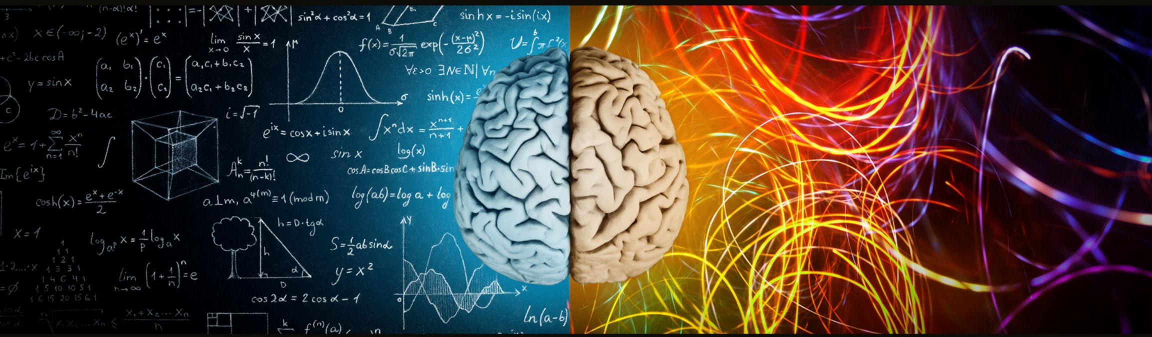

The Psychology of LEARNING AND MOTIVATION

Volume 78

Speaking, Writing and Communicating

Volume Editors

KARA D. FEDERMEIER

JESSICA L. MONTAG

VOLUME SEVENTY EIGHT

The Psychology of LEARNING AND MOTIVATION

Speaking, Writing and Communicating

Series Editor

KARA D. FEDERMEIER

*Department of Psychology, Program in Neuroscience,
and the Beckman Institute for Advanced Science and Technology,
University of Illinois, Champaign, IL, United States*

VOLUME SEVENTY EIGHT

The Psychology of LEARNING AND MOTIVATION

Speaking, Writing and Communicating

Edited by

KARA D. FEDERMEIER
Department of Psychology, Program in Neuroscience, and the Beckman Institute for Advanced Science and Technology, University of Illinois, Champaign, IL, United States

JESSICA L. MONTAG
Department of Psychology, University of Illinois, Champaign, IL, United States

ELSEVIER

ACADEMIC PRESS
An imprint of Elsevier

Academic Press is an imprint of Elsevier
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
525 B Street, Suite 1650, San Diego, CA 92101, United States
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
125 London Wall, London, EC2Y 5AS, United Kingdom

First edition 2023

ISBN: 978-0-443-13409-8
ISSN: 0079-7421

For information on all Academic Press publications
visit our website at https://www.elsevier.com/books-and-journals

Publisher: Zoe Kruze
Acquisitions Editor: Mariana Kuhl
Developmental Editor: Naiza Ermin Mendoza
Production Project Manager: Vijayaraj Purushothaman
Cover Designer: Christian J. Bilbow

Typeset by STRAIVE, India

Contents

Contributors

Laurel Brehm
Department of Linguistics, University of California Santa Barbara, Santa Barbara, CA, United States

Ruth E. Corps
Psychology of Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Paola E. Dussias
Department of Spanish, Italian, and Portuguese, Pennsylvania State University, State College, PA, United States

Arella E. Gussow
Department of Psychology, University of Wisconsin-Madison, Madison, WI, United States

Eleanor Heggdal Lønes
Division of Psychology, University of Dundee, Dundee, United Kingdom

Yuki Kamide
Division of Psychology, University of Dundee, Dundee, United Kingdom

Amanda C. Kelley
Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, United States

Alissa Melinger
Division of Psychology, University of Dundee, Dundee, United Kingdom

Jorge R. Valdés Kroff
Spanish and Portuguese Studies, University of Florida, Gainesville, FL, United States

