There was a significant difference between these two conditions ($p = .009$), with slower responses in the Impossible, related than Impossible, unrelated condition. However, again, this difference was similar between participants in the two age groups ($F_1(1,80) = 1.676$, $p = .199$, $\eta_p^2 = 0.021$, $BF_{01} = 2.101$, error = 0.015%).

Given the lower split-half internal consistency in Chapter 2 for the Mismatch effect, we also examined this here. The (Spearman-Brown corrected) split-half internal consistency of the interference effect stemming from Unlikely constructions was $r_{SB} = 0.35$, 95%CI [0.11, 0.55] indicating weak to at best moderate internal consistency of this effect within participants. Weak to moderate internal consistency was also observed for the interference effect stemming from Impossible, related constructions ($r_{SB} = 0.41$, 95%CI [0.2, 0.58]); as well as for the Impossible, unrelated constructions ($r_{SB} = 0.25$, 95%CI [-0.02, 0.47]).

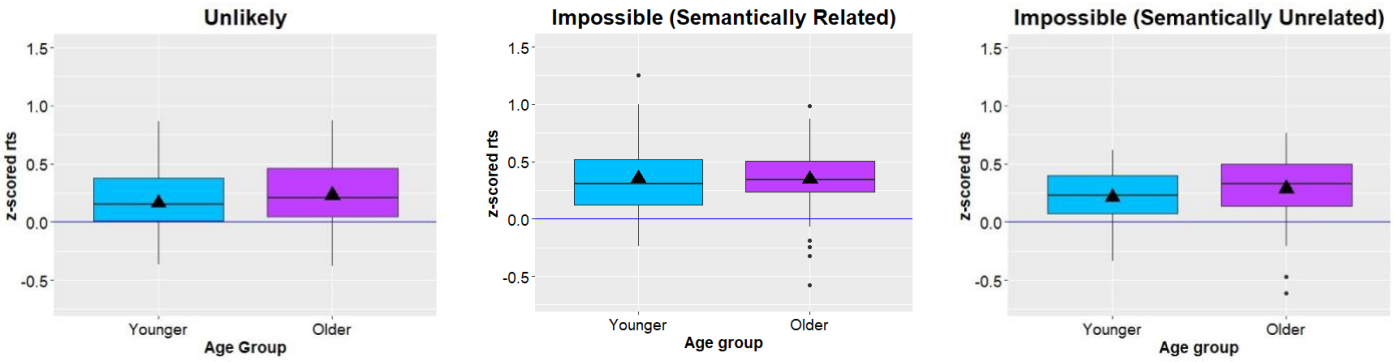

Figure 5.2 Box plots showing the interference effects associated with each of the experimental manipulations (longer RTs relative to the Neutral context). Scores above the zero line (blue line) indicate an interference effect, with slower responses to the mismatch than Neutral contexts. Box plot height denotes interquartile range, vertical lines below plots denote 25th percentile, while lines above denote 75th percentile. Black horizontal lines denote the median, triangles denote the mean. Black dots represent potential outliers.

Naming times and likeliness ratings

Similar to Chapter 2, we also ran a further analysis examining the potential role of individual items' likeliness ratings, as provided by each participant. Given that these used linear mixed-effects analyses, we first ran an analysis examining effects of Condition and Age group. Similar to the ANOVAs reported above, significant Mismatch effects were observed for the two impossible conditions (*p*s < .02). The Mismatch effect was close to significance for the unlikely condition (*p*s < .065). The difference between the related and unrelated conditions did not reach significance in this analysis (*p*s = .12). None of the effects interacted with Age group (*p*s > .10). Focusing next on the analyses with likeliness ratings, in contrast to Chapter 2, the linear mixed effect analysis showed that likeliness ratings were not significantly associated with raw RTs (β = -3.351, *SE* = 4.012, *t* = -0.835, *p* = .404) or z-scored RTs (β = -0.010, *SE* = 0.013, *t* = -0.811, *p* = .417). In line with the ANOVA analyses above not showing differences between the Unlikely and Impossible conditions (A and C), this suggests the size of the mismatch effect was not modulated by how (un)likely/expected a target word was. Similar to the ANOVAs, age group did not interact with the likeliness ratings in either of the models (both *p*s > 0.29).

Picture naming RTs and semantic control

With regards to semantic control, accuracy cost (difference in accuracy scores between congruent and incongruent trials) was numerically higher in the older adult group ((*M* = -13.618%, *SD* = 16.563) than in the younger adult group (*M* = -10.569%, *SD* = 11.067). However, this between-groups difference in accuracy cost was not statistically significant (*t*(80) = -.980, *p* = .330). The Semantic control RT cost (difference in RTs between congruent and incongruent trials) was significantly higher in the older adult group though (*M* = -1300.34ms, *SD* = 794.32) than in the younger adult group (*M* = -652.188ms, *SD* = 426.36; *t*(61.282) = -4.604, *p* = <.001) suggesting the older adults experienced greater difficulty with semantic control. However, this semantic-control score was not significantly correlated with the language-production scores of interest: Research Q2, the RT difference between the Unlikely and Impossible-unrelated conditions (*r*(82) = -.028, *p*=.806); nor with, Research Q3, the RT difference between the Impossible, semantically related and unrelated trials (*r*(82) = .076, *p*=.498).

Picture naming Accuracy

Finally, because accuracy was not at ceiling (in either age group), we ran some further analyses on the picture naming accuracy data. In the first analysis, we compared the three experimental conditions to the neutral baseline, to assess whether there was an accuracy mismatch effect. Accuracy was significantly poorer in the Impossible, semantically related condition (B) than in the Neutral condition (β = -1.010, *SE* = 0.376, *z* = -2.687, *p* = .007). In contrast, no accuracy mismatch effect was observed for conditions A (Unlikely; β = -0.303, *SE* = 0.386, *z* = -0.785, *p* = .433) and C (Impossible, semantically unrelated; β = -0.333, *SE* = 0.385, *z* = -0.867, *p* = .386).

There was a significant main effect of age, with older adults exhibiting poorer accuracy across all conditions in comparison to the younger adults (β = -0.768, *SE* = 0.191, *z* = -4.027, *p* <.001). However, there was no interaction between condition and age group (all *ps* > .320).

The second analysis compared the Impossible, semantically unrelated condition (C) to the Unlikely condition (A, Research Q2) and the Impossible but semantically related condition (B, Research Q3). Accuracy did not differ significantly between the Unlikely and impossible conditions (β = 0.030, *SE* = 0.295, *z* = 0.103, *p* = .918) but was significantly lower in the semantically related than unrelated condition (β =-0.677, *SE* = 0.283, *z* = -2.389, *p* = .017).

Table 5.6 Accuracy scores (picture naming)

Condition	Younger adults	Older adults
A Unlikely	95.93 (5.55)	92.76 (5.05)
B Impossible (related)	92.76 (7.18)	89.11 (6.71)
C Impossible (unrelated)	96.50 (4.47)	92.93 (5.59)
D Neutral	97.40 (3.84)	95.77 (3.87)

Mean proportion of correct answers (and standard deviations) produced by age group and condition

5.4 Discussion

In this chapter we examined mismatching sentence context-target word relationships and their effects on younger and older adults' language production. To that end, participants completed a sentence completion task, wherein they responded to verbally presented

contexts (that predicted a specific word) by naming pictures that were either A) Unlikely but possible, B) Impossible, semantically related, or C) Impossible, semantically unrelated completions of them. Participants also named pictures following a set of D) Neutral contexts to measure mismatch effects.

Our study had three main goals; the first of these was to explore if processing these different types of mismatching sentence contexts creates interference in younger and older adults' subsequent language production. Interference effects (slower naming times compared to the Neutral baseline condition) were observed within all three of the mismatching conditions, suggesting these types of contexts can indeed slow down subsequent language production. The second goal of our study was to examine the effect of *'target expectedness'* on language production (i.e., testing whether the mismatch effect depends on how unlikely/ unexpected a target word is). We did not observe a difference between the Unlikely and Impossible conditions. Finally, the third aim of our study was to examine the effect of *'semantic relatedness'* on language production. Responses were slower and less accurate when participants had to name an unexpected word that was semantically related (rather than unrelated) to the expected target word.

We did not observe significant age-group differences for any of the effects reported above.

5.4.1 Interference during mismatched target production and the effect of target expectedness (Research Questions 1 and 2)

Our study builds on the findings of Bannon et al. (2025) being one of the few other studies investigating the role of context-target mismatches within language *production*. Like their findings (but contrary to Chapter 2), we also found that both smaller and larger mismatches between target words and the preceding sentence can create interference during language production. This interference relative to a neutral baseline can be explained through the conflict that arises when the word that a speaker is required to produce is different to the one they predicted based on the sentence context. In such cases, the predicted word must be inhibited in favour of retrieving the required one. Importantly, in contrast to Bannon et al. (2025), our study provides evidence for these findings by directly examining language production, as opposed to relying on participants' button presses to gauge production times.

The fact that naming in condition A (Unlikely but possible completions of the preceding sentence) was also associated with interference costs (although this only reached significance in the ANOVA analysis) was somewhat surprising given that in Chapter 2, as well as in Haigh et al. (2022), these types of constructions did not elicit (noticeable) interference. The likeliness data collected in Chapter 2 and the current study suggest that this difference in findings could be due to the fact that the Unlikely, possible constructions used in the current study were a little *less* likely (more mismatched with the preceding contexts) while the neutral constructions were a little *more* likely compared to Chapter 2. Thus, in terms of likeliness ratings, the mismatch might have been larger in the current study than in Chapter 2. However, in the absence of a direct relationship between likeliness ratings and production RTs in the current study, this interpretation remains tentative. Another possible explanation for this discrepancy could be the lack of a semantic relationship between targets, the context and expected words within this condition in this chapter (since this overlap is argued to aid retrieval, rather than hinder it as may have been the case in Chapter 2, Mahon et al., 2007).

Bannon et al. (2025) furthermore found that target words in the most strongly mismatched contexts were associated with the largest interference effects. This is also supported by comprehension studies (e.g., Federmeier et al.,2002; Federmeier & Kutas, 1999b) showing that neural effects can be modulated by the target word's likeliness/unexpectedness. Our study did not find RT or accuracy differences between the unlikely and impossible conditions, when (as pre-registered) directly comparing the two conditions that were comparable in their semantic relationship with the expected target. Furthermore, additional analyses showed no relationship between production RTs and the target word's likeliness rating provided by the participant.

Crucially, in the current study we ensured that the comparison of target expectedness/ likeliness was not confounded with semantic relatedness. It is possible that in studies where semantic relationships (between required words and expected words/ sentence context) have not been controlled for, these relationships could have had an indirect effect on processing times/neural activity. For example, Federmeier et al. (2002) compared various groups of stimuli that varied in their likeliness but also in their within- or between-semantic category relationship to the expected target. Furthermore, Bannon et al. (2025) do not report controlling for semantic overlaps within their low constraint (neutral type) sentences, nor ensuring that semantic relationships were balanced between their high and low constraint

(i.e., more and less likely target word) constructions. Thus it is difficult to eliminate the potential contribution of these relationships to the observed likeliness/expectedness effects in previous studies.

Notably, despite the fact that sentences in condition A (Unlikely) elicited an interference effect for both younger and older adults in the current study (unlike in Chapter 2), we did not observe age-group differences in the degree of interference caused by this condition. This suggests that the absence of a significant age-group difference in the mismatch effect in Chapter 2 was unlikely to have been because the mismatched sentences were not mismatched/interfering enough.

5.4.2 Semantic relatedness (Research Question 3)

Our final research question asked how semantic relationships between unviable sentence endings and the preceding context/ the expected sentence ending influence interference effects. We found that a stronger semantic overlap created *greater* interference in terms of naming RTs as well as accuracy. This is in line with some previous research findings, particularly those obtained from studies utilising the picture-word interference paradigm (e.g., Taylor & Burke, 2002). We hypothesised that semantic overlap with the unexpected word could either create facilitation or interference. Collectively, these data align with the direction of our hypothesis that the presence of conflicting semantically salient information during lexical selection creates *interference* by competing with required target words. In the current study, priming/prediction mechanisms would have prepared participants to produce the expected endings of the preceding sentences. However, needing to produce a word that was strongly associated with the expected target instead (e.g., producing "***nurse***" when "***ambulance***" was expected) seems to have created *competition* between the two words (causing delays in production), rather than *facilitating* its retrieval (as would be expected within some models of language production, e.g., Mahon et al., 2007). The Inhibition deficit hypothesis accounts for the greater interference effect observed within Condition B (semantically related condition) through greater competition between co-active representations requiring greater suppression via inhibitory control mechanisms. Alternately, interference can also be explained through 'activation-based' models such as the transmission deficit hypothesis, wherein interference arises when insufficient top-down excitation reaches phonological nodes of a word a speaker wishes to produce. Lexical items that share semantic and phonological

similarities to the target might receive relatively more excitation that the target, resulting in competition between the two (Oberle & James, 2013).

The finding that greater semantic overlap created greater competition contradicts our alternative hypothesis, which is rooted in a vein of research reporting facilitatory effects for processing words that are strongly associated with expected upcoming words. These effects are accounted for by prediction/priming mechanisms through which automatic spreading activation between primed/expected words and their neighbours *eases* the retrieval of those related words too (e.g., '**salt**' being easier to process following the sentence "*I take my coffee with cream and*" owing to its relatedness to the expected ending '**sugar**', Broderick et al., 2021).

Aside from the fact that these differential findings might be explained by differing underlying mechanisms involved in language comprehension and production, it is worth highlighting here that the sentences we used in condition B were semantically related to the sentence context and/or the expected target word. The word2vec scores were slightly higher for the relationship with the sentence context than the expected target word, but both relatedness scores were higher than those of the unrelated condition. Previous studies have typically only manipulated the target words relationship to the expected word (e.g., Broderick et al., 2021; Federmeier & Kutas, 1999a; Federmeier, 1999b, Federmeier et al., 2002, Federmeier, 2007, and Federmeier and Kutas, 1999). Further research may be useful in elucidating the individual (facilitatory or inhibitory) contributions of unexpected target words being related to the preceding sentence contexts versus the expected target word.

Finally, it is worth noting the surprisingly significantly higher likeliness ratings provided by both younger and older adults for the stimuli in condition B (Impossible, semantically related) compared to Condition C (Impossible, semantically unrelated). Although, as intended, the likeliness ratings of target words in both of these conditions were as close to floor as possible, and not statistically different to one another in the pilot analyses, in the main study, the likeliness ratings were higher for Condition B (although still close to floor). If anything, we expected more likely target words to be associated with faster naming, while the reversed was observed for Condition B. However, it is possible that the higher likeliness of the target words in this condition may have contributed to creating more competition with the expected words, thus leading to greater interference. The difference in likeliness ratings between Conditions B and C, however, was far smaller than the likeliness difference between conditions A and C.

Given that no RT differences were found between conditions A and C, the small likeliness differences between B and C is unlikely to explain the observed RT difference between those conditions. Furthermore, the analyses including the actual likeliness ratings showed no significant relationship with naming RTs, again suggesting the RT effect discussed here is more likely explained by our semantic relationship manipulation than by small likeliness differences.

Moving onto the effects of age, we found that older adults, similar to the younger adults, experienced hindrance from semantic relatedness, rather than facilitation. We expected that if interference occurred more strongly in the semantically related condition, its effects might be *greater* in the older adult group due to widely reported age-related declines in semantic control (Hoffman, 2018); or *weaker* if older adults do not utilise semantic context in the same way younger adults do (e.g., Broderick et al., 2021). However, we found that older adults experienced a similar degree of interference from the semantic relatedness manipulation, opposing both arguments.

With regards to semantic control, we found that scores differed significantly between age groups in the expected direction (with older adults being more strongly affected by the incongruent feature association trials than the younger adults were). However, this difference score was not significantly related to the effects of interest in our study (i.e., differences in RTs between conditions A and C, and B and C) despite our expectation that semantic control might interplay into these effects, especially also when looking at semantic relatedness effects. This lends some support to the possibility that neither older nor younger adults strongly relied on semantic control to manage the interference, including interference from semantically related unexpected words, during the language production task.

5.5 Conclusion

Studies on language comprehension show that both the degree of mismatch and the semantic relationship between target words and the contexts within which they occur can have significant implications for language processing. The current study explored these effects within the context of language production in healthy ageing. The interference effects suggested that in production too, both younger and older adults' word production can be influenced by sentence contexts that are incongruent with target words. Despite this interference being present, older adults did not show greater difficulty with the unexpected words than the younger adults. This suggests that despite showing greater semantic-control

difficulties in the separate semantic control task, the two age groups did not differ in how they experienced and potentially managed interference during language production. Finally, we showed that a strong semantic overlap between context-target resulted in strong interference effects in both healthy younger and older adults. Collectively, these findings show that context-target mismatches which have previously been observed to create interference within comprehension contexts, can also create interference within language *production*. The findings also lend further support to the argument that connections between semantic representations in the mental lexicon are preserved in old age (since these are necessary for semantic competition to arise). One point that requires further clarification though is how semantic control processes play into these contextual interference effects.

Chapter 6: General discussion

6.1 Thesis Overview

Spoken language is the primary method of communication used by humans to interact with each other. As people age, retrieving the words they need for communication becomes more difficult (Condret-Santi et al., 2013; Lovelace & Twohig, 1990). Conversation itself imposes many demands on speakers, requiring them to not only *produce* words; but to also engage in multiple other processes concurrently. During (and prior to) each conversational turn, speakers are tasked with comprehending their partner's utterance; planning and formulating a verbal response that is appropriate and relevant to this; estimating when their partner will finish their ongoing turn so they can time their own utterance; and eventually producing their response (Bogels et al., 2018 De Ruiter et al., 2006; Magyari et al., 2014; Riest et al., 2015). In addition to this, the nature of the speech produced by their partner may affect how efficiently speakers can produce subsequent words. Speech contexts that are matched with required words typically ease their retrieval, while mismatched contexts might instead interfere with their retrieval (e.g., Bannon et al., 2025; Shao & Rommers, 2020).

Ageing is associated with specific changes in terms of semantic and cognitive functioning (e.g., Hoffman, 2018) that might affect how older adults respond to matched and mismatched sentence contexts. Increased accumulated semantic knowledge might help maintain or improve their ability use of matched contexts to retrieve predicted/primed words. On the other hand, decrements in semantic and inhibitory control might negatively affect older adults' ability to manage competition between primed and target concepts in mismatched sentence contexts. Thus, across the empirical chapters in this thesis, I have sought to study word production processes in response to spoken sentence contexts, to capture conversational processes more fully than possible in previous studies focusing on naming of individual pictures. Specifically, I compared the effects of different variations of matched and mismatched sentence contexts on word production, in groups of healthy younger and older adults, while also studying potential contributions of various theoretically and empirically implicated semantic and cognitive abilities.

My initial findings (reported in Chapter 2) showed that younger and older adults are similarly able to use simple matched constructions to aid in the retrieval of required words. Conversely, I did not find that simple *mismatching* contexts cause interference in younger or

older adults' word production. Following this, in Chapter 3, I extended these findings by asking the same participants to complete a series of tests measuring various cognitive and semantic capabilities. Finding significant age-related differences in some of the expected abilities, I then examined the association between participants' performance on the cognitive tests and the context effects measured in Chapter 2. In Chapter 4, I moved onto exploring resource theories (which postulate that cognitive resources such as working memory decline as a result of ageing). Specifically, I investigated if/how this decline affects older adults' ability to leverage more complex, as compared to simpler, matched contexts (as more complex sentence contexts were postulated to exert more strain on working memory resources). I also examined whether the position of semantic primes within matched contexts affected speech planning mechanisms differently in younger and older adults. Finally, in Chapter 5, I turned my attention to mismatching contexts, focussing on the effects of semantic control declines in older adults. Here, I studied the effects of the magnitude of the mismatch between contexts and target words, as well as the degree of overlap between context/expected word and target words.

In this concluding chapter, I will first present a summary of each empirical chapter; focussing on the research questions I addressed, the work I carried out, and how my findings relate to relevant literature and theoretical frameworks. After this, I will coalesce the work carried out across all chapters to discuss wider theoretical and practical implications as well as limitations and recommendations for future research.

6.2 Summary

6.2.1 Chapter 2

Previous studies demonstrating age-related declines in word production have typically adopted single naming tasks, which record the accuracy and efficiency with which participants produce the names of pictured target words (in isolation) (e.g., Goulet et al., 1994). Despite its merits as a reliable experimental method, picture naming in isolation does not capture the myriad of additional influences that exist within natural everyday speech. One of the primary aims of this chapter was therefore to try to understand if word production differs between isolated single words and in more natural speech contexts. A second goal of this chapter was to study the effects of matched sentence contexts on younger and older adults' word production. Younger adults have been widely shown to utilise matched sentence contexts to

facilitate their production and comprehension of predicted/primed words (e.g., Haigh et al., 2022; Shao & Rommers, 2020, as reviewed in Chapter 1). On the basis that semantic knowledge stores remain intact and even expand with ageing (Carrol, 2023; Hoffman, 2018; Hoffman et al., 2018; Kavé & Yafé, 2014; Verhaeghen, 2003), and the finding that automatic semantic priming processes remain stable in older adults (Balota et al., 1999; Faust et al., 2004), it was postulated that this group's word retrieval may be facilitated by matched contexts too. The final aim of this chapter was to understand if mismatched sentence contexts, which prime words that conflict with the target, can *interfere* with language production. Some previous work (mostly focussing on comprehension) with younger adults suggests that producing/processing words in contexts they are mismatched with is associated with delays in behavioural responses (Bannon et al., 2025) and larger N400 amplitudes within EEG studies (e.g., Federmeier et al., 2002). Based on empirical findings that demonstrate age-related declines in semantic and inhibitory control processes, which allow for the management/inhibition of unwanted active information (Hoffman, 2018); as well as relevant theoretical frameworks (e.g., the inhibition deficit hypothesis, Hasher & Zacks, 1988), older adults were expected to be more strongly affected by mismatching contexts than younger adults.