This page intentionally left blank

CHAPTER ONE

What's an error anyway? Speaker- and listener-centered approaches to studying language errors

Laurel Brehm*
Department of Linguistics, University of California Santa Barbara, Santa Barbara, CA, United States
*Corresponding author: e-mail address: lbrehm@ucsb.edu

Contents

Abstract

Everybody makes errors when they speak and when they listen. Everybody also has the ability to recover meaning from the error-containing utterances that other people produce. Considering both speaker and listener perspectives to speech errors is therefore important for a full picture of the phenomenon. Importantly, speakers and listeners do not always align in their understanding of an error: Sometimes listeners can recover what was intended by the speaker, but sometimes they cannot. In this chapter, I review research on speech errors from both speaker and listener perspectives, focusing on when errors happen in speaking and in listening and when speakers and listeners come to disagree about what an utterance means. This provides a bridge between several approaches to the study of language errors, variability, and repairs, and it sheds new light on the question of 'what's an error anyway?.'

Psychology of Learning and Motivation, Volume 78
ISSN 0079-7421
https://doi.org/10.1016/bs.plm.2023.02.001

1. Introduction

It is a fairly trivial statement that language as used in the real world contains errors—utterances that were not produced as intended. Everybody makes them, and some of us make more errors than others. Freud built his career on what are fundamentally language errors, and there are public figures famous for their frequent language errors. Consider the historical figure Reverend Spooner, who made so many malapropisms that an error class, 'Spoonerisms,' was named after him, or the American politicians Dan Quayle and George H.W. Bush, who made spelling and speech errors, respectively, that generated a lot of press in the news and popular media. This shows how language errorscan be remarkably salient—and occasionally, how these can have big consequences for the error-maker.

The systematic study of language production began with speech error research, and this research has caused us to draw many generalizations about the patterns of mistakes that people make when they speak. Early speech error research aimed to understand the production system by finding systematicities in the errors people make (e.g., Fromkin, 1971; Garrett, 1975), leading to a fairly good field-wide consensus about what the component processes are to language production and what architecture would be suitable to model these computationally (e.g., Dell, 1986; Levelt, Roelofs, & Meyer, 1999).

However: despite the fact that speech errors are easy to observe, and the fact that the field of language production has a reasonably clear idea of why they happen, it is also true that the question of 'what is a speech error' is not as trivial to answer as it seems on the surface. I proposed a speaker-centered perspective of speech and language errors in the first paragraph: an error is an utterance that was not produced as intended. This means that to truly understand what counts as an error, the speech error researcher (or any other listener) needs to be able understand the intent of the speaker. This can be deceptively challenging, as it involves recognizing that the speaker's dialect may differ in important ways from the listener's that make it hard to tell what should be categorized as an error at all.

Furthermore, listeners also make contributions to dialogues, and it is important for our understanding of speech errors to remember that listener perspectives do not always align with speaker perspectives. This is supported by recent psycholinguistic research showing that comprehenders are fairly skilled at repairing the errors they see and hear (Gibson, Bergen, &

Piantadosi, 2013; Levy, 2008), but that they also make their own errors in perception (Bond, 2021; Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira, 2003). It is also supported by recent sociolinguistic research showing how variation can lead to potential discrimination and misperception: a speaker of a more standard or higher-prestige dialect (consider the type of English spoken by White newscasters) may misperceive another speaker's correct utterance as an error especially often if the other speaker has a non-standard or lower prestige dialect (consider casual African-American Vernacular English or casual Southern American English; e.g., Craft, Wright, Weissler, & Queen, 2020).