Participants named target pictures following a matched, mismatched, and neutral sentence context, and once in isolation. The data showed that participants in both groups were not significantly facilitated nor hindered by (neutral) sentence contexts relative to producing the same words in isolation, after accounting for the effects of task order. This finding appears to speak against cognitive resource theories that postulate that stronger task demands (i.e., producing words in speech contexts) can strain cognitive resources more than less demanding tasks (producing words in isolation), especially in older individuals (Kemper et al., 2001b; Kemper & Kemptes, 1999), which could be predicted to hinder word production in context compared to in isolation. At the same time, these data suggest that any previously observed differences between age groups in word production tasks requiring naming of individual pictures versus tasks eliciting connected speech are unlikely just because of the presence of context, and more likely related to the different measures used across studies (cf. Kavé & Goral, 2017).

Both age groups were equally facilitated by matched contexts. These data suggest that older adults can continue to benefit from priming in sentence contexts and/or continue to

form predictions in a similar manner to younger adults. This lends further support to prior findings that semantic networks and priming mechanisms are preserved in old age (Balota et al., 1999; Kavé & Yafé, 2014). The finding that neither older nor younger adults did not experience interference from mismatched contexts echoes previous behavioural studies implementing similar sentential constructions (sentence contexts that were not too strongly mismatched with target words; e.g., Haigh et al., 2022). Although other work has shown interference even with the weaker types of mismatches used here (Bannon et al., 2025), interference effects reported in the literature have been found to be greater in magnitude and more robust when stronger mismatching contexts are used (both behaviourally and neurally; and for both language comprehension *and* production, e.g., Bannon et al., 2025). Taken together, these findings suggest that within the context of language processing/production, the context-target mismatch threshold for eliciting reliable interference effects might be quite high. Further support for this notion is provided by the finding that even older adults did not experience any degree of interference from these less strongly mismatched contexts, despite showing weaker semantic control and general inhibitory skills, which should make it more difficult for them to supress active, distracting information during ongoing language/cognitive operations.

6.2.2 Chapter 3

In Chapter 3, employing the same participants that took part in Chapter 2, I examined performance on a range of semantic and cognitive tasks, including measures of semantic knowledge, semantic control, inhibitory control, and working memory. The data largely supported the expected differences between age groups, in that older adults exhibited greater semantic knowledge but (to some extent, although not consistently) poorer semantic and domain general control. However, contrary to our expectations, older adults performed similarly to younger adults in terms of working memory (perhaps owing to issues with administering this task online). In this chapter I also tested the postulated relationships between producing words in different types of sentence contexts and relevant semantic and cognitive abilities. Specifically, I examined how participants' performance on the semantic and cognitive tasks predicted three aspects of their behaviour: their match effect (RT difference between naming in matched and neutral contexts), their mismatch effect (RT difference between naming in mismatched and neutral contexts), and their context effect (RT difference

between naming without context and within neutral contexts). The data surprisingly showed that none of the cognitive/semantic abilities tested held as significant predictors of these effects. This suggests the cognitive abilities assessed may not directly be related to the computed sentence context effects from Chapter 2, or even to word retrieval processes more generally (as supported by exploratory analyses not showing a relationship with overall naming times) across younger and older adults..

Despite the absence of a correlation with the sentence context effects, however, it is possible that preserved or increased semantic knowledge (as observed at the group level for older adults) is a prerequisite for older adults to continue use priming and/or prediction mechanisms during speech processing. This is since the mechanisms that underlie these processes (such as spreading activation, or formulating predictions based on associations between words and concepts that often present concurrently) are dependent on connections within semantic networks to occur. The fact that older adults exhibited an equivalent 'match effect' to younger adults in Chapter 2 suggests that the use of predictive/priming mechanisms continues in old age. The role of declining semantic and cognitive control in older participants' decreased ability to manage conflicting information during language production was not clear based on the findings in this chapter. Previous findings have supported theoretical frameworks such as the Inhibition deficit hypothesis, showing that older adults experience difficulties in managing co-active information during speech production (evidenced by difficulties in adhering to conversational topics, and producing relevant utterances in ongoing speech). Furthermore, the extent of these production difficulties is related to individuals' semantic and cognitive control performance (Hoffman et al., 2018). In Chapter 3, while older participants did exhibit some reductions in semantic and cognitive control (supporting the inhibition deficit hypothesis), this did not relate to how well they could manage co-active representations when producing words in mismatched contexts. This finding further reiterates that perhaps the information comprised within the mismatched sentence-target pairs was not *conflicting enough* to induce interference, such that even participants who exhibited the poorest performance in terms of semantic/cognitive control did not experience interference within mismatched trials in comparison to those with better control. Finally, with regards to the role of resources such as working memory capacity in producing words in conversation, the evidence from Chapter 3 was inconclusive. It is possible that this relationship would be more

evident if more demanding sentence contexts (requiring more cognitive resources) were used rather than the simple contexts used in Chapter 2.

6.2.3 Chapter 4

In Chapter 4, I moved on to examining prediction and priming within matched contexts while accounting for some of the additional demands speakers face while engaging in conversation. These include the demands associated with processing more complex sentence contexts (Gibson, 1998, 2000) as well as planning their next turn in the conversation (Bogels et al., 2018). Specifically, in this chapter I compared younger and older adults' production of expected words following longer/more complex sentence contexts to shorter/simpler sentence contexts. The presented work was rooted in prior findings showing that older adults experience increased difficulties in comprehending (Feier & Gerstman, 1980; Lash, 2010; Walsh & Baldwin, 1977) and producing (Kemper et al., 1989) complex sentence constructions. These difficulties are thought to relate to age-related declines in resources such as working memory, which are responsible for maintaining information during ongoing cognitive/language processes (Liu & Wang, 2019). These resources are arguably more strongly taxed when processing demands are high. Participants' production of an accurate response was sensitive to sentence length, with more accurate responses produced following shorter constructions. This could suggest participants indeed needed more cognitive resources, as also shown through the relationship with working memory scores, to respond accurately to more complex sentences. However, this pattern was not reflected in the RTs and did not differ between age groups, providing no support for theories that attribute older adults' difficulties in cognitive or language tasks to reduced cognitive resources.

To assess speech *planning* mechanisms, I manipulated where in the sentence constructions semantic primes were placed (these were placed either at the start or towards the end of a sentence). We found that both younger and older participants were faster to produce words when primes were placed earlier on within the sentence context. This supports theoretical frameworks arguing speakers begin planning their next turn in the conversation as soon as they have obtained enough information do so (Bogels et al., 2015a; Bogels et al., 2018), and further suggests this extends to older as well as younger adults.

6.2.4 Chapter 5

In my final empirical chapter, I turned my attention towards mismatching sentence contexts. Quite often speakers are required to produce words that are different to the ones they expect to produce based on the sentence context (Luke and Christianson, 2016). In this chapter, I therefore focussed on two questions addressing the effects of target word likeliness and semantic relatedness to the prior context and the predicted word on word production. Previous work with younger adults has demonstrated that in contexts where the likelihood of a word occurring is lower, words are produced more slowly than in contexts where they are relatively more expected/likely (Bannon et al., 2025). Furthermore, within ERP contexts, more unexpected words are associated with larger N400 effects (Thornhill & van Petten, 2012). In my study, I did not replicate these findings. In contrast (and contrary to Chapter 2), both strongly and weakly mismatched context-target pairs were associated with delays in RTs relative to a neutral baseline. Importantly, the degree of interference created by each of these did not appear to further influence production. Although this finding does not align with prior work that focussed on this effect within language production contexts, some consideration must be given to the methodology used in that research. For example, Bannon et al. (2025) gauged production times through participants' button pressing responses, which may not be a valid measure of word production times. Crucially, we here showed that older adults do not experience more interference than younger adults when such interference is actually present. This strongly suggests that the older adults tested continued to be able to activate the required target words while potentially suppressing interference from other competing words.

Addressing my second research aim in this experiment, it is known that semantic relationships can create both facilitation (through priming mechanisms, Faust et al., 2004) and interference (through competition, Hoffman., 2018). In this chapter then, I sought to understand the effects of targets sharing a semantic relationship to the prior (mismatched) context/ the predicted word. In line with the interference account, the findings showed that this overlap hindered language production times more strongly than unrelated mismatched contexts in both groups. This suggests that active but unnecessary semantic information can cause interference within language production (rather than facilitating responses thorough priming mechanisms).

6.3 Theoretical and methodological contributions of the empirical work

6.3.1 The effects of sentence contexts on word production

The work presented in this thesis shows that the sentence contexts that speakers produce words in can have a significant impact on language production processes. For example, both younger and older participants' production of target words was facilitated by matched sentence contexts (Chapter 2); while it was hindered by mismatched ones (Chapter 5) to a similar degree across younger and older participants.

Facilitation through prediction and/or priming in matching contexts

Focussing first on the effects of matched contexts, these have widely been shown to facilitate the production of predicted/primed words in both younger and older adult participants (Bannon et al., 2025; Shao & Rommers, 2020). When studied comparatively to one another, older adults derive as much (Haigh et al., 2022; Kliegl et al., 2004), if not *more* (Pichora-Fuller et al., 1995; Rayner et al., 2006) facilitation from matching contexts than younger adults do. The findings reported in Chapter 2 align with the former pattern of results in that older adults gained a similar degree of facilitation from these contexts to younger adults. This facilitation (both in terms of processing and production speed) of words that are matched with the preceding context is believed to be underpinned by two key mechanisms: semantic priming and prediction. While semantic priming refers to the automatic activation of concepts that are related to ones that speakers/listeners encounter within their own or their partner's speech (Meyer & Schvaneveldt, 1971), the automaticity of prediction mechanisms is debated. A theory that brings together the conflicting viewpoints on this subject is that proposed by Pickering and Gambi (2018), which argues that prediction comprises both automatic and effortful phases. Specifically, the earlier (automatic) stage of prediction draws on semantic priming processes, whereas the latter (non-automatic) stage involves formulating more precise expectations based on a variety of sources to make more accurate judgements about upcoming words. These sources are agued to include include knowledge of previous linguistic experiences, such as knowing which words frequently occur together, and which objects accompany particular verbs, as well as knowledge of their conversational partner's habits and likes/dislikes (e.g., Corps et al., 2022, Experiment 1; Kamide et al., 2003, Experiment 2; Pickering & Gambi, 2018).

The finding that the use of matched contexts is stable/increases in old age may be attributed to the well-documented preservation/growth of semantic networks across the lifespan (e.g., Carrol, 2023; Hoffman, 2018; Hoffman et al., 2018; Kavé & Halamish, 2015; Kavé & Yafé, 2014; Verhaeghen, 2003). Maintenance of the connectivity between related semantic/conceptual nodes in these networks is thought to allow aged individuals to continue benefitting from semantic priming: a consequence of spreading of activation between active semantic representations and their neighbours (Meyer & Schvaneveldt, 1971). Existing evidence for preserved (automatic) semantic priming effects in old age is obtained from studies wherein older participants are required to produce or process single words following a prime word e.g., furniture – '**sofa**' (Balota et al., 1999; Faust et al., 2004). Based on the hypothesis that prediction is more difficult for older adults than priming is (as a result of the non-automatic aspects of it; Ito et al., 2016, Ito et al., 2018, experiment 2; Huettig and Janse, 2016; cf. Pickering & Gambi, 2018), it is possible that the continued use of matched contexts in old age is due to preserved semantic priming effects.

Unlike in single word studies though, within sentential contexts, teasing apart the effects of priming/prediction is more difficult due to the amount of information available. Focussing on the matched stimuli used in Chapter 2, while some contexts comprised a definitive prime word that was semantically associated to the target word, others did not (e.g., consider "*what was the dog frightened of*?" - "**fireworks**"). Thus, it is likely that alongside leveraging priming mechanisms, older adults were also engaging prediction to formulate expectations about upcoming words. If prediction is indeed more effortful (and thus more difficult for older adults), this is somewhat surprising given that older adults are expected to have reduced cognitive resources needed for formulating predictions at the same time as managing the demands of ongoing language processing. This finding thus lends support to the viewpoint that prediction too comprises (at least some) automatic components. An alternative possibility is that the older adults who took part in Chapter 2 had not undergone significant age-related changes in terms of cognitive resources such as working memory and therefore they may have been able to engage in prediction as effectively as the younger adults did (supported by the lack of an age difference in performance on this task in Chapter 3).

Sentence and task complexity might be a modulating factor of age-group differences. On the one hand, it is worth nothing that the sentential contexts used in Chapter 2 were generally short, simple constructions that likely did not tax cognitive resources greatly.

Matched contexts that are more demanding may tax cognitive resources more strongly, thus potentially increasing effort in older adults if they are engaging prediction mechanisms while also processing complex spoken input. On the other hand, if older adults can successfully use priming and/or prediction mechanisms to leverage matched sentence contexts, such benefits might be more likely to arise in more demanding tasks ;as they have previously been shown to leverage greater facilitation from these contexts compared to younger adults in other demanding circumstances, (i.e., in noise, Pichora-Fuller et al., 1995). However, the findings from Chapter 4 (showing no interaction between sentence length, prime position, and/or age group) suggest sentence length as such is unlikely to explain the (absence of) age-group differences observed in terms of the match effect. Taken together then, the presented evidence suggests that older adults may be able to engage both automatic and effortful priming/prediction processes. Further research might focus on elucidating how well older adults can engage more effortful processes when the sentence context places stronger demands, beyond sentence length as such, on cognitive resources.

Interference, semantic control and mismatching contexts

Moving onto mismatched sentence contexts (studied in Chapters 2 and 5), I found that these interfered with language production equally in younger and older adults' language production (Chapter 5; although in Chapter 2, these contexts did not hinder production in either age group). It was expected that the degree of interference these contexts caused for individuals' production would be related to their semantic control ability (part of which includes inhibiting irrelevant representations that become automatically activated due to spreading activation within semantic networks). However, we consistently found that performance on a semantic control task did not relate to the size of participant's mismatch effect (See Chapters 3 and 5). Since semantic control has been shown to decline with age, I also expected older adults to be more strongly affected by mismatched contexts due to increased difficulty in managing competing representations. On the contrary, I did not observe a significant age effect in terms of the mismatch effect.

In Chapter 2, participants responded to questions that predicted/primed a certain word by producing unlikely (but still viable) answers. However, one thing to bear in mind is that here neither the target words' semantic relationship with the primed/predicted word nor with the sentence context was directly targeted. For example, consider *"what did she put the*

sleeping baby into?" – "**bag**". The target word (bag) does not share a strong semantic relationship with the expected answer (e.g., cot) or to components of the context (e.g., sleeping baby). As a result, this might explain why the mismatch contexts in Chapter 2 did not relate to semantic control ability. Furthermore, since the target words following these contexts in Chapter 2 were selected such that they were not impossible within the context, it is possible that producing these words did not create *competition* with the expected target (as they were, albeit to a much lesser degree than he target word, still *possible in* the context).

Conversely, in Chapter 5, wherein there *was* a strong semantic overlap between the target, the expected ending and/or the sentence context; the size of mismatch effect was still not related to semantic control score. One possibility then is that the crossover between language processing and production within conversational contexts does not include the involvement of semantic control processes. In particular, there may not be an overlap in semantic processes between different modes of language, such that spreading activation triggered by concepts that a listener *hears* within their partner's speech may not affect their *production* of related concepts within their subsequent speech turn. This explanation is unlikely however, since participants still experienced interference from these relationships within their language production (Chapter 5); and they *leverage* these relationships during language production following matched contexts (Chapter 2). A more likely explanation is that the type of semantic control engaged by the mismatched contexts in Chapters 2 and 5 differs from that assessed by the chosen standard measures of semantic control. For example, note that the feature association task (adopted from Hoffman, 2018; used in Chapters 3 and 5) tested participants' ability to suppress dominant semantic relationships (e.g., between probe: salt and distractor: pepper) while selecting a target (dove) which shared a featural relationship (e.g., colour) with the probe word. In Chapter 3, participants' composite semantic control score also included an additional measure (global association), which assessed their ability to draw the relationships between weakly semantically associated words (e.g., ring and iron). Arguably, neither of these measures might directly target the type of semantic control needed to produce words that are unviable (but related to the) sentence context or conflict with those predicted/primed by preceding sentence contexts. Supporting this, some neuroimaging studies have revealed that different kinds of semantic control involved in different tasks contexts draw on different neural networks, lending support to the idea that there are varieties of semantic control which may change at different rates or in different ways during

ageing (Badre, 2005; Krieger-Redwood, 2025). As a result, future work should study a range of semantic control measures that tap into different underlying abilities in relation to language production, to better understand what (if any) aspects of semantic control are involved in language production.

Relationships to previous (EEG) literature

The reported findings concerning matched and mismatched contexts appear to contradict a vein of literature which has focussed on participants' neural activity during online processing of predictive contexts. The findings from this literature overall suggest that older adults are less effective in formulating predictions compared to younger adults. An example is their reduced N400 benefit, relative to younger adults, from matched contexts when processing endings that are predicted or primed by those contexts (Federmeier et al., 2002, Federmeier & Kutas, 2005; Gunter et al., 1992; Hamberger et al., 1995; Wlotko & Federmeier, 2012; Woodward et al., 1993). Specifically, while younger adults' N400 responses are significantly reduced when processing words predicted by the sentence context, older adults' N400 responses in these contexts are smaller and slower (Ford et al, 1996; Wlotko & Federmeier, 2012). Furthermore, older adults do not always demonstrate facilitation (i.e., reduced N400s) for the processing of words semantically related to the sentence context (Federmeier et al. 2002). In contrast, as discussed earlier, when employing tasks that focus on behavioural measures (including those employed in this thesis), older adults do demonstrate continued use of prediction and priming. Taken together, these findings suggest that the effects of priming/prediction during sentence processing appear to diminish in terms of neural responses with age, while remaining evident in terms of participants' behavioural responses during language comprehension and production. This may be explained by differences in the underlying mechanisms that are captured by each of these measures. For example, EEG studies focus often focus on online processing of speech or text signals, without participants needing to respond to that input. Behavioural studies require participants to comprehend and respond in some way to the information contained within those signals, as is often needed in daily life interactions.

One possible interpretation is that EEG studies are more sensitive to age-group differences and/or that such differences arise most strongly within the first 400-600 milliseconds of speech processing. While ERP studies can detect such, potentially more subtle,

effects at different and early time points, behavioural studies only capture the "end point". In particular when studying ageing, it is possible older adults use other mechanisms or strategies to overcome small delays in the earliest processing stages, resulting in their *responses to* that input being comparable to those of younger adults.

The process of language comprehension itself has also been argued to trigger prediction through the language production system itself (see Pickering & Gambi, 2018, prediction through production). It is therefore possible that production is influenced by the prediction mechanisms used during preceding comprehension, leading to differences between comprehension (typically employed in ERP studies) and the current production studies. Previous behavioural work using similar stimuli as Chapter 2 to study mismatch effects during comprehension (Haigh et al., 2022), however, did not show age-group differences either. This suggests potential prediction differences between comprehension and production are unlikely to explain the absence of age-group differences observed here.

Methodological contributions of the presented work

One of the key methodological contributions of the presented work is that across the chapters, I derived the magnitude of facilitatory and inhibitory effects of matched and mismatched sentence contexts on word production by comparing participants' word production efficiency within these contexts to a *neutral baseline condition*. This allowed me to accurately determine and tease apart the facilitatory and inhibitory effects of the employed sentence contexts. This builds on previous (comprehension) work, which has often compared language processing/production within matched contexts to mismatched ones, or different types of matched contexts directly to one another. it is difficult to reliably gauge exactly how much (if any) facilitation or interference at all is caused by the different contexts in these studies.

A second goal was to understand how ageing affects word *production* in different types of more naturalistic sentence contexts. The literature is currently divided into two key areas that each employ different measures, making it difficult to compare the two. The first comprises studies focussed on single word retrieval in isolation (without context), which gauge measures such as response accuracy and times taken to produce target words. The findings from these studies are suggestive of age-related declines in word production efficiency and accuracy (e.g., Connor et al., 2004; Evrard, 2002; Gordon & Kindred, 2011; Gordon & Kurczek,

2014; Kavé & Goral, 2017; Tsang & Lee, 2003); critically, despite their use of reliable measures, these tasks lack ecological validity. The other area of the pre-existing literature focuses on the production of connected speech. These studies have investigated a variety of variables including word productivity (number of words produced), lexical diversity (variety of words produced, and type-token ratio), and word retrieval failures within more naturalistic scenarios; and have often not found evidence for age-related declines (see Kavé & Goral, 2017 for a review). Although these measures focus on language production within naturalistic scenarios, they cannot provide precise insight to mechanisms relating to the retrieval of individual words; and may allow participants to mask any word-related difficulties making them harder to detect.

Across the studies I have reported, I employed paradigms which involved participants producing single words following the presentation of a pre-recorded verbal sentence context (adapted from Shao & Rommers, 2020), to try to mimic word production within typical, everyday conversation. In Chapters 2 and 5 this involved participants producing the names of pictured targets after a sentence context was presented. This methodology allowed me to determine the words that participants needed to produce. From a researcher perspective this is extremely useful since younger and older adults were tested on their ability to produce the same words; and thus comparisons between age groups could be made. At the same time, this allowed me to manipulate context to address my research questions.

In Chapter 4, with the aim of making my experimental paradigm even more naturalistic and to test the notion that older adults might use compensatory strategies, such as the use of circumlocutions to mask their difficulties in retrieving words (Kavé & Goral, 2017; Nicholas et al., 1985), I allowed participants to choose which words they produced to respond to the preceding sentence. This also allowed me to test if older adults pick easier words than younger adults. As expected, words that were accepted as being accurate but were different to the target word we were expecting participants to produce, were higher in frequency than the expected target words across age groups. However, these alternative words were similar in frequency between the younger and older participants, suggesting that older adults did not select comparably easier words as such, which may be expected if lower frequency words are relatively more difficult for them to retrieve than they are for younger adults. These findings speak directly to theoretical standpoints such as the Transmission Deficit Hypothesis, which postulates that ageing disproportionally affects the production of low frequency words in

older individuals (Burke et al., 1991). Since lower frequency targets are less imageable, it is difficult to embed them within picture naming tasks; thus, the paradigm adopted within Chapter 4 was unique in that it allowed for this observation while also being able to measure reaction times and manipulate the role of context. Despite these merits of the task employed within Chapter 4, the paradigm was flawed in a number of ways, as already discussed in the 'limitations' section of Chapter 4

6.3.2 Absence of age group differences in context effects: theoretical implications

Expectedly, older adults were *slower* in their verbal production of target words overall across the reported chapters. However, we did not find that *context* had a differential impact on word production in the two age groups. The overall RT difference between the two groups might be explained in a variety of different ways such as general age-related slowing (Salthouse, 1996), or semantic neighbourhood density (i.e., words with a greater number of semantic relatives are arguably harder to retrieve than those with fewer relatives, due to the competition between these co-active representations during lexical retrieval. Furthermore, this difficulty may be exacerbated by age Mirman, 2011). There may also be other factors that could have contributed to these differences, including phonological properties of target words, such as number of phonological neighbours, constitutional phoneme frequency, and other phonological properties (James & Burke, 2000). Older adults may be more strongly affected (negatively) by properties such as sparse phonological connectivity for target words, since these connections may be further weakened by age, thus making word retrieval more challenging (Burke et al., 1991, 2004; Burke & Shafto, 2004). However, given that this was beyond the scope of the thesis research, it is not possible to delineate the effects of these phonological properties based on the data collected.

Beyond overall slowing, my thesis focused on effects of context. One important implication of the presented work is that it addresses some of the key theories of language and ageing. Below, I will thus focus on these theories and explain how my findings relate to them.

Resource theories

Resource theories propose that language production difficulties arise in old age because of natural declines in vital cognitive resources such as processing speed (Birren, 1965; Cerella,

1985) and working memory. The data reported in this thesis somewhat support these theories. For example, the between age-groups word production RT comparisons carried out in Chapters 2 and 5 showed that older adults' overall word production speed was slower than younger adults'. However, contrary to the predictions of resource theories, in Chapter 4, we observed that older and younger adults both were faster to produce words following longer, more complex sentences than they were following short simple sentences. According to resource theories, it would be expected that when the processing load is higher (within the more complex sentences), fewer resources would be spared for production, which consequentially could lead to slower word production times, *especially within older individuals.* Also in disagreement with resource theories is the finding that none of the tested resources such as working memory or cognitive/semantic control were related to word production times generally (Chapter 3) or to any of the contextual effects computed within the different chapters (i.e., the match, mismatch and context effects in Chapter 2, the sentence length effect in Chapter 4, or the mismatch effects reported in Chapter 5). Furthermore, neither younger nor older adults were negatively affected by producing words within contexts relative to producing words in isolation within Chapter 2.

Semantic knowledge and matched contexts

The work presented in this thesis supports previous work (e.g., Carrol, 2023; Kavé & Yafé, 2014; Verhaeghen, 2003) showing that the semantic system is preserved in old age (supported by the finding that older adults outperformed younger adults on the synonym judgement task, in Chapter 3). Semantic connections are thought to be paramount to enabling semantic priming effects, as the activation of encountered concepts/words triggers spreading activation to related ones. Supporting this, semantic priming effects are widely shown to remain intact in old age (Balota et al., 1999; Faust et al., 2004). For example, viewing or listening to prime words facilitates younger and older adults' production of semantically related words. My findings show that these priming effects can also occur within sentence contexts (see Chapter 4), such that semantic cues comprised within streams of speech that individuals comprehend can facilitate their subsequent production of associates of those primes. Also of relevance here is the finding that in *mismatching* contexts, older and younger adults were both hindered to a larger extent when there was a semantic overlap between unviable sentence endings and the preceding sentence context than when there was not. Semantic interference is an

alternate effect of spreading activation mechanisms within semantic networks (e.g., Taylor & Burke, 2002); thus reiterating that these networks are indeed preserved in old age. The preservation of these networks may also be paramount to *prediction* since words that may be likely to occur together may also be associated through semantic connections. It is not clear from the reported findings, which of these two processes participants were using, or if they were using both; thus, future research should seek to more precisely elucidate which of these two mechanisms older adults can draw upon within conversational circumstances.

Surprisingly, I did not find a direct relationship between the size of individuals' semantic knowledge and the degree of facilitation they leveraged from the matched contexts in Chapter 2. Although not directly related to the match effect, it is possible that these increments in semantic knowledge served as a pre-requisite for the preservation in older adults using these contexts to *facilitate* the production of primed/predicted words. This contradicts the proposal that age related increases in semantic knowledge could create *more* difficulties in word retrieval because of competitiveness between concurrently activated representations (Ramscar et al., 2014). This is surprising considering the additional widely documented declines in older adults' semantic and inhibitory control, which could make managing these co-active representations particularly challenging for older adults. One explanation for the lack of support for this proposal is that taken collectively, the studies reported in this thesis found inconclusive evidence for age related declines in semantic control processes (perhaps due to participating older adults representing a subgroup of this population, who due to educational and lifestyle factors might be protected against some forms of cognitive decline, discussed further in "Limitations of presented work and future research directions"). Future research employing older adults from different demographic backgrounds could therefore help to further evaluate the argument of Ramscar et al. (2014).

The Transmission Deficit Hypothesis is one theoretical viewpoint which argues that a feature of the ageing lexicon is a preserved semantic system characterised by well-connected networks of conceptual knowledge (Burke et al., 1991, 2004; Burke & Shafto, 2004); as well as preserved connectivity between semantic and lexical levels of representation in the mental lexicon. This is argued to be since multiple semantic connections converge onto single lexical items, promoting *accessibility to*, and *activity within* these representations. In contrast (due to sparser connectivity between lexical and phonological levels of representation) the Transmission Deficit Hypothesis argues that this is the locus of deficit for older adults' word

production difficulties. In the current thesis, the matched contexts targeted the relationship between semantic and lexical levels of representation in the lexicon; and this indeed showed that this connection was preserved since the word forms of items that were predicted/primed (on a semantic level) were easier to retrieve that those that were not. However, to test the assumption that the deficit lies in the connectivity between lexical and phonological levels of representation should be a goal for future research.

Semantic control/ Inhibition and the mismatch effect

According to models such as the Inhibition deficit hypothesis (Hasher & Zacks, 1988), older adults experience difficulties in managing competing information during cognitive tasks because of the weakening of inhibitory/ semantic specific control mechanisms. In conversation, these mechanisms may be important for allowing speakers/listeners to inhibit words that are predicted/primed by the context, but not actually needed, in favour of producing different, required words (as was the case in mismatched contexts). In line with the Inhibition deficit hypothesis and previous research, the older adults exhibited some declines in inhibition and semantic control (see Chapters 3 and 5) but neither younger nor older adults exhibited a mismatch effect, and performance on the inhibitory/ semantic control measures was not related to the size of individuals' mismatch effect. On the one hand, it is possible that the older adults who volunteered to take part in the reported studies did not experience (strong enough) declines in inhibitory/semantic control for them to be evident within their language production. This is supported by the finding that declines in semantic control in particular were inconsistent in Chapter 3 (older adults performed more poorly than younger adults on some measures, while *outperforming* them on others). Therefore, the participants that chose to take part in the studies may have represented a small group of overachieving older adults rather than representing older adults more generally (see "Limitations and future directions"). On the other hand, it is possible that the administered measures of semantic control measured a different type of semantic control to that which was tapped within mismatched contexts (see "Interference, semantic control and mismatching contexts").

6.3.3 Alternation between language comprehension and production in conversation

The finding that participants begin planning their speech as soon as they can (while their partner is still taking their turn in the conversation) suggests that speakers' language

comprehension and production operations overlap (See Chapter 4). This has implications for how cognitive resources may be split between the two (especially since comprehension and production are known to engage the same underlying brain regions and neuronal and attentional resources; Jongman, 2021; Menenti et al., 2011; Segaert et al., 2012; Shitova et al., 2017). These resources may be strained further considering that keeping up with the conversational pace may require speakers to formulate predictions about their partner's and their own upcoming speech (Pickering & Garrod, 2013). Additionally, if the speech produced by their partner (which they need to process) or what they themselves need to produce is complex (i.e., grammatically or comprising lower frequency words), this may also increase the demand on these resources.

Considering that such resources may already be limited in old age (see "resource theories") we expected that older adults might experience relatively more difficulty in managing these demands. Within their language production performance this may have presented as delayed production times if this group strategically saved production operations till later (just before their next conversational turn). In contrast, we found that older adults alternated between comprehension and production operation with ease, both in trials where processing demands were stronger and when they were weaker, with the only difference with the younger adult group being an increase in RTs (which were independent of contextual complexity). One thing to note is that participants' RTs were not associated to their working memory score; this could indicate that the sentences contexts may not have been designed in a way that targeted this resource in the expected way. Use of different measures of working memory, or of different resources may be more appropriate to capturing the role of these in the management of comprehension/production demands within conversation.

As briefly discussed above, one process which allows speakers to keep up with the quick pace of turn taking in conversation is *planning* of their next utterance (Bogels et al., 2015a). Two differing viewpoints exist in relation to this process. The first proposes that participants begin planning speech for their next conversational turn while their partner is still taking their turn in the conversation, as soon as they have derived enough information from their speech to do so (termed 'early planning', e.g., Bogels et al., 2018). The second viewpoint advocates that speakers do not begin planning their speech until their next turn is imminent (termed 'late planning', e.g., Sjerps & Meyer, 2015). The work presented in Chapter 4 extends previous (EEG) findings obtained from younger participants supporting the notion of early

language planning (Bogels et al., 2015a; Bogels et al., 2018). In Chapter 4, both younger and older participants were faster to produce upcoming words when semantic primes were placed earlier rather than later in the preceding context. While earlier planning may strain cognitive resources such as working memory in older adults, later planning may have implications for another resource (i.e., speed). Since older adults are slower to produce words, leaving response planning too late may leave too little time for planning the next response, which may interrupt the conversational flow. Thus, planning their response earlier may aid older adults in keeping up with the conversation.

6.4 Limitations of presented work and future research directions

One of the primary goals of the work presented in this thesis was to contribute to the body of literature surrounding word retrieval/production changes that accompany healthy ageing, within more naturalistic contexts. A fundamental prerequisite for achieving this goal was the adoption of experimental paradigms that fulfilled the criteria of eliciting verbal production within conversation-like contexts, while also allowing for the precise derivation of measures such as target naming times and accuracy. The picture naming paradigm (adopted from Shao & Rommers, 2020), which was used in Chapters 2 and 5, met both criteria, allowing me to compare how different types of sentence contexts affected participants' subsequent retrieval of the names of pictured objects.

Despite these strengths, this experimental paradigm is still somewhat limited. One of its key limitations is its focus on the production of single words, rather than dialogues (which is what speakers usually produce within spontaneous everyday conversations). In comparison to producing single words, constructing multiword sentences as well as organising them in accordance with syntactic and grammatical rules (Thompson et al., 2015) is more demanding. The production of dialogues as opposed to single words may also lead to significant age effects (in terms of context effects, which were absent in the current study). For example, producing a single word (rather than a multiword utterance) to respond to mismatching contexts might have made the management of conflicting information within these contexts easier for older adults; since participants could allocate more resources to inhibiting/suppressing unwanted representations. If they were required to produce dialogues instead, this may have spared fewer resources for this, increasing the chances of older adults performing more poorly than younger adults.

In terms of matching contexts, within natural, connected speech, speakers may be required to simultaneously manage multiple active semantic/ predictive relationships (e.g., if there are associations between more than one word within the contexts that one hears and those that they respond with, or if different clauses predict different words). Production demands may also be higher (than producing single words) if listeners need to respond to multiple parts of the context in one response. For example, consider the *following question: "Since it was my birthday on Sunday, first I went to Smith's bakery and bought myself a chocolate cake, then I dropped the cake off home before I headed into town, and there were sales in a lot of shops; so I decided to pick up some bits for my holiday in June. Would you like a slice of cake?"* One may respond to this question by saying *"Oh Yes please! And Happy birthday! I didn't realise your birthday was in June! And oh my gosh, Smiths cakes are the best, bet you enjoyed a good chunk of that cake when you got home from all your shopping! What did you buy*!?" Within Chapter 4, although we increased the processing demand on participants, the context overall only predicted/primed a single word; furthermore, participants responded with a single word to the entire context. The age-related decline in working memory within speech production may be more defined when demands relating to both processing *and* production (based on that processing) are increased.

Although everyday speech consists of partners exchanging dialogues, rather than responding to one another with single word responses, one thing to note is that capturing this interaction is more difficult experimentally. Specifically, it is difficult to design an experimental task that measures *natural* language production processes. For example, observing spontaneous conversation does not allow sufficient control over which words participants will choose to produce (in the paradigms used throughout this thesis, all participants either produced the same words, or produced words that were matched on important characteristics such as frequency and length, which can contribute to how easy/difficult they are to produce, allowing for between-groups comparisons). Furthermore, spontaneous speech does not comprise a reference point relative to which participants' production efficiency/speed (RT) can be gauged; nor does it allow any insight to response accuracy since participants have more opportunity to mask word retrieval difficulties by employing strategies such as circumlocutions or choosing to produce a different word instead. Spontaneous speech paradigms are limited to employing measures that might not target these subtle aspects of word production that may decline with age; and instead, are required to focus on alternative

measures (i.e., word productivity and type token ratio, which can also *improve* with age). In addition, spontaneous speech does not allow control over the speech that conversational partners produce, making it difficult to draw conclusions regarding how processing different types of preceding contexts (matched, mismatched, or neutral) affects subsequent language production processes differently.

It is also worth considering that producing the names of pictured objects may not be a valid measure of natural word retrieval processes. Normally during conversation, words are retrieved from memory without the presence of any external cues. The presentation of a picture of an object may make producing it easier. Furthermore, since participants were required to produce the names of the pictured objects, they had little flexibility in producing the words they naturally might in those contexts. Thus, we attempted to address these limitations in Chapter 4, wherein participants completed sentences missing their final word with the words *they believed completed them.* This gave them more flexibility to produce the words they naturally would use. However, this paradigm presented its own unique challenges. For example, there was variability between the words produced by different participants following the same contexts. In contrast (as mentioned previously), picture naming studies measure all participants' ability to produce *the same words,* making it easier to make comparisons between participant groups.

One other drawback of the sentence completion paradigm used in Chapter 4 is that participants were instructed to wait until the audio file comprising the sentence context had finished playing (and a fixation cross appeared on the screen) before they started producing a verbal response (since their response could not be recorded while the audio was still playing). A number of participants started producing their verbal responses while the audio recording was still playing, however, making it impossible to determine their word production onset times (resulting in them having to be excluded from the study). This questions the validity of the sentence completion task since participants were required to suppress their responses until the fixation cross appeared

Finally, focussing on the effects of a conversational partner's speech on one's own language production (as was the case in all the empirical chapters) does not account for how the sentence contexts *produced by speakers themselves* influence their own production of subsequent words. Fundamentally, producing utterances after a partner's speech captures the effect of preceding *comprehension* on word production while producing words as part of one's

own conversational turn would capture the effects of preceding production on subsequent language production.

Finally, moving from the tasks to the participants, one other potential limitation of the presented work relates to the representativeness of the (older) participant samples included in the reported studies. The older participants showed some expected age-related developments (i.e., semantic knowledge) and declines (i.e., in terms of overall naming times, and on some measures of semantic and inhibitory control). Generally however, older participants across all four empirical chapters performed better than expected (with their performance often being comparable to that of younger adults). Aside from the obvious similar age effects in terms of the context effects, this also extends to performance on tasks such as those measuring semantic control; and short-term working memory (even when this was designed to be more challenging than a simple digit span task, see Chapter 4). Preservation of these abilities might relate to the older adults who took part in the studies. For example, the older participants were typically highly educated (most had attained at least a bachelor's degree). Furthermore, they were recruited either through prolific or through a university departmental database. Highly educated older participants who sign up to platforms such as prolific and to university department participants pools to take part in psychology research may belong to a specific demographic rather than representing all/ most older adults. Those who take part in such research may be interested in their cognitive health more generally and may participate in other activities that promote this. The reported findings may therefore not extend to other groups of older adults.

Taken together, future research should focus on recruiting older adults from a variety of different backgrounds to understand language production and ageing more generally. It should also focus on trying to manipulate the demands of both language processing and production mechanisms to mimic real life conversation more closely (rather than focussing on the production of single words).

6.5 Conclusion

In this thesis, I sought to further our understanding of language production mechanisms within the ageing brain. The presented work fills a gap in existing literature by seeking to explain the everyday effects of the cognitive, semantic, and language changes that accompany ageing, on word retrieval mechanisms. One of the main contributions of this thesis is showing

that while cognitive processes slow down with ageing, older adults process language contexts in the same way that younger adults do. While previous work focussing on language has often highlighted the ways in which ageing hinders this (e.g., lack of ability to continue using prediction mechanisms/ slowing down of linguistic processes); my work highlights that language processes within everyday interactions are largely preserved in old age. For example, older adults continue to benefit from matching sentence contexts that predict or prime the words they need to use in their utterances. In line with this, the older adults also showed larger semantic/conceptual knowledge stores.

Surprisingly, older adults were not hindered more by more difficult language contexts (mismatching ones, as well as ones that were otherwise designed to be complex/demanding) comparatively to younger adults; even though they exhibited declines in resources such as semantic and cognitive control. This suggests that contextually guided language processes may not depend on these controlled cognitive mechanisms; and that prediction during conversation may unfold relatively automatically (rather than being an effortful process).

Further research with older adults from more diverse backgrounds, ages, and education levels is recommended since the older adults participating in the studies reported in this thesis may reflect a narrow demographic which may be more likely to have preserved cognitive functioning in old age. Nonetheless, the reported findings suggest that at least some older adults can manage more difficult language contexts as well as younger adults can.

On a final note, this thesis underlines the importance of studying language production within more naturalistic contexts. Although the methodologies adopted in this thesis have their own limitations, this is an important first step towards better understanding the influence of sentences produced by a conversation partner on the speaker's own word retrieval processes, Future research should aim to further develop and refine these methodologies to allow better understanding of language processes in spoken interactions between people

Appendix A: Chapter 2 Stimuli

Stimuli were divided into two sets. Sentences that formed a "matched" context for one target in Set 1 were a "mismatched" context for another target in Set 2. The neutral sentences were the same in the two sets. Participants either named the Set 1 targets or the Set 2 targets.

Although Sets 1 and 2 were not compared, we ensured the target words in the two sets were comparable in terms of log frequency (NWatch, Davis, 2005; Set 1: $M = 1.577$, $SD = 0.547$; Set 2: $M = 1.458$, $SD = 0.601$; $t(78) = 0.931$, $p = 0.355$), number of syllables (Set 1: $M = 1.500$, $SD = 0.679$; Set 2: $M = 1.525$, $SD = 0.679$; $t(78) = -0.165$, $p = 0.870$); and number of phonemes (Set 1: $M = 4.150$, $SD = 1.511$; Set 2: $M = 4.175$, $SD = 1.412$; $t(78) = -0.076$, $p = 0.939$).

The sentences were also matched in terms of number of words between the matched/mismatched sentences ($M = 6.412$, $SD = 1.651$) and neutral sentences ($M = 6.525$, $SD = 1.867$; $t(118) = -0.337$, $p = 0.737$). Within each sentence we identified the key words (all sentence subjects, nouns, verbs, and adjectives). Mean log frequency and number of syllables of the key words (averaged across the key words per sentence) were also comparable between the matched/mismatched sentences (M frequency = 1.772, $SD = 0.622$; M syllables = 1.578, $SD = 0.602$) and neutral sentences (M frequency = 1.909, $SD = 0.548$; M syllables = 1.772, $SD = 0.582$; frequency: $t(118) = -1.181$, $p = 0.240$; syllable length: $t(118) = -1.686$, $p = 0.094$)

Table A1 Chapter 2 stimuli set 1

Set 1				
	Targets	**MATCHED**	**MISMATCHED**	**NEUTRAL**
1	Present	What did she wrap before Christmas?	What did the football player break?	What did grandfather perceive?
2	Potato	What did the dinner lady mash?	What did the monkey eat?	What did the cat examine?
3	Knife	What was the defendant accused of stabbing the victim with?	What did the dog lick?	What was the young man holding?

4	Ring	What do married women wear?	What did the queen wear?	What did the girl lift off the table?
5	Bible	What did the priest read?	What did the father read to his daughter before bed?	What did the man pick up?
6	Lamp	What did she turn off before bed?	What did the kitchen assistant switch on before preparing dinner?	What did the man use during the day?
7	Rose	What did he give to his date?	What did the thief steal?	What did the girl request?
8	Bin	What did they put the rubbish into?	What did the little girl put the sand into?	What did the children cleanse?
9	Ice	What did he add to the cold drink?	What did the sweet pastry chef melt?	What did the assistant pour?
10	Ship	What did the captain navigate back to the harbour?	What did the pilot crash?	What did the character destroy?
11	Duck	What did the toddler feed the bread to?	Who was the new mum feeding?	What distracted the little girl away from her toys?
12	Water	What did he spill?	What did the baker pour into the jar?	What did the student chill for a few hours?
13	Bottle	What did they store the water in?	What did the commuter drink coffee from?	What did the guy handle?
14	Fireworks	What was the dog frightened of?	What did she light on the dinner table?	What did they put in the box?
15	Bread	What did the baker make with yeast?	What did she eat on her birthday?	What did the pupil make in her spare time?
16	Mountain	What did the alpinist climb?	What did the construction worker climb?	What did they reach?
17	Bandage	What did he wrap around the wound?	What did the midwife wrap the newborn in?	What did the consumer snatch from the man?
18	Horse	What did the jockey ride?	What did she ride to work?	What object did he win in the competition?