This chapter therefore focuses on speech error research from both speaker and listener perspectives, with special focus on how differences in perspective can create differences in interpretation. Consider for example the process of inferring the intended meaning from each of the three utterances in Table 1: 'I make mis*m*akes' (a speech error that the author observed in a conversation about speech errors), /mIs/ ('miss,' a dialectal pronunciation of 'mist' associated with African American Vernacular English; see Staum-Casasanto, 2008), and 'The *key* to the cabinets *are* rusty' (an elicited error following from Bock & Miller, 1991). Note that here and following, I highlight errors with italics and put examples in plain English text when possible, using broad phonetic notation such as /I/ or narrow phonetic notation such as [I] only where necessary to make the point.

Table 1 Mismatches in utterance recoverability between speaker and listener.

	Speaker says what was intended	**Speaker does not say what was intended**
Listener infers what was intended	Utterance = Speaker understands = Listener understands	Utterance: 'I make mis*m*ake*s*' Speaker understands: I make mistakes Listener understands: I make mistakes
Listener does not infer what was intended	Utterance: /mIs/ Speaker understands: Mist Listener understands: Miss	Utterance: 'The *key* to the cabinets *are* rusty' Speaker understands: The key to the cabinets is rusty Listener understands: The keys to the cabinets are rusty

In the first case, 'I make mis*m*akes,' the intended meaning is likely clear to both speaker and listener because the source of the error is clear: the speaker re-produced (perseverated) the /m/ that occurs in the previous word and previous syllable. Here, both parties can easily infer that the utterance 'I make mistakes' was intended, coming to the same understanding of the utterance, even though this understanding diverges from what was actually said. This error-containing utterance is remarkably easy to understand because of the hidden cognitive machinery that causes and allows us to repair errors. I discuss the production of examples like these in the 'Speaker-centered approaches' section below, and their interpretation in the 'Listener-centered approaches' section below.

In the second case, /mIs/, the speaker is producing their intended target—their own correct realization of the word 'mist'—but an unknowing listener may not derive the correct interpretation. This is because the 'standard' pronunciation in a dialect does not hold for all speakers, and moreover, what gets considered to be the standard pronunciation intersects with privilege, power, and other important social factors. These utterances may have unfortunate overlap with the 'slips of the ear' that are more commonly studied as errors in speech perception (Bond, 2021). In this chapter, I argue that sociolinguistic variation is important to consider for listener errors: we can make correct speaker-specific inferences about why utterances occurred, but our inferences can also lead us astray. I discuss examples like these in the 'Listener-centered approaches' section below.

In the third case, 'The *key* to the cabinets *are* rusty,' it is clear that an error occurred but it is unclear how to repair the utterance. Plural verbs like '*are*' sometimes occur in error in items like these, which might lead the listener to infer that the intended utterance was 'The key to the cabinets is rusty.' However, because the observed error occurs in a dependency between a noun and a verb, it would also be valid to infer that 'The keys to the cabinets are rusty' was the intended utterance. Errors like these have important consequences for the listener because it is unclear what should be inferred; I discuss this in the 'Listener-centered approaches' section below. It is also important to consider mismatches in understanding from a conversational perspective in order to examine how real-world dialogues are affected by error; I further discuss how conversational devices allow some degree of collaborative understanding in the 'Listener-centered approaches' section below.

The roadmap of this chapter is therefore as follows. I begin where speech error research began, sketching out the types of sound (phonetic), word (lexical), and morphosyntax errors from a typical speaker-centered language production perspective. As sketched out in Fig. 1, I describe the overall

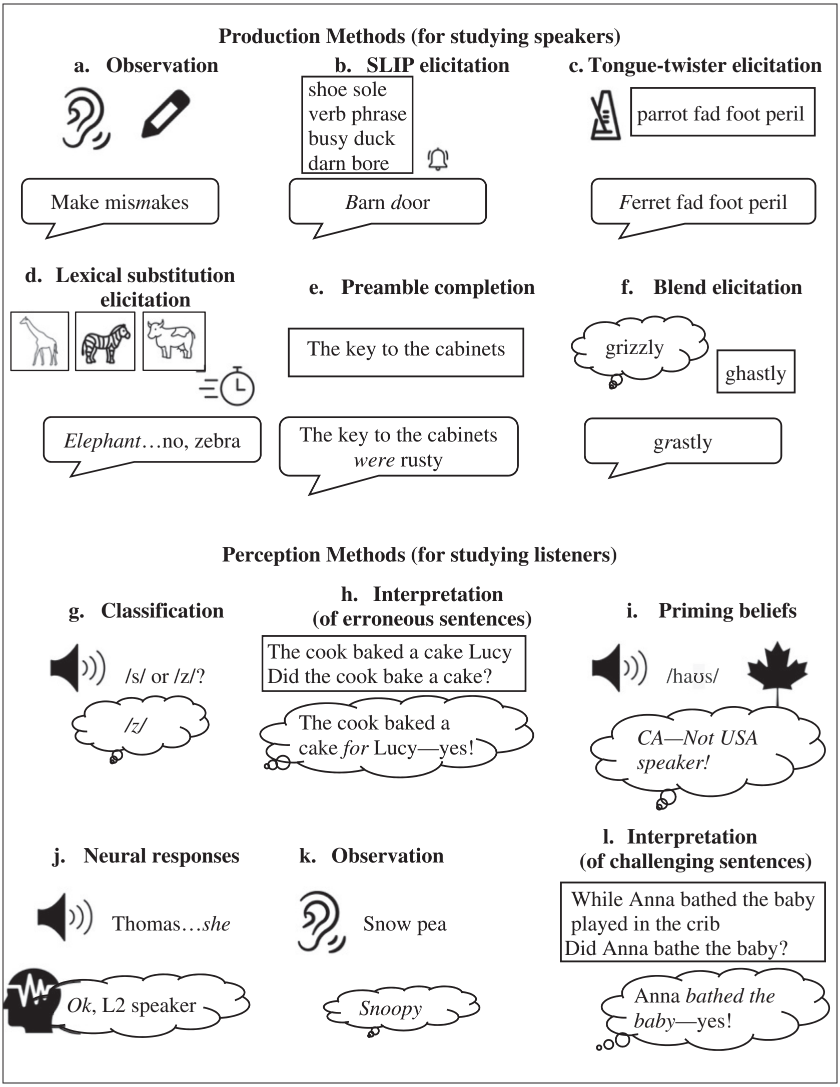

Fig. 1 A dozen ways to study speech errors. Visualizations of selected paradigms used in the reviewed literature to understand the production and perception of speech and language errors; the number 12 is specifically chosen in homage to Baars (1992). The top of each panel represents the paradigm; the bottom represents the response used as data. Italic text represents errors made by speakers and listeners. Items inside a rectangular box represent visual stimuli presented by computer, ear represents stimuli presented by an interlocutor, loudspeaker represents auditory stimuli presented by computer, and speech/thought bubbles reflect the speech/thought of the participant. All icons are taken from the Noun Project (www.thenounproject.com).

patterns, the methods used to elicit errors, and the theories or models built upon them. I then consider the role of the listener and of the conversational dyad in how errors affect communication, interfacing with work in sociolinguistics and conversation analysis. As I discuss, combining both perspectives is informative for considering how to define the concept of a speech error. I end with some suggestions for future work that integrates both perspectives, focusing on the conversational danger zone where listener and speaker do not come to the same understanding of what an utterance means.

2. Speaker-centered approaches: An error is what you didn't mean to say

2.1 From corpora to models of speech production

Early speech error research focused on building corpora: collections of naturally-observed errors. In these corpora, researchers like Meringer and Mayer (1895), Fromkin (1971), Garrett (1975), Nooteboom (1969), MacKay (1970), and Shattuck-Hufnagel (1979) documented the speech errors that they and their colleagues noticed in their daily lives. This type of paradigm appears in Fig. 1A. From error corpora, generalizations can be made about representations and timing in language production. If an error is observed in naturally occurring speech, it implies that (1) the unit(s) partaking in the error, such as phonetic features, sounds, words, and structures, are represented as entities in the mind, and that (2) the erring units are co-active in the mind at a particular point in time. This is, as many have noted, a much more sophisticated view on language errors than the view articulated by Freud (1901, reprinted 1958; as discussed in Dell, 1986 and Fromkin, 1971), but one that shares some commonalities: the things we don't mean to say sometimes get said anyway because they exist in our minds. However, from a psycholinguistic perspective, speech errors actually occur because of the nature of the mental representations of language and not because of any deep, dark, suppressed thoughts.