19	Calendar	Where did mum write the scheduled family events?	What did the artist illustrate?	What did the lady create?
20	Hair	What did the barber chop?	What did the butcher chop?	What did she inspect vigilantly?
21	Carrot	What did the rabbit eat?	What did the mouse nibble on?	What did the associates purchase for their endeavour?
22	Purse	What did she carry her money in?	What did she keep the cookies in?	What did the onlooker have?
23	Painting	What did the artist admire?	What did the seamstress create?	What did the attendee comment on at the event?
24	Bag	Where did the athlete keep his kit?	What did she put the sleeping baby into?	What did the guest fix at the weekend?
25	Bed	Where did he sleep?	What did the automobile salesman sell?	What did the visitor purchase at the establishment?
26	Egg	What did the chef crack?	What did the baby shake?	What did the father ask his son to bring?
27	Bank	Which institute did the thief rob?	Where did the mum drop off her daughter?	Which organisation did the gentleman contact to discuss the matter?
28	Tenant	Who did the landlord evict?	Who drew the sketch?	Who did the team get in touch with?
29	Beer	What did the student purchase at the bar?	What did grandma sweeten?	What did grandpa discard?
30	Ball	What did the school children play with?	What did the pirate discover?	What did the children share?
31	Tree	What did the squirrel climb?	What was the guard dog sitting by?	What did the committee remove?
32	Rope	What did the sailor knot?	What was the little boy learning to tie?	What did they forget to put in the bag?
33	Wood	What did the carpenter cut?	What did the beautician file?	What did the girl damage?
34	Glue	What did he use to stick the pieces of paper together?	What was the kid chewing?	What did the teenager retrieve from the box?
35	Finger	What did he put the ring on?	What did the gardener cut?	What did the new occupant find?

36	Chimney	What was the smoke coming out of?	What did the maid sweep?	What did the guardian assess?
37	Prisoner	Who did the guard escort back to the cell?	Who did the doctor treat?	Who did the spectator meet?
38	Gardener	Who was trimming the hedge?	Who did the resident alert?	Who did the candidate consult?
39	Tin	What were the baked beans stored in?	What did she ladle the soup into?	What did the parent hand over to the supervisor?
40	Hand	What did the candidate shake at the start of the interview?	What did she put the coat on?	What did she break?

targets = pictures named by participants following each of the corresponding, presented matched mismatched and neutral questions

Table A2 Chapter 2 stimuli set 2

Set 2				
	Targets	**MATCHED**	**MISMATCHED**	**NEUTRAL**
1	Leg	What did the football player break?	What did she wrap before Christmas?	What did grandfather perceive?
2	Banana	What did the monkey eat?	What did the dinner lady mash?	What did the cat examine?
3	Bone	What did the dog lick?	What was the defendant accused of stabbing the victim with?	What was the young man holding?
4	Crown	What did the queen wear?	What do married women wear?	What did the girl lift off the table?
5	Fairy tale	What did the father read to his daughter before bed?	What did the priest read?	What did the man pick up?

6	Oven	What did the kitchen assistant switch on before preparing dinner?	What did she turn off before bed?	What did the man use during the day?
7	Money	What did the thief steal?	What did he give to his date?	What did the girl request?
8	Bucket	What did the little girl put the sand into?	What did they put the rubbish into?	What did the children cleanse?
9	Butter	What did the sweet pastry chef melt?	What did he add to the cold drink?	What did the assistant pour?
10	Aeroplane	What did the pilot crash?	What did the captain navigate back to the harbour?	What did the character destroy?
11	Baby	Who was the new mum feeding?	What did the toddler feed the bread to?	What distracted the little girl away from her toys?
12	Jam	What did the baker pour into the jar?	What did he spill?	What did the student chill for a few hours?
13	Flask	What did the commuter drink coffee from?	What did they store the water in?	What did the guy handle?
14	Candle	What did she light on the dinner table?	What was the dog frightened of?	What did they put in the box?
15	Cake	What did she eat on her birthday?	What did the baker make with yeast?	What did the pupil make in her spare time?
16	Ladder	What did the construction worker climb?	What did the alpinist climb?	What did they reach?
17	Blanket	What did the midwife wrap the newborn in?	What did he wrap around the wound?	What did the consumer snatch from the man?
18	Bike	What did she ride to work?	What did the jockey ride?	What object did he win in the competition?

19	Book	What did the artist illustrate?	Where did mum write the scheduled family events?	What did the lady create?
20	Meat	What did the butcher chop?	What did the barber chop?	What did she inspect vigilantly?
21	Cheese	What did the mouse nibble on?	What did the rabbit eat?	What did the associates purchase for their endeavour?
22	Jar	What did she keep the cookies in?	What did she carry her money in?	What did the onlooker have?
23	Dress	What did the seamstress create?	What did the artist admire?	What did the attendee comment on at the event?
24	Cot	What did she put the sleeping baby into?	Where did the athlete keep his kit?	What did the guest fix at the weekend?
25	Car	What did the automobile salesman sell?	Where did he sleep?	What did the visitor purchase at the establishment?
26	Rattle	What did the baby shake?	What did the chef crack?	What did the father ask his son to bring?
27	School	Where did the mum drop off her daughter?	Which institute did the thief rob?	Which organisation did the gentleman contact to discuss the matter?
28	Artist	Who drew the sketch?	Who did the landlord evict?	Who did the team get in touch with?
29	Tea	What did grandma sweeten?	What did the student purchase at the bar?	What did grandpa discard?
30	Treasure	What did the pirate discover?	What did the school children play with?	What did the children share?
31	Door	What was the guard dog sitting by?	What did the squirrel climb?	What did the committee remove?
32	Laces	What was the little boy learning to tie?	What did the sailor knot?	What did they forget to put in the bag?
33	Nails	What did the beautician file?	What did the carpenter cut?	What did the girl damage?
34	Gum	What was the kid chewing?	What did he use to stick the pieces of paper together?	What did the teenager retrieve from the box?

35	Grass	What did the gardener cut?	What did he put the ring on?	What did the new occupant find?
36	Floor	What did the maid sweep?	What was the smoke coming out of?	What did the guardian assess?
37	Patient	Who did the doctor treat?	Who did the guard escort back to the cell?	Who did the spectator meet?
38	Policeman	Who did the resident alert?	Who was trimming the hedge?	Who did the candidate consult?
39	Bowl	What did she ladle the soup into?	What were the baked beans stored in?	What did the parent hand over to the supervisor?
40	Hook	What did she put the coat on?	What did the candidate shake at the start of the interview?	What did she break?

targets = pictures named by participants following each of the corresponding, presented matched mismatched and neutral questions

Appendix B: Chapter 3 Supplementary graphs

*p< 0.05 (two-tailed); **p< 0.01 level (two-tailed); ***p<.001 (two-tailed).

Synonym judgement

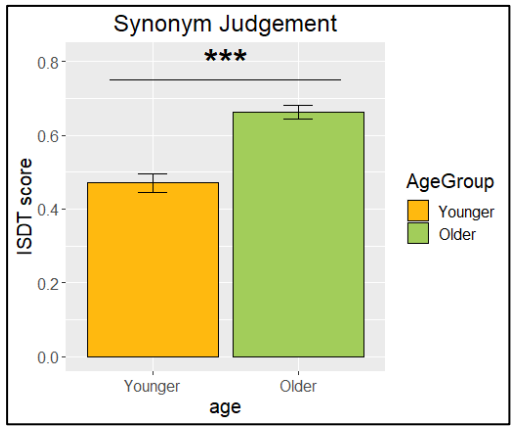

Figure B1: Bar chart displaying the synonym judgement mean ISDT (considering hits and false alarms) score per age group in the synonym judgement task.

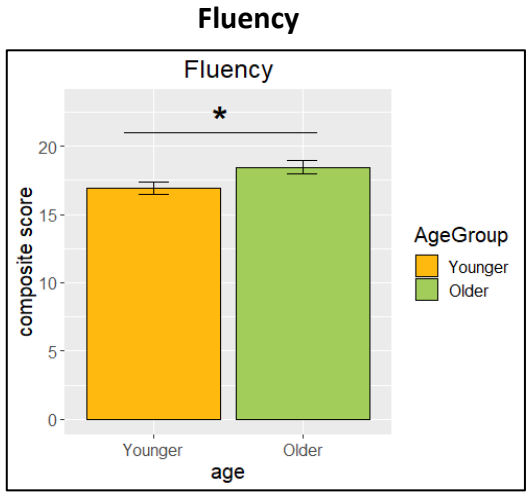

Figure B2: Bar chart displaying the mean fluency composite score per age group in the verbal fluency task.

As pre-registered, the verbal fluency score used in the regression analysis was computed across the two tasks. However, further analyses showed the age-group difference was only significant for the letter fluency task (*M* older adults = 17.6, *SD* = 3.9; *M* younger adults = 13.9, *SD* = 3.3; $t(79) = 4.538$, $p < .001$). There was no significant difference on the semantic fluency task (*M* older adults = 19.2, *SD* = 4.1; *M* younger adults = 20.0, *SD* = 3.9; $t(79) = -0.924$, $p = .358$).

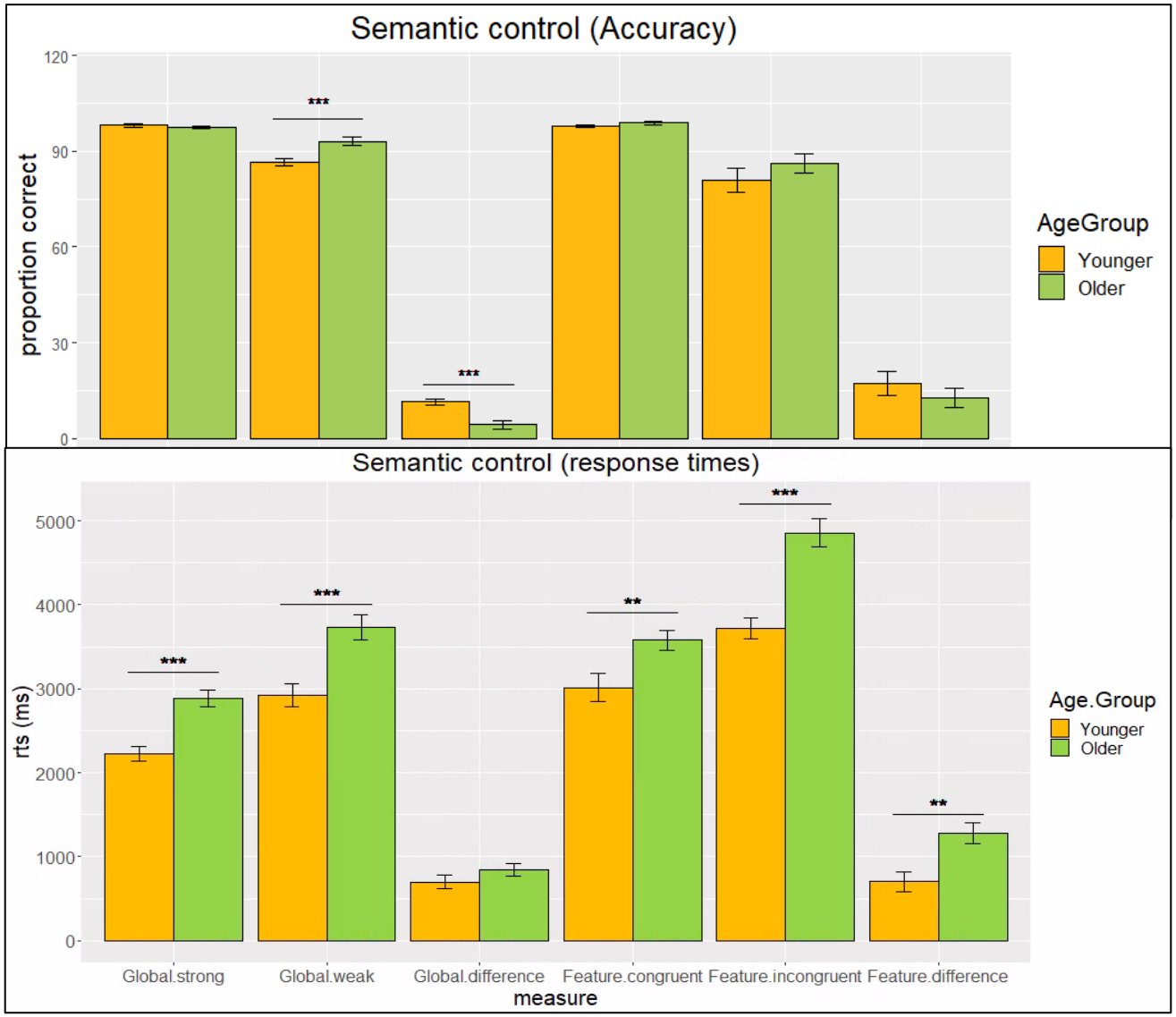

Figure B3: Supplementary Bar chart showing semantic control (accuracy, top; RTs, bottom) in the global association and feature association task. On the left, strong and weak global association trials are shown, with the global difference reflecting the difference between strong and weak trials. On the right, congruent and incongruent feature association trials are shown, with the feature difference reflecting the difference between congruent and incongruent trials.

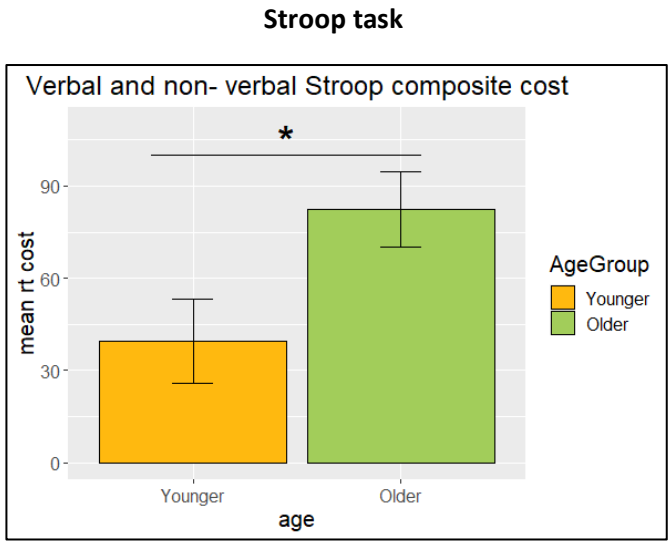

Figure B4: Supplementary Bar charts displaying mean Stroop interference costs (across the verbal and non-verbal tasks) per age group.

In the regression, as pre-registered, we created one Stroop cost across the verbal and non-verbal task. However, further checks showed the age-group difference was only significant in the verbal Stroop task (*M cost* older adults = 99ms, *SD* = 137; *M cost* younger adults = 39ms, *SD* = 104; $t(80) = 2.216$, $p = .030$), with no significant difference in the non-verbal task (*M cost* older adults = 66ms, *SD* = 66; *M cost* younger adults = 40ms, *SD* = 128; $t(81) = 1.173$, $p = .244$)

Digit Span

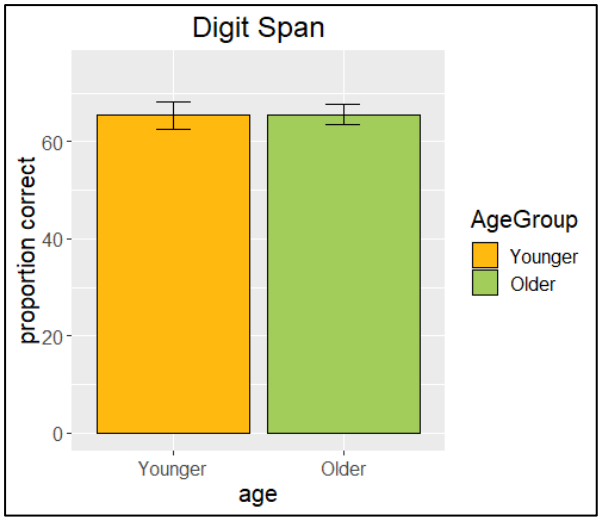

Figure B5: Supplementary Bar charts displaying performance (percentage correct) on the digit span task per age group.

Appendix C: Chapter 4 Stimuli

Stimuli were split into four sets; each participant was allocated to one of these. Each set included one (out of four) of the variations of each basal sentence (corresponding to the conditions: short, prime beginning; short, prime end; long, prime beginning; long, prime end).

Table C1 Chapter 4 stimuli set 1

	SET 1				
	Sentence	**Prime**	**Expected answer**	**Length**	**Prime position**
1	the bride's father walked her down the	bride's	AISLE	short	Beginning
2	sung by the attendees on the queen's arrival, was the	queen's	ANTHEM	short	End
3	Neil Armstrong, who most people know of and many admire, was a very famous and influential	Neil Armstrong	ASTRONAUT	long	Beginning
4	the famous man at the event that was organised by a large company, was asked by his fans for his	famous man	AUTOGRAPH	long	Beginning
5	to determine, considering that he was feeling only mildly unwell, the patient's cause of death, they conducted an	cause of death	AUTOPSY	long	End
6	the sport using a shuttlecock is called	Shuttlecock	BADMINTON	short	End
7	the Neil Armstrong instrument used to play the song was a	Scottish instrument	BAGPIPE	short	Beginning
8	you can buy fresh bread from the	fresh bread	BAKERY	short	End
9	wearing, after her mother had handed it down to her, the tutu was the	tutu	BALLERINA	long	End
10	the monkey was eating the yellow fruit called a	yellow fruit	BANANA	short	End
11	around the wound he wrapped a	wound	BANDAGE	short	Beginning

12	she gripped, before the big opening ceremony that was about to start, while descending the stairs, the wooden	stairs	BANISTER	long	End
13	they used holy water, which was prepared during the morning preparations by the institution, in the ceremony called a	holy water	BAPTISM	long	Beginning
14	even during the winter months, despite the temperature and scarce resources, the troops fought on the	troops fought	BATTLEFIELD	long	End
15	a round Christmas tree decoration is called a	Christmas tree	BAUBLE	short	Beginning
16	the children stayed up, though they struggled to get ready on time every morning, in the evening past their	children stayed up	BEDTIME	long	Beginning
17	on the street, people asking for money are called	asking for money	BEGGARS	short	End
18	they revoked his religious membership because he was not a true	religious	BELIEVER	short	End
19	on her way to work, during the morning while it was raining, she popped a tyre, while riding her	popped a tyre	BIKE	long	End
20	every morning at 5am on the dot, the workers threw the paper waste into the	waste	BIN	long	End
21	the famous snack which i remember my mother keeping in a glass jar when i was growing up, Custard Creams, are a type of	custard creams	BISCUIT	long	End
22	chalk was used by the teachers to write the main messages from the session, on the	chalk	BLACKBOARD	long	Beginning
23	keeping warm the new born was a knitted	keeping warm	BLANKET	short	Beginning
24	his uncomfortable new shoes gave him	uncomfortable new shoes	BLISTERS	short	Beginning
25	famous people are often protected by	Protected	BODYGUARDS	short	End
26	they built on Guy Fawkes' day a	Guy Fawkes'	BONFIRE	Short	End

27	the green, tree-shaped vegetable that my children don't like is called	green, tree-shaped vegetable	BROCCOLI	short	Beginning
28	on her arm, whilst playing outdoors during the summer holidays, she had acquired a tender blue-ish purple spot called a	blueish-purple spot	BRUISE	long	End
29	he sold the meat, after getting the delivery at nine-thirty that morning, as he was a	meat	BUTCHER	long	Beginning
30	she marked important dates on the	important dates	CALENDAR	short	End
31	they rode through the desert on a	desert	CAMEL	short	End
32	school dinner, although many people do not enjoy it too much, can be bought in the place called a	school dinner	CANTEEN	long	Beginning
33	the food-group, to which components of a number of recipes, such as bread and pasta belong, is called	bread and pasta	CARBOHYDRATES	long	End
34	sung by the Christmas choir were	Christmas choir	CAROLS	short	End
35	her potion was boiling in a	potion	CAULDRON	short	Beginning
36	into the bowl of milk, in the morning he poured	bowl of milk	CEREAL	short	Beginning
37	the food item, which he knew he would be able to use in quite a few different recipes, he grated, was	grated	CHEESE	long	End
38	out of the shells, while everyone was watching with great delight, hatched little yellow	hatching	CHICKS	long	End
39	fear of closed, narrow spaces is a psychological condition called	fear of closed, narrow spaces	CLAUSTROPHOBIA	short	Beginning

40	the computer mouse, which is often present in many different environments, makes a type of sound called a	computer mouse	CLICK	long	Beginning
41	his broken leg made him walk on	broken leg	CRUTCHES	short	Beginning
42	ice used in drinks is usually in the shape of	Ice	CUBES	short	Beginning
43	the black and white spotted dog snacking on an apple was a	black and white spotted dog	DALMATION	short	Beginning
44	to understand it better, because were sure they had come across it before, but had never studied it deeply enough, the word was looked up in a	looked up	DICTIONARY	long	End
45	Jurassic Parc is a well-known film about	Jurassic Parc	DINOSAURS	short	Beginning
46	the classic ring-shaped bakery item, which she always enjoyed eating throughout the entire year, she had for dessert, was a	ring shaped bakery item	DOUGHNUT	long	Beginning
47	the duke, due to familial and societal pressure and because he craved true companionship, was marrying the	duke	DUCHESS	long	Beginning
48	the baby sucked, while being irritated, tired and hungry in the pram, on the	baby sucked	DUMMY	long	Beginning
49	the companions in a famous story featuring some compelling characters, which is about Snow White, were seven	Snow White's	DWARFS	long	End
50	the client, who owned her own business through which she had acquired a lot of clients, and had generated a lot of income, had the beautician pluck her	beautician plucked	EYEBROW	long	End
51	"once upon a time" is used at the start of a story called a	“once upon a time”	FAIRYTALE	short	Beginning
52	the palace cooks, just in time to have all required items delivered well in advance, prepared a grand	palace cooks	FEAST	long	Beginning

53	the bonfire night display, which was organised for the locals using the money donated throughout the year, featured colourful	bonfire night	FIREWORKS	long	Beginning
54	the commuter's coffee, which he was not enjoying the taste of today, was kept warm for a few hours by putting it into a	commuter's coffee	FLASK	long	Beginning
55	when walking, pedestrians use the	pedestrians	FOOTPATH	short	End
56	into a container, the brilliant greenish coloured aromatic scented liquid, was poured through a	poured the liquid	FUNNEL	long	End
57	the study of environments is a subject called	study of environments	GEOGRAPHY	short	Beginning
58	the pieces, which were used on the open day, were put together firmly in front of the group with a sticky substance called	sticky substance	GLUE	long	End
59	the bride was kissed, for the first time, by her	bride	GROOM	short	Beginning
60	protecting, because the instructor explained that it needed to be covered properly in case of any accidents, the motorcyclist's head, was a	motorcyclist's	HELMET	long	End
61	a shape with six sides is a	six sided	HEXAGON	short	End
62	the cake, after very skillful preparation by an excellent worker, was covered with a smooth layer of sugary	cake	ICING	long	Beginning
63	sleeplessness is a condition which is also called	sleeplessness	INSOMNIA	short	Beginning
64	another word, which is frequently used to describe people, for "envy" is	envy	JEALOUSY	long	End
65	the denim garment, after carefully considering multiple options that were available to her that day, that she wore on her legs, was a pair of	denim	JEANS	long	Beginning