Fromkin (1971) describes the error patterns in a corpus that she collected; here, and throughout the chapter, I follow the field's convention and list out the theorized utterance target, followed by the observed error. Fromkin observes a variety of error types. The most common errors she observes are contextual errors—i.e., errors where the source is inside the utterance. These include anticipations, where the error intrudes from the future ('cup of coffee' → 'cu*ff* of coffee'), perseverations, where the error intrudes from the past ('Chomsky and Halle' → 'Chomsky and *Ch*alle'), and exchanges (also known as metathesis or spoonerisms), where two elements swap places ('the zipper is narrow' → 'the *n*ipper is *z*arrow'). There are also non-contextual errors, where the error does not depend on the content of the utterance. These include deletions of elements (also known as omissions or haplologies; 'tremendously' → '*tremenly*'), substitutions of an element from outside the current utterance ('chamber music' → 'chamber *maid*'), and blends of two often-similar elements (also known as portmanteaus; clarinet + viola → '*clarinola*').

Contextual and non-contextual errors can occur on phonemes (sounds), words, and morphology, reflecting the multi-stage representations we build in our minds in order to produce language (See Fig. 2). Studying the sites where errors occur provides a great deal of insight into the representation of language in the mind. Fromkin notes that phonemes can be the site of errors; consonant intruding on consonant ('week long race' → '*r*eek long race') and vowel intruding on vowel ('ad hoc' → '*o*dd h*a*ck'). This shows that phonemes exist as independent units in the mind, and that speakers obey the rules for combining them, even in errors. Phonemes occur in syllables,

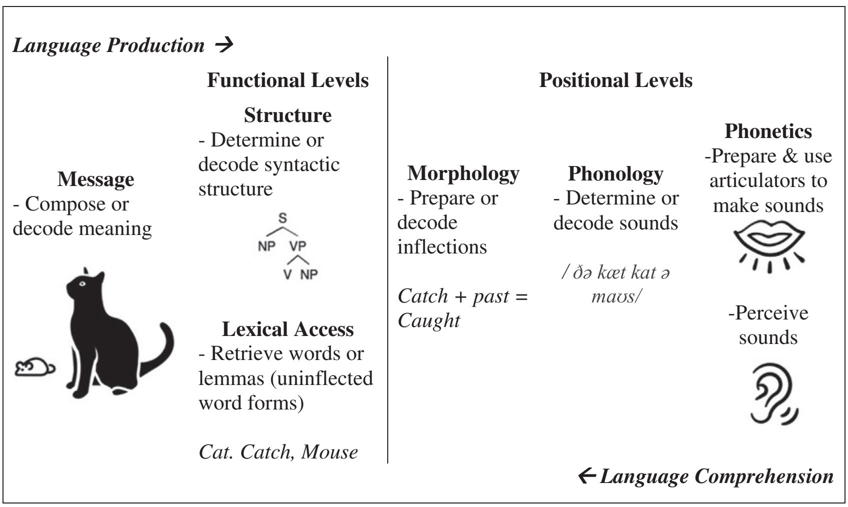

Fig. 2 A schematic of psycholinguistic levels of representation for speaking and listening for the utterance 'The cat caught a mouse.' This schematic draws upon Levelt (1989), Bock and Ferreira (2014), and Garrett (1975), adapted to consider both production and comprehension. For production, read the figure from left to right, and for comprehension, read it from right to left. Processes in the same column (structural encoding and lexical access) likely need to be done simultaneously. Pronunciation reflects the author's dialect. Icons are from thenounproject.com.