66	a fruit flavoured dessert which should be kept refrigerated, can be eaten in a variety of different ways, and is wobbly is called	wobbly	JELLY	long	End
67	the dog slept, after a long and eventful day including a walk by the river, in the garden at night, in a	dog slept	KENNEL	long	Beginning
68	with food, people often eat a famous tomato flavoured sauce called	tomato flavoured sauce	KETCHUP	short	End
69	water was boiled for the tea in the	boiled	KETTLE	short	End
70	they typed in the office on a	typed	KEYBOARD	short	Beginning
71	the babies, on a warm and sunny July afternoon, in my friend's back garden, birthed by the cat were	cat	KITTENS	long	End
72	the sleep-inducing purple plant, though some people would say it can have quite an overpowering smell, in her bedroom is called	sleep-inducing purple plant	LAVENDER	long	Beginning
73	expressed through their wedding vows was their	wedding vows	LOVE	short	End
74	the organ damaged by smoking is the	smoking	LUNG	short	End
75	the rabbit and hat trick at the one-off community event, which took place in front of a large audience, was performed by a	rabbit and hat trick	MAGICIAN	long	Beginning
76	relieving tension, which you can do at home or get someone to do professionally for you, through rubbing muscles and joints is called	rubbing muscles	MASSAGE	long	End
77	he sung on stage, in front of a large audience, who had purchased expensive tickets to be admitted into the event, into a	sung	MICROPHONE	long	Beginning

78	the bacteria were studied by magnifying them under a	bacteria	MICROSCOPE	short	beginning
79	the insect bites on their arms came from	insect bites	MOSQUITO	short	beginning
80	restraining the biting dog's jaw was a	dog's jaw	MUZZLE	short	End
81	to hang the criminal, who had committed many offences and shown very little regret for them, they used a	hang	NOOSE	long	beginning
82	the pastry was baked in the	baked	OVEN	Short	End
83	for the study, which focussed on a previously unexplored topic and for which they had quite a strict criteria, they had begun recruiting	study	PARTICIPANTS	long	beginning
84	the oyster shell had inside it beautiful	oyster shell	PEARLS	short	beginning
85	purchasing the camera was a professional	camera	PHOTOGRAPHER	short	End
86	for example, although it is not as large as some of the others in the same category, Pluto is a	Pluto	PLANET	long	End
87	hats were worn, as they were excited to finally see the rewards of their efforts, by the graduates, as were	graduates	ROBES/ GOWNS	long	End
88	to chase away the birds, as he could not think of a better solution to take care of his produce, the farmer used a	chase away the birds	SCARECROW	long	beginning
89	he wore, to keep his neck warm, a knitted	neck	SCARF	short	End
90	they cut out the paper shapes using a pair of	cut out	SCISSORS	short	Beginning
91	in the bowls, which they always presented whenever they had guests over during the summer months, they put ice-cream using a	ice-cream	SCOOP	long	End
92	hair is cleaned, usually a few times a week depending on the person, using water with	hair	SHAMPOO	long	Beginning
93	he washed himself after the run in the	washed himself	SHOWER	short	Beginning

94	the little girl was pulled along in the snow, by the mother, on a	snow	SLEDGE	short	End
95	pencils and paper, which were needed to take part in the activities they had signed up for, were bought from the shop selling	pencils	STATIONERY	long	Beginning
96	the transparent tube the freshly drawn blood was in, was a	blood	SYRINGE	short	End
97	to observe his study object, the astronomer looked in the distance through a	astronomer	TELESCOPE	short	End
98	the doctor, as it had been fluctuating since the last meeting, measured the temperature with a	temperature	THERMOMETER	long	End
99	the gym goers were running at a steady pace on the	gym-goers	TREADMILL	short	beginning
100	the crowned champion, who was rather tired and desperately needed a shower, held in his hand a	crowned-champion	TROPHY	long	beginning
101	she applied for a tourist	tourist	VISA	short	End

sentence = sentences missing their final word which were presented to the participant (auditorily), prime = the word(s) within the sentences which were assumed to prime the sentence ending, expected ending= the endings primed by the sentences which we expected participants to produce, length = sentence length (short or long), prime position = whether the semantic prime was placed at the beginning or the end of the sentence.

Table C2 Chapter 4 stimuli set 2

	SET 2				
	Sentence	**Prime**	**Expected answer**	**Length**	**Prime position**
1	the father walked the bride down the	bride's	AISLE	short	End

2	on the queen’s arrival, as always at the events she was attending, the attendees sung the	queen's	ANTHEM	long	Beginning
3	the very famous man, who most people know of and many admire, Neil Armstrong was an influential	Neil Armstrong	ASTRONAUT	long	End
4	his fans at the event that was organised by a large company, asked the famous man for his	famous man	AUTOGRAPH	long	End
5	cause of death of the patient was determined through an	cause of death	AUTOPSY	short	Beginning
6	a shuttlecock, which can be used by players of all levels, is used in a sport called	shuttlecock	BADMINTON	long	Beginning
7	they played the song on the Scottish instrument called a	Scottish instrument	BAGPIPE	short	End
8	fresh bread, together with various other items that are popular at lunch time, can be bought from the	fresh bread	BAKERY	long	Beginning
9	the tutu was worn by the	tutu	BALLERINA	short	Beginning
10	the yellow fruit, prepared by the man with the hat in the morning, and eaten by the monkey, was a	yellow fruit	BANANA	long	Beginning
11	wrapped around the wound was a	wound	BANDAGE	short	End
12	while descending the stairs, she gripped the wooden	stairs	BANISTER	short	Beginning
13	the ceremony, for which the institution prepared during the morning preparations the holy water, was called a	holy water	BAPTISM	long	End
14	the troops fought even during the winter months, on the	troops fought	BATTLEFIELD	short	Beginning
15	the round decoration found on Christmas trees is a	Christmas tree	BAUBLE	short	End
16	in the evening, though they struggled to get ready on time every morning, the children stayed up past their	children stayed up	BEDTIME	long	End
17	people asking for money, who may also benefit from other types of support, on the street, are called	asking for money	BEGGARS	long	Beginning

18	his religious membership in the organisation, which had been founded by a very strict leader, who was intolerant of the members' mistakes, was revoked, because he was not a true	religious	BELIEVER	long	Beginning
19	she popped a tyre on her way to work, while riding her	popped a tyre	BIKE	short	Beginning
20	paper waste was thrown into the recycling	waste	BIN	short	Beginning
21	Custard Creams are a famous snack which are a type of	custard creams	BISCUIT	short	Beginning
22	to write the main messages from the session, the teachers used chalk on the	chalk	BLACKBOARD	long	End
23	the newborn was kept warm by a knitted	keeping warm	BLANKET	short	End
24	wearing the uncomfortable new shoes gave him	uncomfortable new shoes	BLISTERS	short	End
25	protection of famous people, like members of the royal family, who may be high risk, is often done by	protection	BODYGUARDS	long	Beginning
26	on Guy Fawkes' day, which was seen as a reason to get together with friends and family, and enjoy their company for a night, they built a	Guy Fawkes'	BONFIRE	long	Beginning
27	my children don't like the green, tree-shaped vegetable called	green, tree-shaped vegetable	BROCCOLI	short	End
28	the tender blueish-purple spot on her arm was a	blueish-purple spot	BRUISE	short	Beginning
29	the seller, after getting the delivery at nine thirty that morning, of the meat was a	meat	BUTCHER	long	End
30	important dates, although her husband john truthfully couldn't care less about them, were marked on the	important dates	CALENDAR	long	Beginning

31	through the desert, they rode, for a very long time and during an extremely hot day, on a	desert	CAMEL	long	Beginning
32	where you buy, although many people do not enjoy it too much, school dinner from, is called a	school dinner	CANTEEN	long	End
33	bread and pasta belong to the food-group called	bread and pasta	CARBOHYDRATES	short	beginning
34	the Christmas choir, during the evening when the entire community came together, were singing	Christmas choir	CAROLS	long	beginning
35	she boiled her potion in a	potion	CAULDRON	short	End
36	in the morning, he poured into the bowl of milk some	bowl of milk	CEREAL	short	End
37	he grated the food item called	grated	CHEESE	short	beginning
38	hatching out of the shells were little yellow	hatching	CHICKS	short	beginning
39	the psychological condition where people fear closed, narrow spaces is called	fear of closed, narrow spaces	CLAUSTROPHOBIA	short	End
40	the type of sound, which is often present in many different environments, made by the computer mouse, is a	computer mouse	CLICK	long	End
41	he walked on his broken leg with	broken leg	CRUTCHES	short	End
42	drinks oftain contain ice in the shape of	ice	CUBES	short	End
43	snacking on an apple was a black and white spotted dog called a	black and white spotted dog	DALMATION	short	End
44	they looked up the word to understand it better, in a	looked up	DICTIONARY	short	beginning
45	the well-known film Jurassic Parc is about	Jurassic Parc	DINOSAURS	short	End
46	she had for dessert, which she always enjoyed eating throughout the entire year, the classic ring-shaped bakery item called a	ring shaped bakery item	DOUGHNUT	long	End

47	marrying, due to familial and societal pressure, and because she craved true companionship, the duke was the	duke	DUCHESS	long	end
48	in the pram, the irritated, tired and hungry baby sucked on the	baby sucked	DUMMY	long	End
49	Snow White's companions were seven	Snow White's	DWARFS	short	beginning
50	the beautician plucked the client's	beautician plucked	EYEBROW	short	beginning
51	a story that starts with "once upon a time" is called a	"once upon a time"	FAIRYTALE	short	End
52	prepared, just in time to have all required items delivered well in advance, by the palace cooks, was a grand	palace cooks	FEAST	long	End
53	the display, which was organised for the locals using the money donated throughout the year, on bonfire night featured colourful	bonfire night	FIREWORKS	long	End
54	for a few hours, though he was not enjoying the taste of it today, the commuter's coffee was kept warm, by putting it into a	commuter's coffee	FLASK	long	End
55	pedestrians, who have just come out of the university and are heading home, walk on the	pedestrians	FOOTPATH	long	beginning
56	he poured the liquid into a container through a	poured the liquid	FUNNEL	short	beginning
57	the subject involving studying environments is called	study of environments	GEOGRAPHY	short	End
58	they used a sticky substance to put the pieces firmly together called	sticky substance	GLUE	short	beginning
59	for the first time, the bride was kissed by her	bride	GROOM	short	End
60	the motorcyclist's head was protected by a	motorcyclist's	HELMET	short	beginning

61	a six sided shape, which features in various drawings and pictures, is a	six sided	HEXAGON	long	beginning
62	covering, after very skillful preparation by an excellent worker, the cake, was a smooth layer of sugary	cake	ICING	long	End
63	a condition where people are sleepless is also called	sleeplessness	INSOMNIA	short	End
64	"envy" is another word for	envy	JEALOUSY	short	beginning
65	on her legs, after carefully considering multiple options that were available to her that day, she wore a garment made of denim called a pair of	denim	JEANS	long	end
66	a wobbly dessert which is fruit flavoured is called	wobbly	JELLY	short	beginning
67	at night, after a long and eventful day including a walk by the river, the dog slept in the garden, in a	dog slept	KENNEL	long	End
68	a famous tomato flavoured sauce, that some people really enjoy, while others complain is too salty, which is often eaten with food, is called	tomato flavoured sauce	KETCHUP	long	beginning
69	she boiled water for the tea, which was consumed by herself and the rest of the team during the day, in the	boiled	KETTLE	long	beginning
70	in the office, they typed on a	typed	KEYBOARD	short	End
71	the cat birthed lots of baby	cat	KITTENS	short	beginning
72	in her bedroom, though some people would say it can have quite an overpowering smell, the sleep-inducing purple plant is called	sleep-inducing purple plant	LAVENDER	long	End
73	through their wedding vows, while feeling nervous and excited, they expressed their	wedding vows	LOVE	long	beginning
74	smoking damages the organ, which plays a pivotal role in the organism's functioning, called the	smoking	LUNGS	long	Beginning

75	performing at the one-off community event, which took place in front of a large audience, the trick with the rabbit and hat, was a	rabbit and hat trick	MAGICIAN	long	End
76	rubbing muscles and joints to relieve tension is called	rubbing muscles	MASSAGE	short	Beginning
77	on stage, in front of a large audience, who had purchased expensive tickets to be admitted into the event, he sung into a	sung	MICROPHONE	long	End
78	to study them, the bacteria were magnified under a	bacteria	MICROSCOPE	short	End
79	on their arms were insect bites from	insect bites	MOSQUITO	short	End
80	the biting dog's jaw, as people were starting to feel threatened and thus started to leave, was restrained with a	dog's jaw	MUZZLE	long	beginning
81	the criminal, who had committed many offences and showed very little regret, was hung using a	hang	NOOSE	long	end
82	she baked the pastry, which she decided she would add other things to later, in the	baked	OVEN	long	beginning
83	recruitment was begun, while they were focussing on a previously unexplored topic, for which they had quite a strict criteria, for the study's	study	PARTICIPANTS	long	End
84	inside the oyster shell were beautiful	oyster shell	PEARLS	short	End
85	the camera, which was brand new, of very high quality, and was rather expensive, was purchased by a professional	camera	PHOTOGRAPHER	long	beginning
86	Pluto is an example of a	Pluto	PLANET	short	beginning
87	the graduates wore hats as well as	graduates	ROBES/ GOWNS	short	beginning
88	the farmer, as he could not think of a better solution to take care of his produce, chases away the birds using a	chase away the birds	SCARECROW	long	End

89	to keep his neck warm, because of the changes in temperature, he wore a knitted	neck	SCARF	long	beginning
90	the paper shapes were cut out using a pair of	cut out	SCISSORS	short	End
91	ice-cream was put in the bowl using a	ice-cream	SCOOP	short	beginning
92	to clean, usually a few times a week, depending on the person, hair, you use water and	hair	SHAMPOO	long	End
93	after the run, he washed himself in the	washed himself	SHOWER	short	End
94	in the snow, during the middle of the day whilst most of city's residents were nowhere to be seen, the mother pulled along the little girl on a	snow	SLEDGE	long	beginning
95	they bought, as they were needed to take part in the activities they had signed up for, pencils and paper, from the shop selling	pencils	STATIONERY	long	End
96	the freshly drawn blood, which needed to be sent to the external team, to be examined by their staff, was currently in the transparent tube called a	blood	SYRINGE	long	beginning
97	the astronomer, who was out in his garden at 2am while most people were sleeping, was observing his study object in the distance, through a	astronomer	TELESCOPE	long	beginning
98	the temperature was measured by the doctor with a	temperature	THERMOMETER	short	beginning
99	at a steady pace, the gym-goers were running on the	gym-goers	TREADMILL	short	End
100	in his hand, though he was rather tired and desperately needed a shower, the crowned champion held a	crowned-champion	TROPHY	long	End
101	the tourist, who hadn’t been able to do this for a few years due to family responsibilities, applied for a	tourist	VISA	long	beginning

sentence = sentences missing their final word which were presented to the participant (auditorily), prime = the word(s) within the sentences which were assumed to prime the sentence ending, expected ending= the endings primed by the sentences which we expected participants to produce, length = sentence length (short or long), prime position = whether the semantic prime was placed at the beginning or the end of the sentence.

Table C3 Chapter 4 stimuli set 3

	SET 3				
	Sentence	**Prime**	**Expected answer**	**Length**	**Prime position**
1	the bride's father, who was extremely proud and felt particularly emotional today, walked her down the	bride's	AISLE	long	Beginning
2	sung by the attendees, as always at the events she was attending, on the queen's arrival, was the	queen's	ANTHEM	long	End
3	Neil Armstrong was a very famous and influential	Neil Armstrong	ASTRONAUT	short	Beginning
4	the famous man was asked by his fans for his	famous man	AUTOGRAPH	short	Beginning
5	to determine the patient's cause of death, they conducted an	cause of death	AUTOPSY	short	End
6	the sport, which can be played by players of all levels, using a shuttlecock, is called	shuttlecock	BADMINTON	long	End
7	the Scottish instrument, which is part of the country's heritage and beloved by many, used to play the song, was a	Scottish instrument	BAGPIPE	long	Beginning
8	you can buy, together with various other items that are popular at lunch time, fresh bread from the	fresh bread	BAKERY	long	End
9	wearing the tutu was the	tutu	BALLERINA	short	End

10	the monkey, after the man with the hat prepared it in the morning, was eating a yellow fruit called a	yellow fruit	BANANA	long	End
11	around the wound, the man who had just recently arrived back home, wrapped a	wound	BANDAGE	long	Beginning
12	she gripped, while descending the stairs, the wooden	stairs	BANISTER	short	End
13	they used holy water in the ceremony called a	holy water	BAPTISM	short	Beginning
14	even during the winter months, the troops fought on the	troops fought	BATTLEFIELD	short	End
15	a round Christmas tree decoration, which is sold in a variety of different colours and is often sparkly, is called a	Christmas tree	BAUBLE	long	Beginning
16	the children stayed up in the evening, past their	children stayed up	BEDTIME	short	Beginning
17	on the street, people, who may also benefit from other types of support, asking for money are called	asking for money	BEGGARS	long	End
18	they revoked from the organisation, which had been founded by a very strict leader who was intolerant of the member's mistakes, his religious membership, because he was not a true	religious	BELIEVER	long	End
19	on her way to work, she popped a tyre while riding her	popped a tyre	BIKE	short	End
20	they threw the paper waste into the recycling	waste	BIN	short	End
21	the famous snack Custard Creams are a type of	custard creams	BISCUIT	short	End
22	chalk was used by the teachers to write on the	chalk	BLACKBOARD	short	Beginning
23	keeping warm the new born, who was delivered healthily after a relatively short period of labour, was a knitted	keeping warm	BLANKET	long	beginning
24	his uncomfortable new shoes, after they had been gifted to him by his family on his last birthday, gave him	uncomfortable new shoes	BLISTERS	long	Beginning

25	famous people, like members of the royal family who may be high risk, are often protected by	protection	BODYGUARDS	long	End
26	they built, which was seen as a reason to get together with friends and family, and enjoy their company for a night, on Guy Fawkes day a	Guy Fawkes'	BONFIRE	long	End
27	the green, tree-shaped vegetable, which I try to prepare each day, but that my children don't like, is called	green, tree-shaped vegetable	BROCCOLI	long	Beginning
28	on her arm, the tender blueish-purple spot was a	blueish-purple spot	BRUISE	short	End
29	he sold meat as he was a	meat	BUTCHER	short	Beginning
30	she marked, although her husband John truthfully couldn't care less about them, the important dates on the	important dates	CALENDAR	long	end
31	they rode for a very long time, and during an extremely hot day, through the desert on a	desert	CAMEL	long	End
32	school dinner can be bought in the place called a	school dinner	CANTEEN	short	Beginning
33	the food-group to which bread and pasta belong is	bread and pasta	CARBOHYDRATES	short	End
34	sung, during the evening when the entire community came together, by the Christmas choir were	Christmas choir	CAROLS	long	End
35	her potion, which she was constantly watching carefully, was boiling in a	potion	CAULDRON	long	Beginning
36	into the bowl of milk, even though he was aware that he should probably have something else in the morning, he poured	bowl of milk	CEREAL	long	Beginning
37	the food item he grated was	grated	CHEESE	short	End
38	out of the shells hatched little yellow	hatching	CHICKS	short	End

39	fear of closed, narrow spaces, which 12.5 percent of the population suffers from, and which can be tackled through therapeutic intervention, is a psychological condition called	fear of closed, narrow spaces	CLAUSTROPHOBIA	long	Beginning
40	the computer mouse makes a type of sound called a	computer mouse	CLICK	short	Beginning
41	his broken leg, which caused him some hassle, and stopped him from doing some of the things he wanted to do, made him walk with	broken leg	CRUTCHES	long	Beginning
42	ice, which you can make at home, or can be bought from the shop, when used in drinks, is usually in the shape of	ice	CUBES	long	Beginning
43	the black and white spotted dog snacking on an apple, that he stole from the fruit bowl when no one was watching, was a	black and white spotted dog	DALMATION	long	beginning
44	to understand it better, the word was looked up in a	looked up	DICTIONARY	short	End
45	Jurassic Parc , a well known film, starring many famous actors, which I went to watch in the cinema when it first came out, is about	Jurassic Parc	DINOSAURS	long	Beginning
46	the classic, ring-shaped bakery item she had for dessert was a	ring shaped bakery item	DOUGHNUT	short	Beginning
47	the duke was marrying the	duke	DUCHESS	short	Beginning
48	the baby sucked in the pram, on the	baby sucked	DUMMY	short	Beginning
49	the companions of Snow White were seven	Snow White's	DWARFS	short	End
50	the client had the beautician pluck her	beautician plucked	EYEBROW	short	End
51	"once upon a time" is a classic phrase used by many famous authors writing this type of work, at the start of a story called a	"once upon a time"	FAIRYTALE	long	Beginning

52	the palace cooks prepared a grand	palace cooks	FEAST	short	Beginning
53	the bonfire night display featured colourful	bonfire night	FIREWORKS	short	Beginning
54	the commuter's coffee was kept warm for a few hours by putting it into a	commuter's coffee	FLASK	short	Beginning
55	when walking, after just coming out of the university and heading home, pedestrians use the	pedestrians	FOOTPATH	long	End
56	into the container the liquid was poured through a	poured the liquid	FUNNEL	short	End
57	the study of environments, which a lot of students find quite interesting, is a subject called	study of environments	GEOGRAPHY	long	Beginning
58	the pieces were firmly put together with a sticky substance called	sticky substance	GLUE	short	End
59	the bride, who had spent months preparing with her closest confidants, and was feeling a little anxious now that it was finally happening, was kissed for the first time by her	bride	GROOM	long	beginning
60	protecting the motorcyclist's head was a	motorcyclist's	HELMET	short	End
61	a shape, which features in various drawings and pictures, with six sides, is a	six sided	HEXAGON	long	End
62	the cake was covered with a layer of sugary	cake	ICING	short	Beginning
63	sleeplessness, something that a lot of people suffer from, and is often attributed to the stressors of modern life, is a condition which is also called	sleeplessness	INSOMNIA	long	Beginning
64	another word for "envy" is	envy	JEALOUSY	short	End
65	the denim garment she wore on her legs was a pair of	denim	JEANS	short	Beginning
66	a fruit flavoured dessert which is wobbly is called	wobbly	JELLY	short	End
67	the dog slept in the garden at night in a	dog slept	KENNEL	short	Beginning