which leads to some constraint on where errors occur: because vowels and consonants live in different parts of a syllable, they therefore cannot be exchanged. Full syllables, though, do not tend to participate in English errors. This suggests that syllables may not be stored as independent units in the mind; though note that later cross-linguistic work (e.g., Alderete & O'Séaghdha, 2022) suggests that the syllable may be more important in other languages. Words are also frequently the site of errors, and here, word class tends to be the same for the target and error, such that nouns intrude on nouns ('bottom of page five' → '*bottle* of page five') and verbs intrude on verbs ('hand out' → '*hang* out'). This suggests that words are psychologically real elements, and also that syntactic categories, like nouns and verbs, are psychologically real; analogous to the vowel and consonant differences, these word classes have different roles in a sentence and cannot be exchanged. Finally, morphemes from complex words can also be the site of errors ('infinitive clauses' → 'infinit*y* clauses'), providing evidence that uninflected word forms, or 'lemmas,' are also psychologically real elements.

The implication is that all of these entities—phonemes, lemmas, sounds; consonants, vowels, nouns and verbs– are psychologically real, decomposable parts of language that are used in processing, and that different aspects of these representations are active at different stages in time, creating different levels of planning for production. Producing an utterance begins with the activation of a set of semantic and syntactic features that can capture a desired message, and ends with the activation of the muscles needed to articulate a sound, and errors tend to occur within a level. This is why errors tend to be, relatively speaking, well-formed: To paraphrase Fromkin herself, anomalous utterances are remarkably non-anomalous.

Next, Fromkin also notes the role of similarity in many of the errors from her corpus. There seems to be an occasional semantic (meaning-driven) component to lexical errors, where words are swapped for synonyms or antonyms ('This room is too damn *hot*…cold'), and where semantically similar items can end up fused together in a blend ('mainly/mostly' → '*monly*'). This implies that the word selection process is sensitive to meaning: We retrieve the lemmas we need by virtue of the meaning of the message we want to convey. Fromkin also outlines the role of phonetic similarity (similarities in how sounds sound or how they are produced) in phonemic errors. Consonants that share many phonetic features are more likely to participate in errors ('reveal' → 're*f*eal'). In this example, note that /v/ and /f/ are both produced in the same place in the mouth (as labio-dentals, with the front teeth and bottom pushed together) and in the same manner (as a fricative, or breathy, sound). While Fromkin suggested that the phonetic feature effect in errors is evidence that errors occur at the feature level (part of sound), not at the phoneme level (sound), later work (e.g., Shattuck-Hufnagel & Klatt, 1979) shows evidence that these seeming feature errors are most often better explained as phonemic errors. The implication, therefore, is that although phonetic features contribute to increasing similarity between phonemes and therefore increase the likelihood of errors, the erring unit is the entire sound. Similarity in both meaning and sound, therefore increases the chances of errors happening.

Garrett (1975) adds to Fromkin's analysis, noting that there are also systematicities in the distance of elements participating in errors. Errors more often than not occur with the source and the error site in the same clause (consider 'night life' → 'kni*fe* li*ght*' and 'give my *bath* a hot *back*'), and sound errors occur with the source and the error much closer to each other on average than word errors do (consider the two examples above compared to 'Every time I put one of these buttons *off*, another one comes *on*'). This leads Garrett to propose a clear separation between an early functional level of planning and a late positional level of planning. Lemma- and phrase-level errors, as well as blends, occur at the functional level, and sound errors and morpheme shifts arise at the positional level. This separation of sentence planning into functional and positional levels has had a remarkable influence on the field of sentence production in the intervening decades; see Vigliocco and Hartsuiker (2002) for extensive discussion on the necessity and separability of each level in the system based upon evidence from speech errors, Bock and Ferreira (2014) for a discussion of research in other domains of sentence production focusing on functional and positional-level effects, mainly in English, and Jaeger and Norcliffe (2009) for discussions on functional and positional-level effects through a cross-linguistic lens.