68	with food, some people often really enjoy eating, while others complain is too salty, a famous tomato flavoured sauce called	tomato flavoured sauce	KETCHUP	long	end
69	water, which was consumed by herself and the rest of the team, during the day, was boiled for the tea in the	boiled	KETTLE	long	End
70	they typed in the office, which was quiet despite it being a weekday, on a	typed	KEYBOARD	long	Beginning
71	the babies birthed by the cat were	cat	KITTENS	short	End
72	the sleep-inducing purple plant in her bedroom is called	sleep-inducing purple plant	LAVENDER	short	Beginning
73	expressed, while feeling nervous and excited, through their wedding vows was their	wedding vows	LOVE	long	End
74	the organ which plays a pivotal role in the organism's functioning, damaged by smoking is the	smoking	LUNG	long	End
75	the rabbit and hat trick was performed by a	rabbit and hat trick	MAGICIAN	short	Beginning
76	relieving tension through rubbing muscles and joints is called	rubbing muscles	MASSAGE	short	End
77	he sung on stage into a	sung	MICROPHONE	short	Beginning
78	the bacteria, which the team leader was very interested in, and instructed all of his team to start working on immediately, were studied by magnifying them under a	bacteria	MICROSCOPE	long	Beginning
79	the insect bites on their arms which were smothered with cream and then covered up by the consultant, came from	insect bites	MOSQUITO	long	Beginning
80	restraining, as people were starting to feel threatened and thus started to leave, the biting dog's jaw was a	dog's jaw	MUZZLE	long	End

81	to hang the criminal, they used a	hang	NOOSE	short	Beginning
82	the pastry, which she decided she would add other things to later, was baked in the	baked	OVEN	long	End
83	for the study, they had begun recruiting	study	PARTICIPANTS	short	Beginning
84	the oyster shell, which the little boy had accidentally found whilst on holiday with his parents, had inside it beautiful	oyster shell	PEARLS	long	Beginning
85	purchasing the brand new, very high quality, and rather expensive camera was a professional	camera	PHOTOGRAPHER	long	End
86	for example, Pluto is a	Pluto	PLANET	short	End
87	hats were worn by the graduates, as were	graduates	ROBES/ GOWNS	short	End
88	to chase away the birds, the farmer used a	chase away the birds	SCARECROW	short	Beginning
89	he wore, because of the change in temperature, and to keep his neck warm, a knitted	neck	SCARF	long	End
90	they cut out the paper shapes, which they intended to use for the upcoming project they anticipated would extend over the course of the summer months, using a pair of	cut out	SCISSORS	long	Beginning
91	in the bowls, they put ice-cream using a	ice-cream	SCOOP	short	End
92	hair is cleaned with	hair	SHAMPOO	short	Beginning
93	he washed himself after the run, as it had been a few days since he had last done so, in the	washed himself	SHOWER	long	Beginning
94	the little girl, during the middle of the day whilst most of the city's residents were nowhere to be seen, was pulled along in the snow by the mother on a	snow	SLEDGE	long	End
95	pencils and paper were bought from the shop selling	pencils	STATIONERY	short	Beginning

96	the transparent tube, which needed to be sent to the external team to be examined by their staff, which the freshly drawn blood was in, was a	blood	SYRINGE	long	End
97	to observe his study object from his garden at 2am, while most people were sleeping, the astronomer looked in the distance through a	astronomer	TELESCOPE	long	End
98	the doctor measured the temperature with a	temperature	THERMOMETER	short	End
99	the gym-goers, after considering the range of classes they could access during opening hours, were running at a steady pace on the	gym-goers	TREADMILL	long	beginning
100	the crowned champion held in his hand a	crowned-champion	TROPHY	short	beginning
101	she applied, even though she hadn't been able to do this for a few years due to family responsibilities, for a tourist	Tourist	VISA	long	End

sentence = sentences missing their final word which were presented to the participant (auditorily), prime = the word(s) within the sentences which were assumed to prime the sentence ending, expected ending= the endings primed by the sentences which we expected participants to produce, length = sentence length (short or long), prime position = whether the semantic prime was placed at the beginning or the end of the sentence.

Table C4 Chapter 4 stimuli set 4

	SET 4				
	Sentence	**Prime**	**Expected answer**	**Length**	**Prime position**
1	the father, who was extremely proud and felt particularly emotional today, walked the bride down the	bride's	AISLE	long	End

2	on the queen’s arrival, the attendees sung the	queen's	ANTHEM	short	beginning
3	the very famous man Neil Armstrong was an influential	Neil Armstrong	ASTRONAUT	short	End
4	his fans asked the famous man for his	famous man	AUTOGRAPH	short	End
5	cause of death of the patient, considering that he was feeling only mildly unwell, was determined through an	cause of death	AUTOPSY	long	Beginning
6	a shuttlecock is used in a sport called	shuttlecock	BADMINTON	short	Beginning
7	they played the song, which is part of the country's heritage and beloved by many, on the Scottish instrument called a	Scottish instrument	BAGPIPE	long	End
8	fresh bread can be bought from the	fresh bread	BAKERY	short	Beginning
9	the tutu which her mother had handed down to her, was worn by the	tutu	BALLERINA	long	Beginning
10	the yellow fruit eaten by the monkey was a	yellow fruit	BANANA	short	Beginning
11	wrapped by the man who had recently arrived back home, around the wound, was a	wound	BANDAGE	long	End
12	while descending the stairs before the big opening ceremony that was about to start, she gripped the wooden	stairs	BANISTER	long	Beginning
13	the ceremony they used holy water in is called a	holy water	BAPTISM	short	End
14	the troops fought even during the winter months despite the temperature and scarce resources, on the	troops fought	BATTLEFIELD	long	Beginning
15	a round decoration, which is sold in a variety of different colours and is often sparkly, found on Christmas trees is a	Christmas tree	BAUBLE	long	End
16	in the evening, the children stayed up past their	children stayed up	BEDTIME	short	End
17	people asking for money on the street are called	asking for money	BEGGARS	short	Beginning

18	his religious membership was revoked because he was not a true	religious	BELIEVER	short	Beginning
19	she popped a tyre, during the morning while it was raining, on her way to work, while riding her	popped a tyre	BIKE	long	Beginning
20	the paper waste was thrown every morning at 5am on the dot, by the workers into the	waste	BIN	long	Beginning
21	Custard Creams are a famous snack, which I remember my mother keeping in a glass jar when I was growing up, which are a type of	Custard Creams	BISCUIT	long	Beginning
22	to write, the teachers used chalk on the	chalk	BLACKBOARD	short	End
23	the newborn, who was delivered healthily after a relatively short period of labour, was kept warm by a knitted	keeping warm	BLANKET	long	End
24	wearing, after they had been gifted to him by his family on his last birthday, his uncomfortable new shoes, gave him	uncomfortable new shoes	BLISTERS	long	End
25	protection of famous people is often done by	protection	BODYGUARDS	short	Beginning
26	on Guy Fawkes' day, they built a	Guy Fawkes'	BONFIRE	short	Beginning
27	my children don't like, although I try to prepare it each day, the green, tree-shaped vegetable called	green, tree-shaped vegetable	BROCCOLI	long	End
28	the tender blueish-purple spot, which she had acquired whilst playing outdoors during the summer holidays, on her arm, was a	blueish-purple spot	BRUISE	long	Beginning
29	the seller of the meat was a	meat	BUTCHER	short	End
30	important dates were marked on the	important dates	CALENDAR	short	Beginning
31	through the desert they rode on a	desert	CAMEL	short	Beginning
32	where you buy school dinner from is called a	school dinner	CANTEEN	short	End

33	bread and pasta, which are key components of a number of recipes, belong to the food-group called	bread and pasta	CARBOHYDRATES	long	Beginning
34	the Christmas choir were singing	Christmas choir	CAROLS	short	Beginning
35	she boiled, while she was constantly watching carefully, her potion in a	potion	CAULDRON	long	End
36	in the morning, even though he was aware that he should probably have something else, he poured into the bowl of milk some	bowl of milk	CEREAL	long	End
37	he grated the food item, which he knew he would be able to use in quite a few different recipes, called	grated	CHEESE	long	beginning
38	hatching, while everyone was watching with great delight, out of the shells were little yellow	hatching	CHICKS	long	Beginning
39	the psychological condition which 12.5 percent of the population suffers from, and which can be tackled through therapeutic intervention, where people fear closed, narrow spaces is called	fear of closed, narrow spaces	CLAUSTROPHOBIA	long	End
40	the type of sound made by the computer mouse is a	computer mouse	CLICK	short	End
41	he walked, which caused him some hassle, and stopped him from doing some of the things he wanted to do, on his broken leg with	broken leg	CRUTCHES	long	End
42	drinks, which you can make at home or buy from the shop, often contain ice in the shape of	ice	CUBES	long	End

43	snacking on an apple, that he stole from the fruit bowl when no one was watching, was a black and white spotted dog called a	black and white spotted dog	DALMATION	long	End
44	they looked up the word which they were sure they had come across before, but had never studied deeply enough to understand it, in a	looked up	DICTIONARY	long	Beginning
45	the well-known film, starring many famous actors, which I went to watch in the cinema when it first came out, Jurassic Parc is about	Jurassic Parc	DINOSAURS	long	End
46	she had for dessert a classic, ring-shaped bakery item called a	ring shaped bakery item	DOUGHNUT	short	End
47	marrying the duke was the	Duke	DUCHESS	short	End
48	in the pram, the baby sucked on the	baby sucked	DUMMY	short	End
49	Snow White, a famous story featuring some compelling characters, is about a girl whose companions were seven	Snow White's	DWARFS	long	Beginning
50	the beautician plucked, as part of the business she owned, through which she had acquired a lot of clients and had generated a lot of income, the client's	beautician plucked	EYEBROW	long	Beginning
51	a story written by famous authors of this type of work, that starts with the classic phrase "once upon a time" is called a	"once upon a time"	FAIRYTALE	long	End
52	prepared by the palace cooks was a grand	palace cooks	FEAST	short	End
53	the display on bonfire night featured colourful	bonfire night	FIREWORKS	short	End
54	for a few hours, the commuter's coffee was kept warm by putting it into a	commuter's coffee	FLASK	short	End
55	pedestrians walk on the	pedestrians	FOOTPATH	short	Beginning

56	he poured the liquid, which had a brilliant greenish colour and an aromatic scent, into a container through a	poured the liquid	FUNNEL	long	Beginning
57	the subject which a lot of students find quite interesting, involving studying environments is called	study of environments	GEOGRAPHY	long	End
58	they used a sticky substance, which was used on the open day to put the pieces firmly together in front of the group, called	sticky substance	GLUE	long	Beginning
59	for the first time, after months of preparing with her closest confidants, and while feeling a little anxious now that it was finally happening, the bride was kissed by her	bride	GROOM	long	End
60	the motorcyclist's head, which the instructor explained needed to be covered properly in case of any accidents, was protected by a	motorcyclist's	HELMET	long	Beginning
61	a six-sided shape is a	six sided	HEXAGON	short	Beginning
62	covering the cake was a layer of sugary	cake	ICING	short	End
63	a condition which a lot of people suffer from and is often attributed to the stressors of modern life, where people are sleepless is also called	sleeplessness	INSOMNIA	long	End
64	"envy", which is frequently used to describe people, is another word for	envy	JEALOUSY	long	Beginning
65	on her legs, she wore a garment made of denim called a pair of	denim	JEANS	short	End
66	a wobbly dessert which is fruit flavoured, should be kept refrigerated and can be eaten in a variety of different ways is called	wobbly	JELLY	long	Beginning
67	at night, the dog slept in the garden in a	dog slept	KENNEL	short	End

68	a famous tomato flavoured sauce which is often eaten with food is called	tomato flavoured sauce	KETCHUP	short	Beginning
69	she boiled water for the tea in the	boiled	KETTLE	short	Beginning
70	in the office, which was quiet despite it being a weekday, they typed on a	typed	KEYBOARD	long	End
71	the cat, on a warm and sunny July afternoon, in my friend's back garden, birthed lots of baby	cat	KITTENS	long	Beginning
72	in her bedroom, the sleep-inducing purple plant is called	sleep-inducing purple plant	LAVENDER	short	End
73	through their wedding vows they expressed their	wedding vows	LOVE	short	beginning
74	smoking damages the organ called the LUNGS	smoking	LUNGS	short	Beginning
75	performing the trick with the rabbit and hat was a	rabbit and hat trick	MAGICIAN	short	End
76	rubbing muscles and joints, which you can do at home or get someone to do professionally for you, to relieve tension, is called	rubbing muscles	MASSAGE	long	Beginning
77	on stage, he sung into a	sung	MICROPHONE	short	End
78	to study them, because the team leader was very interested, and had instructed all of his team to start working immediately, the bacteria were magnified under a	bacteria	MICROSCOPE	long	End
79	on their arms which were smothered with cream and covered up by the consultant, were insect bites from	insect bites	MOSQUITO	long	End
80	the biting dog's jaw was restrained with a	dog's jaw	MUZZLE	short	Beginning
81	the criminal was hung using a	Hang	NOOSE	short	End
82	she baked the pastry in the	Baked	OVEN	short	Beginning

83	recruitment was begun for the study's	Study	PARTICIPANTS	short	End
84	inside, after the little boy had accidentally found it whilst on holiday with his parents, the oyster shell, were beautiful	oyster shell	PEARLS	long	End
85	the camera was purchased by a professional	camera	PHOTOGRAPHER	short	Beginning
86	Pluto, although it is not as large as some of the others in the same category, is an example of a	Pluto	PLANET	long	Beginning
87	the graduates, who were excited to finally see the rewards of their efforts, wore hats as well as	graduates	ROBES/ GOWNS	long	Beginning
88	the farmer chases away the birds using a	chase away the birds	SCARECROW	short	End
89	to keep his neck warm he wore a knitted	neck	SCARF	short	beginning
90	the paper shapes, which they intended to use for the upcoming project they anticipated would extend over the course of the summer months, were cut out using a pair of	cut out	SCISSORS	long	End
91	ice-cream, which they always presented whenever they had guests over during the summer months, was put in the bowl using a	ice-cream	SCOOP	long	Beginning
92	to clean hair, you use	hair	SHAMPOO	short	End
93	after the run, as it had been a few days since he had last done so, he washed himself in the	washed himself	SHOWER	long	End
94	in the snow, the mother pulled along the little girl on a	snow	SLEDGE	short	Beginning
95	they bought pencils and paper from the shop selling	pencils	STATIONERY	short	End
96	the freshly drawn blood was in the transparent tube called a	blood	SYRINGE	short	Beginning
97	the astronomer was observing his study object in the distance through a	astronomer	TELESCOPE	short	Beginning

98	the temperature, which had been fluctuating since the last meeting, was measured by the doctor with a	temperature	THERMOMETER	long	Beginning
99	at a steady pace, after considering the range of classes they could access during opening hours, the gym-goers were running on the	gym-goers	TREADMILL	long	End
100	in his hand, the crowned champion held a	crowned-champion	TROPHY	short	End
101	the tourist applied for a	tourist	VISA	short	beginning

sentence = sentences missing their final word which were presented to the participant (auditorily), prime = the word(s) within the sentences which were assumed to prime the sentence ending, expected ending= the endings primed by the sentences which we expected participants to produce, length = sentence length (short or long), prime position = whether the semantic prime was placed at the beginning or the end of the sentence.

Appendix D: Chapter 5 Stimuli

The experimental stimuli included 90 sentences missing their final word, which were each paired with three different completions (corresponding to the three mismatched conditions). condition a – possible, unlikely; condition b – impossible, related, condition c – impossible, unrelated. In turn, three stimuli sets were created which each consisted of 30 stimuli pairs corresponding to each of these conditions. Stimuli pairs were not repeated between lists. In addition, all three sets included (the same) 30 neutral stimuli pairs, as well as 30 filler matched sentences.

Target words in conditions a and c were matched on word frequency (NWatch, Davis, 2005, condition a M = 1.15 , SD = 0.57 ; condition c M = 1.07 , SD = 0.57; t(163.940) = 0.967, p = 0.335) and imageability (Bristol/MRC, condition a M = 581.60 , SD = 95.29 ; condition c M = 601.62 , SD = 36.30; t(77.160) = -1.441, p = 0.154). Target words in these two conditions were also matched in terms of length (i.e., number of syllables and phonemes, condition a number of syllables M = 1.71 , SD = 0.66 ; condition c M = 1.86 , SD = 0.84; t(168.119) = -1.282, p = 0.201); number of phonemes , condition a M = 4.83 , SD = 1.66 ; condition c M = 5.31 , SD = 1.99; t(172.459) = -1.752, p = 0.201).

Target words in conditions b and c were also matched on word frequency (NWatch, Davis, 2005, condition b M = 1.16 , SD = 0.60 ; condition c M = 1.07 , SD = 0.57; t(149.421) = 0.953, p = 0.342); as well as in in terms of length (i.e., number of syllables and phonemes, condition b number of syllables M = 1.87 , SD = 0.80 ; condition c M = 1.86 , SD = 0.84; t(177.439) = .091, p = 0.928); number of phonemes , condition b M = 5.13 , SD = 1.95 ; condition c M = 5.31 , SD = 1.99; t(177.938) = -.606, p = .545). Target imageability was slightly lower in condition b than in condition c (NWatch, Davis, 2005 condition b M = 582.61 , SD = 44.59 ; condition c M = 601.62 , SD = 36.30; t(87.624) = -2.228, p = .028).

Target words within each mismatched experimental condition were also matched with targets in the neutral condition in terms of frequency (neutral M = 1.06 , SD = 0.57), imageability (neutral M = 594.82, SD = 30.54), and length (neutral number of syllables M = 1.73, SD = 0.69), (neutral number of phonemes M = 4.90, SD = 1.69), (all ps > 0.2). Furthermore, the sentences used in the mismatched conditions were matched with the neutral sentences in terms of sentence length (number of words) mismatched M = 5.97 , SD = 1.57 ; neutral M = 6.17 , SD = 2.09; t(40.491) = -.482, p = .633); subject frequency mismatched M = 2.69 , SD = 1.24 ; neutral M = 2.45 , SD = 1.01; t(60.587) = 1.011, p = .316); and verb frequency mismatched M = 1.60 , SD = .68 ; neutral M = 1.71 , SD = .68; t(50.327) = -.767, p = .446).

Finally, neutral sentences were matched with the filler (matched) sentences on length (number of words), and subject and verb frequency; as well as on target frequency, imageability and length (all ps > 0.08).