Dell and Reich (1981) highlight two additional effects that shed light on the role of similarity at various processing levels in eliciting errors in corpora. These are the lexical bias and the mixed error effect. The lexical bias is the finding that more errors form real words than would be expected randomly, as highlighted incidentally in the examples listed above (see also Baars, Motley, & MacKay, 1975). The mixed error effect is the finding that errors are often both phonologically and semantically related to the target; this includes blend errors (like 'mainly/mostly' → '*monly*'; Fromkin, 1971) and word substitutions or exchanges ('start' → '*stop*'; Dell, 1986). Two insights come from this observation. The first is a methodological innovation in determining what chance performance should be for errors of various types, which is necessary to determine in order to examine speech errors quantitatively rather than qualitatively. The chance rate for any given speech error is a value that depends on the elements participating in the error. For example, if the first consonant of 'red' is substituted with another consonant, many of the outcomes will be real words (bed, dead, fed…). In contrast, there is only one substitution of the first consonant of 'pants' that makes a real word (chants). This means that just by chance, we would expect more real word errors for 'red' than for 'pants,' and any differences between the two items need to be evaluated with respect to an item-specific baseline expectation. Dell and Reich (1981) use a scrambling procedure along these lines, recombining the onset consonants on the presumed targets observed in a corpus in order to estimate the chance expectation of specific outcomes in speech errors. With the appropriate baseline expectation set, it is clear that word outcomes are overrepresented compared to non-word outcomes in various corpora, that blend errors are more often both phonologically and semantically related than would be expected by chance, and that word exchanges have a phonological component more often than would be expected by chance. This leads to the second insight, and the finding the Dell and Reich (1981) paper is better known for: The independent planning stages outlined by Fromkin (1971) and Garrett (1975) may not be the best representation of speech error data. Instead, the production system likely has some interactivity between levels. Planning is not done fully in sequence, but with some interplay between stages.

Later extensions of these frameworks grew into fully implemented computational models. Dell (1986) outlines a connectionist modeling approach rooted in Garrett (1975) and his own earlier work with Reich. Connectionist modeling uses a simple type of neural network with representational units ('nodes') arranged in separate levels; for psycholinguistic models, these levels represent some or all of the stages in Fig. 2. Activation spreads between nodes based upon how strongly pairs of nodes are connected to each other, and activations are modulated by random noise and the decay of activation over time. Dell (1986) uses two discrete but interacting stages, one representing retrieval of lemmas from semantic features, and one representing retrieval of phonology from lemmas. The nodes within each of these two levels pass activation both backwards and forward through the network, building in interactivity into the framework. This model provides broad coverage for most of the observed speech error corpus data, including the lexical bias effect, phonotactic well-formedness factors including influence of syllable-bias and frequency co-occurrences on error patterns, and speech rate effects including the phenomenon that exchange errors occur more often in fast compared to slow speech. Later extensions of this model provide coverage of mixed errors and other semantic and phonological priming effects (Dell & O'Seaghdha, 1992), as well as the different profiles of error types across various patients with aphasia (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Foygel & Dell, 2000; Schwartz, Dell, Martin, Gahl, & Sobel, 2006).