Table D1 Chapter 5 Stimuli sets 1-3

Experimental Stimuli				
	Sentence	**Naming Target**		
		SET 1	**SET 2**	**SET 3**
1	after dinner he paid the bill to the	swimmer (a)	breakfast (b)	newspaper (c)
2	around her neck, she wore a knitted	stocking (a)	hat (b)	grape (c)
3	around the supermarket, she pushed a	ball (a)	shelf (b)	chimney (c)
4	around the wet hair, she wrapped the	lettuce (a)	straighteners (b)	trolley (c)
5	around the wound, the doctor wrapped the	leaf (a)	stethoscope (b)	jug (c)
6	at the beach, she protected herself from the scorching	barbecue (a)	Ice-creams (b)	trousers (c)
7	at the fairground, they paid to sit on the	rug (a)	candyfloss (b)	bubbles (c)
8	before lunch, he washed his	clipboard (a)	water (b)	kidneys (c)
9	before she left the house, she locked the	diary (a)	key (b)	nose (c)
10	during his birthday party, he sliced the	rubber (a)	boardgame (b)	trampoline (c)
11	flying the plane was the	gardener (a)	airport (b)	alligator (c)
12	for the birthday party, they put helium in the	car (a)	cake (b)	wall (c)

13	for the movie, i purchased a bucket of	oranges (a)	cinemas (b)	microwaves (c)
14	for the train journey, she purchased a	keyboard (a)	rail track (b)	socket (c)
15	he emptied his bowels in the	nest (a)	stomach (b)	policeman (c)
16	he handed his dinner-date a red	ladybird (a)	kitchen (b)	bridge (c)
17	he made the incision with the	thorn (a)	wound (b)	potato (c)
18	he put the rubbish in the	safe (a)	litter (b)	wasp (c)
19	he washed down the cookie with a glass of	perfume (a)	muffins (b)	pompoms (c)
20	i take my coffee with milk and	garlic (a)	cows (b)	cardigans (c)
21	in her teacup, she dunked the	broccoli (a)	kettle (b)	syrup (c)
22	in his armpits, he sprayed	hairspray (a)	shoulders (b)	buttons (c)
23	in the corner of the envelope was a	gumball (a)	post-box (b)	sharpener (c)
24	into the water, the elephant inserted its	ear (a)	rhino (b)	beach (c)
25	on her finger, she wore a	ribbon (a)	toe (b)	helicopter (c)
26	running down the middle of the shirt, were different coloured	shells (a)	jumpers (b)	bags (c)
27	Santa Claus got stuck inside a	bank (a)	bauble (b)	rose (c)
28	Santa delivered the	pizza (a)	Christmas (b)	artist (c)
29	she applied blusher on her	doll (a)	mascara (b)	heart (c)

30	she ate the salad with the	tweezers (a)	rake (b)	first-aid-kit (c)
31	she cried whilst dicing the	tears (b)	plate (c)	soap (a)
32	she inserted the letter in the red	stamp (b)	apple (c)	plant pot (a)
33	she kept the baked cookies in the	flour (b)	needle (c)	guitar (a)
34	she kept the ice in the	snowflake (b)	donkey (c)	shoebox (a)
35	she rubbed his shoulder with her	ribs (b)	fridge (c)	iPad (a)
36	she sat at the desk, on the	office (b)	toothbrush (c)	box (a)
37	she smothered the cake with	cooling-racks (b)	televisions (c)	mud (a)
38	she spread butter on the	yoghurt (b)	pond (c)	purse (a)
39	she stapled the paper with a	printer (b)	till (c)	gun (a)
40	she sweetened the tea with	saucers (b)	onions (c)	icing (a)
41	she went for a drink at her local	ice-cube (b)	clip (c)	school (a)
42	she wrapped the newborn in a	dummy (b)	sieve (c)	blazer (a)
43	sleeping in the hotel was the	suitcase (b)	stool (c)	hen (a)
44	smoking damaged her	cigarettes (b)	screws (c)	grass (a)
45	teaching the class was the	laptop (b)	pie (c)	builder (a)
46	the baby was shaking the	nursery (b)	island (c)	knife (a)

47	the barber shaved his	razor (b)	phone (c)	knee (a)
48	the beautician painted the	massage (b)	glue (c)	ceiling (a)
49	the bodybuilder enjoyed lifting the	muscles (b)	buildings (c)	saucepans (a)
50	the carpenter cut the	saw (b)	sunset (c)	shirt (a)
51	the cook marinated the	waiter (b)	flame (c)	scorpion (a)
52	the court sentenced the	cell (b)	giraffe (c)	princess (a)
53	the crown was studded with	queen's (b)	cats (c)	rocks (a)
54	the devout Muslim covered her	mosque (b)	lightning (c)	fishbowl (a)
55	the doctor prescribed the	ambulance (b)	tulips (c)	vegetables (a)
56	the dragon breathed	fairytales (b)	stoves (c)	confetti (a)
57	the ducks ate all of my	feathers (b)	mug (c)	sponge (a)
58	the finalist won a	judge (b)	grandmother (c)	iron (a)
59	the flight passenger pulled a heavy	plane (b)	staircase (c)	statue (a)
60	the gardener cut the	shovel (b)	toybox (c)	thread (a)
61	the Italian restaurant served	radios (c)	seaweed (a)	colosseums (b)
62	the lifeguard saved the drowning	tie (c)	mouse (a)	pool (b)
63	the monkey was eating the	farmer (c)	candle (a)	zookeeper (b)

64	the novelist was writing a	twig (c)	cheque (a)	pen (b)
65	the nurse fed medicine to the	candy cane (c)	band (a)	crutches (b)
66	the palace cooks prepared a	scientist (c)	bubble-bath (a)	chef (b)
67	the panda chewed the	fairground (c)	steak (a)	zoo (b)
68	the patient was rushed to hospital in a	wheel (c)	boat (a)	nurse (b)
69	the pirate discovered a chest full of	cranes (c)	teeth (a)	ships (b)
70	the priest carried a	mountain (c)	broomstick (a)	church (b)
71	the princess's carriage was pulled along by two	tables (c)	lorries (a)	tiaras (b)
72	the seamstress cut the	ketchup (c)	blackberries (a)	sewing machine (b)
73	the singer sung into a	whale (c)	bottle (a)	drummer (b)
74	the teacher cleaned the blackboard with a	fan (c)	curtain (a)	chalk (b)
75	the trick-o-treaters received lots of	planets (c)	magnets (a)	bats (b)
76	the vaccine was injected in her	cupboard (c)	tail (a)	syringe (b)
77	the wedding guest brought a beautifully wrapped	carousel (c)	carrot (a)	aisle (b)
78	they kept the bird locked up in a	tray (c)	oven (a)	beak (b)
79	they laid the dead body inside a	coat hanger (c)	drawer (a)	skeleton (b)
80	they melted cheese on top of the	jellyfish (c)	piano (a)	milk (b)

81	they poured the guests some	horses (c)	butter (a)	invitations (b)
82	they stuck together the pieces of paper with	lemon (c)	sweets (a)	compasses (b)
83	they stuffed the Christmas	berries (c)	sofa (a)	card (b)
84	to bed, she wore	lungs (c)	boots (a)	mattresses (b)
85	to her bridesmaids, the bride threw the	receptionist (c)	jelly (a)	vicar (b)
86	to measure her temperature, she used a	cloud (c)	hand (a)	sun (b)
87	to protect his head, he wore a	teapot (c)	handkerchief (a)	skull (b)
88	to put ice cream in bowls, they used a	clock (c)	ruler (a)	freezer (b)
89	to the charity they donated the	children (c)	bra (a)	beggar (b)
90	with chopsticks, she ate the	salt (c)	capsule (a)	soy-sauce (b)
91	the father asked his son to bring a	balloon (N)	balloon (N)	balloon (N)
92	the character in the story was busy destroying the	bike (N)	bike (N)	bike (N)
93	the passenger was handling the	blanket (N)	blanket (N)	blanket (N)
94	for their endeavour, they purchased a	book (N)	book (N)	book (N)
95	the man assessed the	bookshelf (N)	bookshelf (N)	bookshelf (N)
96	during the contest, he lost his	briefcase (N)	briefcase (N)	briefcase (N)
97	the neighbour created a	calendar (N)	calendar (N)	calendar (N)

98	during the day, the zookeeper used the	chart (N)	chart (N)	chart (N)
99	the grandma examined the	cherries (N)	cherries (N)	cherries (N)
100	the candidate consulted the	driver (N)	driver (N)	driver (N)
101	at the weekend, her aunt fixed the	earrings (N)	earrings (N)	earrings (N)
102	in the box, they put the	fireworks (N)	fireworks (N)	fireworks (N)
103	in the distance, she saw the	flamingo (N)	flamingo (N)	flamingo (N)
104	the little girl was distracted away from her toys by the	goat (N)	goat (N)	goat (N)
105	to the supervisor, the parent handed the	honey (N)	honey (N)	honey (N)
106	in her spare time, the doctor made	jam (N)	jam (N)	jam (N)
107	the group were sharing the	kebab (N)	kebab (N)	kebab (N)
108	the young man was holding a	lamb (N)	lamb (N)	lamb (N)
109	the girl was requesting a	mop (N)	mop (N)	mop (N)
110	at the establishment, the visitor purchased a	plant (N)	plant (N)	plant (N)
111	the man picked up the	rattle (N)	rattle (N)	rattle (N)
112	off the table, the girl lifted a	ring (N)	ring (N)	ring (N)
113	in the bag, they forgot to put the	rope (N)	rope (N)	rope (N)
114	they finally reached the	ship (N)	ship (N)	ship (N)

115	she cooked for the	skiers (N)	skiers (N)	skiers (N)
116	father discarded the	stapler (N)	stapler (N)	stapler (N)
117	the committee were instructed to remove the	swings (N)	swings (N)	swings (N)
118	the spectator arranged a meeting with the	teacher (N)	teacher (N)	teacher (N)
119	the girl observed the	volcano (N)	volcano (N)	volcano (N)
120	she informed her guest that in the other room there was a	wheelchair (N)	wheelchair (N)	wheelchair (N)

Label following target word: (a), (b), (c), (N) refers to which condition it corresponds to: (a- unlikely, possible, b- related, impossible, c- unrelated, impossible, N= neutral baseline).

Table D2 Chapter 5 Matched (filler) sentences

Matched (filler) sentences		
	Sentence	**Naming target**
121	he put out the cigarette on a glass	ashtray
122	the monkey ate the yellow	banana
123	for support descending the stairs, she gripped the	banister
124	he slept in his	bed
125	the dog was licking the	bone
126	on Guy Fawkes' night, they built a	bonfire
127	she ladled the soup into a	bowl
128	to build a sandcastle, she put the sand into a	bucket

129	he solved the maths problem on a	calculator
130	the princess lived in a	castle
131	the security guard was standing next to the	door
132	the seamstress created a	dress
133	she fed the bread to the	ducks
134	the football player broke his	leg
135	the athlete kept his kit in the	locker
136	the butcher chopped the	meat
137	they studied the bacteria under a	microscope
138	from the bank, they stole a large sum of	money
139	the baby was nursed by his	mum
140	at the salon, they filed the client's	nails
141	the doctor treated the	patient
142	before Christmas she wrapped the	presents
143	she washed her hair with	shampoo
144	the operation was performed by a	surgeon
145	with biscuits, grandma served	tea
146	into the frying pan, she cracked an	egg
147	the commuter drank coffee from a	flask
148	she grew roses in her	garden

149	the construction worker climbed the	ladder
150	before bed she turned off the	lamp

References

Alario, F.-X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 53A*(3), 741–764. https://doi.org/10.1080/027249800410535

Alves, M., Figueiredo, P., & Raposo, A. (2023). Semantic cognition in healthy ageing: Neural signatures of representation and control mechanisms in naming typical and atypical objects. *Neuropsychologia*, *184*, 108545. 10.1016/j.neuropsychologia.2023.108545

Arbuckle, T. Y., & Gold, D. P. (1993). Aging, inhibition, and verbosity. *Journal of Gerontology, 48*(5), 225–232. https://doi.org/10.1093/geronj/48.5.P225

Arbuckle, T. Y., Nohara-LeClair, M., & Pushkar, D. (2000). Effect of off-target verbosity on communication efficiency in a referential communication task. *Psychology and Aging, 15*(1), 65–77. https://doi.org/10.1037/0882-7974.15.1.65

Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, *47*(6), 907-918. DOI: 10.1016/j.neuron.2005.07.023

Balota, D. A., Watson, J. M., Duchek, J. M., & Ferraro, F. R. (1999). Cross-modal semantic and homograph priming in healthy young, healthy old, and in Alzheimer's disease individuals. *Journal of the International Neuropsychological Society*, *5*(7), 626-640. https://doi.org/10.1017/S1355617799577060

Bannon, J., Gollan, T. H., & Ferreira, V. S. (2025). Is predicting during language processing worth it? Effects of cloze probability and semantic similarity on failed predictions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 51*(1), 106–118. https://doi.org/10.1037/xlm0001347

Beese, C., Meyer, L., Vassileiou, B., & Friederici, A. D. (2017). Temporally and spatially distinct theta oscillations dissociate a language-specific from a domain-general processing mechanism across the age trajectory. *Scientific reports*, *7*(1), 11202. 10.1038/s41598-017-11632-z

Bock, J.K., & Warren, R.K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. Cognition, 21(1), 47-67. https://doi.org/10.1016/0010-0277(85)90023-X

Bögels, S., Casillas, M., & Levinson, S. C. (2018). Planning versus comprehension in turn-taking: Fast responders show reduced anticipatory processing of the question. *Neuropsychologia*, *109*, 295-310. 10.1016/j.neuropsychologia.2017.12.028

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific reports*, *5*(1), 12881. 10.1038/srep12881

Boiteau, T. W., Malone, P. S., Peters, S. A., & Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *Journal of experimental Psychology: General*, *143*(1), 295. https://doi.org/10.1037/a0031858

Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*(5), 223-233. 10.1093/geronb/60.5.p223

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*(2), 123–147. https://doi.org/10.1177/00238309010440020101

Boudiaf, N., Laboissière, R., Cousin, E., Fournet, N., Krainik, A., & Baciu, M. (2018). Behavioral evidence for a differential modulation of semantic processing and lexical production by aging: a full linear mixed-effects modeling approach. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn, 25(*1), 1-22. DOI: 10.1080/13825585.2016.1257100

Britt, A. E., Ferrara, C., & Mirman, D. (2016). Distinct effects of lexical and semantic competition during picture naming in younger adults, older adults, and people with aphasia. *Frontiers in Psychology, 7*(813). https://doi.org/10.3389/fpsyg.2016.00813

Broderick, M. P., Di Liberto, G. M., Anderson, A. J., Rofes, A., & Lalor, E. C. (2021). Dissociable electrophysiological measures of natural language processing reveal differences in speech comprehension strategy in healthy ageing. *Scientific Reports*, *11*(1), 4963. 10.1038/s41598-021-84597-9

Brown, A. S., & Nix, L. A. (1996). Age-related changes in the tip-of-the-tongue experience. *The American Journal of Psychology, 109*(1), 79–91. https://doi.org/10.2307/1422928

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology, 7,* Article 1116. https://doi.org/10.3389/fpsyg.2016.01116

Burke, D. M., & MacKay, D. G. (1997). Memory, language and ageing. *Philosophical Transactions of the Royal Society: Biological Sciences*, *352*(1363), 1845–1856. https://doi.org/10.1098/rstb.1997.0170

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(5), 542-579. https://doi.org/10.1016/0749-596X(91)90026-G

Burke, D.M., & Shafto, M.A. (2004). Aging and language production. *Current directions in psychological science: a journal of the American Psychological Society, 13*(1), 21-24. https://doi.org/10.1111/j.0963-7214.2004.01301006.x

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/10.1016/0749-596X(91)90026-G

Capilouto, G. J., Wright, H. H., & Maddy, K. M. (2016). Microlinguistic processes that contribute to the ability to relay main events: Influence of age. *Aging, Neuropsychology, and Cognition, 23*(4), 445–463. https://doi.org/10.1080/13825585.2015.1118006

Carrol, G. (2023). Old Dogs and New Tricks: Assessing Idiom Knowledge Amongst Native Speakers of Different Ages. *Journal of Psycholinguistic Research*, *52*(6), 2287-2302. 10.1007/s10936-023-09996-7

Castro, N., & James, L.E. (2014). Differences between young and older adults' spoken language production in descriptions of negative versus neutral pictures. *Aging, Neuropsychology, and Cognition, 21(*2), 222-238. https://doi.org/10.1080/13825585.2013.804902

Centre for Ageing Better (2023) *State of Ageing 2023*. https://ageing-better.org.uk/summary-state-ageing-2023-4

Cho, S., Nevler, N., Shellikeri, S., Parjane, N., Irwin, D. J., Ryant, N., Ash, S., Cieri, C., Liberman, M., & Grossman, M. (2021). Lexical and acoustic characteristics of young and older healthy adults. *Journal of Speech, Language, and Hearing Research*, *64*(2), 302-314.

Cohen, G., & Faulkner, D. (1986). Memory for proper names: Age differences in retrieval. *British Journal of Developmental Psychology, 4*(2), 187–197. https://doi.org/10.1111/j.2044-835X.1986.tb01010.x

Cole, R. A., & Perfetti, C. A. (1980). Listening for mispronunciations in a children's story: The use of context by children and adults. *Journal of Verbal Learning & Verbal Behavior, 19*(3), 297–315. https://doi.org/10.1016/S0022-5371(80)90239-X

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Condret-Santi, V., Barbeau, E. J., Matharan, F., Le Goff, M., Dartigues, J. F., & Amieva, H. (2013). Prevalence of word retrieval complaint and prediction of dementia in a population-based study of elderly subjects. *Dementia and Geriatric Cognitive Disorders*, *35*(5-6), 313-324. 10.1159/000342594

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

Connor, L. T., Spiro, A. III, Obler, L. K., & Albert, M. L. (2004). Change in object naming ability during adulthood. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 59*(5), P203–P209. https://doi.org/10.1093/geronb/59.5.P203

Cooper, P. V. (1990). Discourse production and normal aging: Performance on oral picture description tasks. *Journal of Gerontology, 45*(5), 210–214. https://doi.org/10.1093/geronj/45.5.P210

Corps, R. E., Brooke, C., & Pickering, M. J. (2022). Prediction involves two stages: Evidence from visual-world eye-tracking. *Journal of Memory and Language*, *122*, 104298. https://doi.org/10.1016/j.jml.2021.104298

Costa, A., Alario, F. -X., & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review, 12*(1), 125–131. https://doi.org/10.3758/BF03196357

De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, *82*(3), 515-535. https://doi.org/10.1353/lan.2006.0130

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning & Verbal Behavior, 20*(6), 611–629. https://doi.org/10.1016/S0022-5371(81)90202-4

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*(4), 801–838. https://doi.org/10.1037/0033-295X.104.4.801

Dennis, P. A., & Hess, T. M. (2016). Aging-related gains and losses associated with word production in connected speech. *Aging, Neuropsychology, and Cognition, 23*(6), 638–650. https://doi.org/10.1080/13825585.2016.1158233

Evrard, M. (2002). Ageing and Lexical Access to Common and Proper Names in Picture Naming, *Brain and Language, 81*(1-3), 174-179. DOI:10.1006/brln.2001.2515

Farrell, M. T., Zahodne, L. B., Stern, Y., Dorrejo, J., Yeung, P., & Cosentino, S. (2014). Subjective word-finding difficulty reduces engagement in social leisure activities in Alzheimer's disease. *Journal of the American Geriatrics Society, 62*(6), 1056-1063. 10.1111/jgs.12850

Faust, M. E., Balota, D. A., & Multhaup, K. S. (2004). Phonological blocking during picture naming in dementia of the Alzheimer type. *Neuropsychology*, *18*(3), 526. https://doi.org/10.1037/0894-4105.18.3.526

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

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469-495. https://doi.org/10.1006/jmla.1999.2660

Federmeier, K. D., & Kutas, M. (1999). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research*, *8*(3), 373-392. 10.1016/s0926-6410(99)00036-1

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

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

Fergadiotis, G., & Wright, H.H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology, 25*(11), 1414-1430. https://doi.org/10.1080/02687038.2011.603898

Ferré, P., Jarret, J., Brambati, S., Bellec, P., & Joanette, Y.(2020). Functional Connectivity of Successful Picture-Naming: Age-Specific Organization and the Effect of Engaging in Stimulating Activities. *Frontiers in Aging Neuroscience, 12*(535770), 1-14. 10.3389/fnagi.2020.535770

Ferreira, F., & Clifton Jr, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348-368. https://doi.org/10.1016/0749-596X(86)90006-9

Ferreira, F., & Çokal, D. (2016). Sentence processing. In *Neurobiology of language* (pp. 265-274). Academic Press.

Feyereisen, P. (1997). The competition between gesture and speech production in dual-task paradigms. *Journal of Memory and Language, 36*(1), 13–33. https://doi.org/10.1006/jmla.1995.2458

Fieder, N., Wartenburger, I., & Abdel Rahman, R. (2019). A close call: Interference from semantic neighbourhood density and similarity in language production. *Memory & cognition*, *47*, 145-168. https://doi.org/10.3758/s13421-018-0856-y

Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning & Verbal Behavior, 18*(1), 1–20. https://doi.org/10.1016/S0022-5371(79)90534-6

Ford, J. M., Woodward, S. H., Sullivan, E. V., Isaacks, B. G., Tinklenberg, J. R., Yesavage, J. A., & Roth, W. T. (1996). N400 evidence of abnormal responses to speech in Alzheimer's disease. *Electroencephalography and clinical Neurophysiology*, *99*(3), 235-246. https://doi.org/10.1016/0013-4694(96)95049-X

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291-325. https://doi.org/10.1016/0010-0277(78)90002-1

Garrett, M. F. (1984). The organization of processing structure for language production: Applications to aphasie speech. In D. Caplan, A. R. Lecours & A. Smith (Eds.), Biological perspectives on language, (pp. 172-193). Cambridge, M.A.: M.I.T. Press.

Gertel, V. H., Karimi, H., Dennis, N. A., Neely, K. A., & Diaz, M. T. (2020). Lexical frequency affects functional activation and accuracy in picture naming among older and younger adults. *Psychology and Aging*, *35*(4), 536.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1-76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 94–126). The MIT Press.

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 640–654. https://doi.org/10.1037/0096-1523.10.5.640

Glosser, G., & Deser, T. (1992). A comparison of changes in macrolinguistic and microlinguistic aspects of discourse production in normal aging. *Journal of Gerontology: Psychological Sciences, 47,* 266–272. 10.1093/geronj/47.4.p266

Gordon, J. K., & Kurczek, J. C. (2014). The ageing neighbourhood: Phonological density in naming. *Language, Cognition and Neuroscience, 29*(3), 326–344. https://doi.org/10.1080/01690965.2013.837495

Gordon, J.K., & Kindred, N.K. (2011). Word retrieval in ageing: An exploration of the task constraint hypothesis. *Aphasiology, 25(*6-7), 774-788. https://doi.org/10.1080/02687038.2010.539699

Gould, O. N., & Dixon, R. A. (1993). How we spent our vacation: Collaborative storytelling by young and old adults. *Psychology and Aging, 8*(1), 10–17. https://doi.org/10.1037/0882-7974.8.1.10

Goulet, P., Ska, B., & Kahn, H. J. (1994). Is there a decline in picture naming with advancing age? *Journal of Speech & Hearing Research, 37*(3), 629–644. https://doi.org/10.1044/jshr.3703.629

Gunter, T. C., Jackson, J. L., & Mulder, G. (1992). An electrophysiological study of semantic processing in young and middle-aged academics. *Psychophysiology*, *29*(1), 38-54. https://doi.org/10.1111/j.1469-8986.1992.tb02009.x

Haigh, P., Hanif, N., & de Bruin, A. (2022). Diving into a pool or volcano? Examining the influence of sentence context and task demands on sentence reading in younger and older adults. *Plos one*, *17*(12), e0279555. https://doi.org/10.1371/journal.pone.0279555

Hamberger, M. J., Friedman, D., Ritter, W., & Rosen, J. (1995). Event-related potential and behavioral correlates of semantic processing in Alzheimer' s patients and normal controls. *Brain and Language*, *48*(1), 33-68. https://doi.org/10.1006/brln.1995.1002

Hameau, S., Nickels, L., & Biedermann, B. (2019). Effects of semantic neighbourhood density on spoken word production. *Quarterly Journal of Experimental Psychology*, *72*(12), 2752-2775. https://doi.org/10.1177/1747021819859850

Hantsch, A., Jescheniak, J. D., & Schriefers, H. (2009). Distractor modality can turn semantic interference into semantic facilitation in the picture–word interference task: Implications for theories of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1443-53. 10.1037/a0017020

Hasher, L., & Zacks, R. T. (1988). *Working memory, comprehension, and aging: A review and a new view.* In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory, Vol. 22* (p. 193–225). Academic Press. https://doi.org/10.1016/S0079-7421(08)60041-9

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

Heine, M. K., Ober, B. A., & Shenaut, G. K. (1999). Naturally occurring and experimentally induced tip-of-the-tongue experiences in three adult age groups. *Psychology and Aging, 14*(3), 445–457. https://doi.org/10.1037/0882-7974.14.3.445

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555-568. https://doi.org/10.1016/j.wocn.2010.08.002

Heller, R. B., & Dobbs, A. R. (1993). Age differences in word finding in discourse and nondiscourse situations. *Psychology and Aging, 8*(3), 443–450. https://doi.org/10.1037/0882-7974.8.3.443

Hilviu, D., Parola, A., Bosco, F. M., Marini, A., & Gabbatore, I. (2024). Grandpa, tell me a story! Narrative ability in healthy aging and its relationship with cognitive functions and Theory of Mind. *Language, Cognition and Neuroscience*, *40*(1), 103–121. https://doi.org/10.1080/23273798.2024.2401027

Hoffman, P. (2018). An individual differences approach to semantic cognition: Divergent effects of age on representation, retrieval and selection. *Scientific reports*, *8*(1), 8145. 10.1038/s41598-018-26569-0

Hoffman, P., Loginova, E., & Russell, A. (2018). Poor coherence in older people's speech is explained by impaired semantic and executive processes. *eLIFE, 7*(e38907), 1-43. doi: 10.7554/eLife.38907

Horton, W. S., Spieler, D. H., & Shriberg, E. (2010). A corpus analysis of patterns of age-related change in conversational speech. *Psychology and Aging, 25*(3), 708–713. https://doi.org/10.1037/a0019424

Hoyau, E., Boudiaf, N., Cousin, E., Pichat, C., Fournet, N., Krainik, A., Jaillard, A., & Baciu, M. (2017). Aging Modulates the Hemispheric Specialization during Word Production. *Frontiers in Aging Neuroscience*, *9*, (125). https://doi.org/10.3389/fnagi.2017.00125

Hummert, M. L., Garstka, T. A., Ryan, E. B., & Bonnesen, J. L. (2004). *The Role of Age Stereotypes in Interpersonal Communication.* In J. F. Nussbaum & J. Coupland (Eds.), *LEA's communication series. Handbook of communication and aging research* (p. 91–114). Lawrence Erlbaum Associates Publishers.

Hustá, C., Zheng, X., Papoutsi, C., & Piai, V. (2021). Electrophysiological signatures of conceptual and lexical retrieval from semantic memory. *Neuropsychologia*, *161*, 107988.

James, L. E., & Burke, D. M. (2000). Phonological priming effects on word retrieval and tip-of-the-tongue experiences in young and older adults. *Journal of experimental psychology. Learning, memory, and cognition*, *26*(6), 1378–1391. https://doi.org/10.1037//0278-7393.26.6.1378

James, L. E., Burke, D. M., Austin, A., & Hulme, E. (1998). Production and perception of "verbosity" in younger and older adults. *Psychology and Aging, 13*(3), 355–367. https://doi.org/10.1037/0882-7974.13.3.355

James, L. E., Schmank, C. J., Castro, N., & Buchanan, T. W. (2018). Tip of the tongue states increase under evaluative observation. *Journal of Psycholinguistic Research*, *47*, 169-178. https://doi.org/10.1007/s10936-017-9524-9

Jefferies, E. (2013). The neural basis of semantic cognition: converging evidence from neuropsychology, neuroimaging and TMS. *Cortex*, *49*(3), 611-625. 10.1016/j.cortex.2012.10.008

Jongman, S. R. (2021). The attentional demands of combining comprehension and production in conversation. In *Psychology of learning and motivation* (Vol. 74, pp. 95-140). Academic Press.

Juncos-Rabadán, O., Pereiro, A. X., & Rodríguez, M. S. (2005). Narrative speech in aging: Quantity, information content, and cohesion. *Brain and Language, 95*(3), 423–434. https://doi.org/10.1016/j.bandl.2005.04.001

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and language*, *49*(1), 133-156. https://doi.org/10.1016/S0749-596X(03)00023-8

Kavé, G., & Goral, M. (2017). Do age-related word retrieval difficulties appear (or disappear) in connected speech? *Aging, Neuropsychology, and Cognition, 24*(5), 508-527. https://doi.org/10.1080/13825585.2016.1226249

Kavé, G., & Halamish, V. (2015). Doubly blessed: older adults know more vocabulary and know better what they know. *Psychology and aging*, *30*(1), 68-73. https://doi.org/10.1037/a0038669

Kavé, G., & Yafé, R. (2014). Performance of younger and older adults on tests of word knowledge and word retrieval: Independence or interdependence of skills? *American Journal of Speech-Language Pathology, 23*(1), 36–45. https://doi.org/10.1044/1058-0360(2013/12-0136)

Kavé, G., Samuel-Enoch, K., & Adiv, S. (2009). The association between age and the frequency of nouns selected for production. *Psychology and Aging, 24*, 17–27. doi:10.1037/a0014579

Kavé, G., Samuel-Enoch, K., & Adiv, S. (2009). The association between age and the frequency of nouns selected for production. *Psychology and Aging, 24*(1), 17–27. https://doi.org/10.1037/a0014579

Keller-Cohen, D., Fiori, K., Toler, A., & Bybee, D. (2006). Social relations, language and cognition in the 'oldest old'. *Ageing & Society*, *26*(4), 585-605. 10.1017/S0144686X06004910

Kemper, S., & Harden, T. (1999). Experimentally disentangling what's beneficial about elderspeak from what's not. *Psychology and Aging, 14*(4), 656–670. https://doi.org/10.1037/0882-7974.14.4.656

Kemper, S., & Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychology and Aging, 16*(2), 312–322. https://doi.org/10.1037/0882-7974.16.2.312

Kemper, S., Schmalzried, R., Hoffman, L., & Herman, R. (2010). Aging and the vulnerability of speech to dual task demands. *Psychology and Aging, 25*(4), 949–962. https://doi.org/10.1037/a0020000

Kemper, S., Thompson, M., & Marquis, J. (2001). Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content. *Psychology and Aging, 16*(4), 600–614. https://doi.org/10.1037/0882-7974.16.4.600

Klaus, J., Mädebach, A., Oppermann, F., & Jescheniak, J. D. (2017). Planning sentences while doing other things at the same time: Effects of concurrent verbal and visuospatial working memory load. *Quarterly Journal of Experimental Psychology*, *70*(4), 811-831. 10.1080/17470218.2016.1167926

Kleiman, G. M. (1980). Sentence frame contexts and lexical decisions: Sentence-acceptability and word-relatedness effects. *Memory & Cognition*, 8, 336–344. https://doi.org/10.3758/BF03198273

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European journal of cognitive psychology*, *16*(1-2), 262-284. https://doi.org/10.1080/09541440340000213

Krieger-Redwood, K., Lanzoni, L., Gonzalez Alam, T. R., Jackson, R. L., Smallwood, J., & Jefferies, E. (2025). Divergent and convergent creativity relate to different aspects of semantic control. *Imaging Neuroscience*, *3*, imag_a_00502. https://doi.org/10.1162/imag_a_00502.

Krethlow, G., Fargier, R., & Laganaro, M. (2020). Age-specific effects of lexical–semantic networks on word production. *Cognitive Science, 44*(11), Article e12915. https://doi.org/10.1111/cogs.12915

Krethlow, G., Fargier, R., Atanasova, T., Ménétré, E., & Laganaro, M. (2024). Asynchronous behavioral and neurophysiological changes in word production in the adult lifespan. *Cerebral Cortex*, *34*(5), bhae187. 10.1093/cercor/bhae187

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621-647. 10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161-163. 10.1038/307161a0

La Heij, W., & Vermeij, M. (1987). Reading versus naming: The effect of target set size on contextual interference and facilitation. *Perception & Psychophysics*, *41*, 355-366. https://doi.org/10.3758/BF03208237

La Heij, W., Dirkx, J., & Kramer, P. (1990). Categorical interference and associative priming in picture naming. *British Journal of Psychology*, *81*(4), 511-525. https://doi.org/10.1111/j.2044-8295.1990.tb02376.x

LaGrone, S., & Spieler, D. H. (2006). Lexical competition and phonological encoding in young and older speakers. *Psychology and Aging, 21*(4), 804–809. https://doi.org/10.1037/0882-7974.21.4.804

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326-338. https://doi.org/10.1016/j.jml.2009.06.004

Laver, G. D., & Burke, D. M. (1993). Why do semantic priming effects increase in old age? A meta-analysis. *Psychology and Aging, 8*(1), 34–43. https://doi.org/10.1037/0882-7974.8.1.34

Lee, J., West, D., Pellegrini, C., Wei, J., Wilcox, S., Neils-Strunjas, J., Rudisill, A. C., Friedman, D. B., Marquez, D. X., Dzwierzynski, J., Balls-Berry, J., Gajadhar, R., & Yang, C. H. (2025). Walking Interventions and Cognitive Health in Older Adults: A Systematic Review of Randomized Controlled Trials. *American journal of health promotion : AJHP*, *39*(7), 1051–1067. https://doi.org/10.1177/08901171251328858

Leeper, L. H., & Culatta, R. (1995). Speech fluency: effect of age, gender and context. *Folia phoniatrica et logopaedica*, *47*(1), 1-14.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–38. https://doi.org/10.1017/S0140525X99001776

Levelt, W.J.M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences, 98*(23), 13464-13471. https://doi.org/10.1073/pnas.23145949

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in cognitive sciences*, *20*(1), 6-14. https://doi.org/10.1016/j.tics.2015.10.010

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology, 6(* 731). https://doi.org/10.3389/fpsyg.2015.00731

Liu, X., & Wang, W. (2019). The effect of distance on sentence processing by older adults. *Frontiers in Psychology, 10,* Article 2455. https://doi.org/10.3389/fpsyg.2019.02455

Long, M. R., Horton, W. S., Rohde, H., & Sorace, A. (2018). Individual differences in switching and inhibition predict perspective-taking across the lifespan. *Cognition, 170,* 25–30. https://doi.org/10.1016/j.cognition.2017.09.004

Lovelace, E. A., & Twohig, P. T. (1990). Healthy older adults' perceptions of their memory functioning and use of mnemonics. *Bulletin of the Psychonomic Society*, *28*, 115— 118.

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, *7*(6), 485-495. https://doi.org/10.3758/BF03198265

MacKay, D. G., & James, L. E. (2004). Sequencing, Speech Production, and Selective Effects of Aging on Phonological and Morphological Speech Errors. *Psychology and Aging, 19*(1), 93–107. https://doi.org/10.1037/0882-7974.19.1.93

Mackenzie, C. (2000). Adult spoken discourse: the influences of age and education. *International Journal of Language & Communication Disorders*, *35*(2), 269-85. 10.1080/136828200247188

Magyari, L., Bastiaansen, M. C. M., de Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience, 26*(11), 2530–2539. https://doi.org/10.1162/jocn_a_00673

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 503–535. https://doi.org/10.1037/0278-7393.33.3.503

Marini, A., Boewe, A., Caltagirone, C., & Carlomagno, S. (2005). Age-related Differences in the Production of Textual Descriptions. *Journal of Psycholinguistic Research, 34*(5), 439–463. https://doi.org/10.1007/s10936-005-6203-z

Marini, A., Petriglia, F., D'Ortenzio, S., Bosco, F. M., & Gasparotto, G. (2025). Unveiling the dynamics of discourse production in healthy aging and its connection to cognitive skills. *Discourse Processes*, *62*(6–7), 479–501. https://doi.org/10.1080/0163853X.2025.2507548

Martin, N., Gagnon, D. A., Schwartz, M. F., Dell, G. S., & Saffran, E. M. (1996). Phonological facilitation of semantic errors in normal and aphasic speakers. *Language and Cognitive Processes, 11*(3), 257–282. https://doi.org/10.1080/016909696387187

Maylor, E. A. (1990b). Recognizing and naming faces: Aging, memory, retrieval and tip of the tongue state. *Journal of Gerontology*, *45*, 215–226. https://doi.org/10.1093/geronj/45.6.P215

Menenti, L., Gierhan, S. M., Segaert, K., & Hagoort, P. (2011). Shared language: overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science*, *22*(9), 1173-1182. 10.1177/0956797611418347

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology, 90(2), 227–234. https://doi.org/10.1037/h0031564

Mirman, D. (2011). Effects of near and distant semantic neighbors on word production. *Cognitive, Affective, & Behavioral Neuroscience*, *11*(1), 32-43. https://doi.org/10.3758/s13415-010-0009-7

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F., Matthew Husband, E., Ito, A., Kazanina, N., Kogan, V., Kohút, Z., Kulakova, E., Mézière, D., Politzer-Ahles, S., Rousselet, G., Rueschemeyer, S. A., … Von Grebmer Zu Wolfsthurn, S. (2020). Dissociable effects of prediction and integration during language comprehension: evidence from a large-scale study using brain potentials. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1791), 20180522. https://doi.org/10.1098/rstb.2018.0522

Oberle, S., & James, L. E. (2013). Semantically-and phonologically-related primes improve name retrieval in young and older adults. *Language and cognitive processes*, *28*(9), 1378-1393.

Paolieri, D., Marful, A., Morales, L., & Bajo, M. T. (2018). The modulating effect of education on semantic interference during healthy aging. *PloS one*, *13*(1), e0191656.

Papoutsi, C., & Piai, V. (2023). The effect of structure on context-driven word production: an EEG study. *PsyArXiv*. doi:10.31234/osf.io/r73k8

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

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002-1044. 10.1037/bul0000158

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329-347. 10.1017/S0140525X12001495

Poon, L. W., & Fozard, J. L. (1978). Speed of retrieval from long-term memory in relation to age, familiarity, and datedness of information. *Journal of Gerontology, 33*(5), 711–717. https://doi.org/10.1093/geronj/33.5.711

Rabbitt, P., Maylor, E., McInnes, L., Bent, N., & Moore, B. (1995). What goods can self-assessment questionnaires deliver for cognitive gerontology? *Applied Cognitive Psychology, 9, Spec Issue,* 127–152. https://doi.org/10.1002/acp.2350090709

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science*, *6*(1), 5-42. 10.1111/tops.12078

Rastle, K. G., & Burke, D. M. (1996). Priming the tip of the tongue: Effects of prior processing on word retrieval in young and older adults. *Journal of Memory and Language, 35*(4), 585–605. https://doi.org/10.1006/jmla.1996.0031

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*(3), 448-65. 10.1037/0882-7974.21.3.448

Reason, J. T., & Lucas, D. (1984). Using cognitive diaries to investigate naturally occurring memory blocks. In J. E. Harris & P. E. Morris (Eds.)

Riest, C., Jorschick, A. B., & de Ruiter, J. P. (2015). Anticipation in turn-taking: mechanisms and information sources. *Frontiers in Psychology*, *6*, 89. https://doi.org/10.3389/fpsyg.2015.00089

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*(1-3), 107–142. https://doi.org/10.1016/0010-0277(92)90041-F

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64*(3), 249–284. https://doi.org/10.1016/S0010-0277(97)00027-9

Roelofs, A. (2003). Shared phonological encoding processes and representations of language in bilingual speakers. *Language and Cognitive Processes, 18*(2), 175–204. https://doi.org/10.1080/01690960143000515

Saling, L. L., Laroo, N., & Saling, M. M. (2012). When more is less: Failure to compress discourse with re-telling in normal ageing. *Acta Psychologica, 139*(1), 220–224. https://doi.org/10.1016/j.actpsy.2011.10.005

Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, *8*(4), 535.

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging, 30*(4), 507–514. https://doi.org/10.1016/j.neurobiolaging.2008.09.023

Salthouse, T. A. (2010). *Oxford psychology series: Vol. 49. Major issues in cognitive aging.* Oxford University Press.

Salthouse, T. A., & Mandell, A. R. (2013). Do age-related increases in tip-of-the-tongue experiences signify episodic memory impairments? *Psychological Science, 24*(12), 2489–2497. https://doi.org/10.1177/0956797613495881

Samu, D., Campbell, K.L., Tsvetanov, K.A., Shafto, M.A., Cam-CAN consortium, Tyler, L.,K. (2017). Preserved cognitive functions with age are determined by domain-dependent shifts in network responsivity. *Nature Communications, 8(*14743). DOI: 10.1038/ncomms14743

Schmitter-Edgecombe, M., Vesneski, M., & Jones, D. W. R. (2000). Aging and word-finding: A comparison of spontaneous and constrained naming tests. *Archives of Clinical Neuropsychology, 15*(6), 479–493. https://doi.org/10.1016/S0887-6177(99)00039-6

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*, 86-102. https://doi.org/10.1016/0749-596X(90)90011-N

Schuberth, R. E., Spoehr, K. T., & Lane, D. M. (1981). Effects of stimulus and contextual information on the lexical decision process. *Memory & Cognition, 9*(1), 68–77. https://doi.org/10.3758/BF03196952

Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(2), 344–354. https://doi.org/10.1037/0278-7393.14.2.344

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language, 24*(2), 232–252. https://doi.org/10.1016/0749-596X(85)90026-9

Schwantes, F. M. (1985). Expectancy, integration, and interactional processes: Age differences in the nature of words affected by sentence context. *Journal of Experimental Child Psychology, 39*(2), 212–229. https://doi.org/10.1016/0022-0965(85)90038-4

Schweich, M., Van der Linden, M., Bredart, S., Bruyer, R., Nelles, B., & Schils, J.-P. (1992). Daily-life difficulties in person recognition reported by young and elderly subjects. *Applied Cognitive Psychology, 6*(2), 161–172. https://doi.org/10.1002/acp.2350060206

Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012). Shared syntax in language production and language comprehension—an fMRI study. *Cerebral Cortex*, *22*(7), 1662-1670. 10.1093/cercor/bhr249

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

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

Shafto, M. A., James, L. E., Abrams, L., Tyler, L. K., & Cam-Can. (2016). Age-related increases in verbal knowledge are not associated with word finding problems in the Cam-CAN cohort: What you know won't hurt you. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *72*(1), 100-106. 10.1093/geronb/gbw074

Shao, Z., & Rommers, J. (2020). How a question context aids word production: Evidence from the picture–word interference paradigm. *Quarterly Journal of Experimental Psychology*, *73*(2), 165-173. doi:10.1177/1747021819882911

Shewan, C. M., & Henderson, V. L. (1988). Analysis of spontaneous language in the older normal population. *Journal of Communication Disorders, 21*(2), 139–154. https://doi.org/10.1016/0021-9924(88)90002-0

Shitova, N., Roelofs, A., Coughler, C., & Schriefers, H. (2017). P3 event-related brain potential reflects allocation and use of central processing capacity in language production. *Neuropsychologia*, *106*, 138-145. 10.1016/j.neuropsychologia.2017.09.024

Sjerps, M. J., & Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition*, *136*, 304-324. https://doi.org/10.1016/j.cognition.2014.10.008

Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging, 14*(3), 458–472. https://doi.org/10.1037/0882-7974.14.3.458

Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance, 22*(2), 461–479. https://doi.org/10.1037/0096-1523.22.2.461

Stanovich, K. E., & West, R. F. (1981). The effect of sentence context on ongoing word recognition: Tests of a two-process theory. *Journal of Experimental Psychology: Human Perception and Performance, 7*(3), 658–672. https://doi.org/10.1037/0096-1523.7.3.658

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General, 112*(1), 1-36. https://doi.org/10.1037/0096-3445.112.1.1

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K. E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10587–10592. https://doi.org/10.1073/pnas.0903616106

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

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*(3), 382-392. https://doi.org/10.1016/j.ijpsycho.2011.12.007

Trunk, D. L., & Abrams, L. (2009). Do younger and older adults' communicative goals influence off-topic speech in autobiographical narratives? *Psychology and Aging, 24*(2), 324–337. https://doi.org/10.1037/a0015259

Tsang, H-L., & Lee, T.M.C. (2003). The effect of ageing on confrontational naming ability. Archives of Clinical Neuropsychology, *18*(1), 81-89. https://doi.org/10.1016/S0887-6177(01)00184-6

Tsang, H., & Lee, T.M.C. (2003). The effect of ageing on confrontational naming ability. *Archives of Clinical Neuropsychology*, *18*(1), 81–89. https://doi.org/10.1016/s0887-6177(01)00184-6

Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging, 17*(3), 453–467. https://doi.org/10.1037/0882-7974.17.3.453

Ulatowska, H. K., Hayashi, M. M., Cannito, M. P., & Fleming, S. G. (1986). Disruption of reference in aging. *Brain and Language, 28,* 24–41. 10.1016/0093-934x(86)90088-x

Verhaegen, C., & Poncelet, M. (2013). Changes in naming and semantic abilities with aging from 50 to 90 years. *Journal of the International Neuropsychological Society, 19(*2), 119-26. 10.1017/S1355617712001178

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

Wei, H. T., Kulzhabayeva, D., Erceg, L., Robin, J., Hu, Y. Z., Chignell, M., & Meltzer, J. A. (2024). Cognitive components of aging-related increase in word-finding difficulty. *Aging, Neuropsychology, and Cognition*, *31*(6), 987-1019.

Wen, H., & Dong, Y. (2023). The effect of ageing on confrontation naming in healthy older adults: A three-level meta-analysis. *Journal of Cognitive Psychology*, *35*(4), 480-508. https://doi.org/10.1080/20445911.2023.2184745

West, R., & Alain, C. (2000). Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. *Psychophysiology*, *37*(2), 179-189.

Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: age-related differences in the use of sentential context. *Psychology and Aging*, *27*(4), 975-988. https://doi.org/10.1037/a0029206

Woodward, S. H., Ford, J. M., & Hammett, S. C. (1993). N4 to spoken sentences in young and older subjects. *Electroencephalography and Clinical Neurophysiology*, *87*(5), 306-320. https://doi.org/10.1016/0013-4694(93)90184-W

Wright, H. H., Koutsoftas, A. D., Capilouto, G. J., & Fergadiotis, G. (2014). Global coherence in younger and older adults: Influence of cognitive processes and discourse type. *Aging, Neuropsychology, and Cognition*, *21*(2), 174-196. https://doi.org/10.1080/13825585.2013.794894

Yoon, J., Campanelli, L., Goral, M., Marton, K., Eichorn, N., & Obler, L. K. (2015). The effect of plausibility on sentence comprehension among older adults and its relation to cognitive functions. *Experimental Aging Research*, *41*(3), 272-302. https://doi.org/10.1080/0361073X.2015.1021646

Zhang, S., Kang, K., Zhong, J., Yu, H., Ip, C. T., & Zhang, H. (2025). Cognitive control states and traits modulate lexical competition during word production. *Scientific Reports*, *15*(1), 25381.