In more recent years, there has been some debate in the field about whether a highly interactive two-stage model best describes other facets of production data, such as picture naming reaction times (e.g., Levelt et al., 1999). In fact, interactivity may not be necessary to describe normal picture naming (Levelt et al., 1999), and highly interactive models have less descriptive validity across a variety of types of error data than weakly interactive models (Rapp & Goldrick, 2000). There has also been debate on the number of levels needed to describe language production (e.g., the existence of the lemma as a psychologically important unit, discussed in Roelofs, 1992, Caramazza, 1997, and Caramazza & Miozzo, 1997), and on the degree to which discrete symbolic units (like those used in the Dell and Levelt models) are necessary to account for error data (see Smolensky, Goldrick, & Mathis, 2014; Vousden, Brown, & Harley, 2000). However, it is important to recognize that these models have more similarities than differences, and that what they share in common arose from the insights derived from error corpora. The general structure of a staged model beginning with a message, then word forms and structure, then morphology, then sounds is shared across models of language production. It is clear that all these processes have to happen before we can speak, and errors can arise in any of these processes because of the failure to retrieve the right elements or the failure to sequence elements in the right order.

2.2 Eliciting speech errors

Error corpora do have downsides, which has led researchers to develop several experimental paradigms to elicit speech errors. First, it is possible that error corpora are biased because a listener needs to recognize and then note down the error; these observed speech errors are therefore heavily influenced by the lens of a listener. As shown in the examples cited throughout this chapter, there are two general biases that occur in most error corpus data: errors tend to occur at the onset of words, and there are more anticipation than perseveration errors. Cutler (1981) suggests that these two phenomena may be critically connected with biases in what listeners are able to observe. Because words are identified starting from their onsets, onset errors may be particularly salient to listeners. Similarly, in a wide variety of language perception tasks, anticipations are notably easier to detect than perseverations (Cohen, 1980).

Next, as Stemberger (1992) highlights, it is also not always clear what the source of an error is. An error like 'night life' (/naIt laIf/) → 'kni*fe* li*ght*' (/naIf laIt/; Garrett, 1975) is typically analyzed as a sound error (the exchange of /t/ and /f/ consonants at the end of each word). However, the same error could arise because the speaker has retrieved two incorrect words from their lexicon. This is a case where the researcher cannot truly know what was intended, and corpora are unfortunately full of many more cases like these. Experimental paradigms can sidestep the error source issue by presenting stimuli that are specifically designed to elicit errors of one type and not another. Because the experimenter has a great deal of control with the stimuli and paradigm, cleaner generalizations can be made about the speaker's intent. With elicited errors, the experimenter can often determine for certain what the speaker's original target was, making it clear that elicited deviations from the target are indeed errors, and making it easier to reconstruct the processes that led to the error.

Elicited errors also have other advantages in terms of data control. Errors in corpora are naturally produced, which can mean that errors of a particular type are relatively rare. Eliciting these errors in the lab can be a better way to systematically study them (see, e.g., Meyer, 1992 for expansion on this reasoning). This can make error elicitation an ideal way to study the less common types of errors, including morphosyntactic errors and blends.

An early speech error elicitation paradigm is the SLIP ('Spoonerisms of Laboratory-Induced Predisposition) paradigm (Baars et al., 1975; Motley & Baars, 1976a; see Fig. 1B). In this paradigm, participants receive a series of quickly-presented word pairs ('sale receipt, verb phrase, shoe sole') that they are cued to produce aloud or to ignore. One or more interference pairs first appears ('comb hair,' 'cold nun'), followed by a target pair such as 'nosy cooks', which has one or more of the interference pair onset consonants exchanged between words. This promotes the elicited exchange error *'cozy nooks'* on the target trial at a rate of about 10% of trials, which is a fairly high elicited error rate. Much like in corpora, the SLIP paradigm elicits a lexical bias, such that word pairs that will slip to words ('darn bore → *'barn door'*) exchange more often than those that slip to non-words ('dart board' → *'bart doard'*; Baars et al., 1975). Much like corpora, the SLIP paradigm also elicits semantic effects: Stimuli are more likely to slip when primed by a synonym ('damp rifle' before 'get one' primes *'wet gun'*; Motley & Baars, 1976b). Finally, errors also seem to perpetuate, such that the same errors repeat over repeated trials (Humphreys, Menzies, & Lake, 2010). This highlights a potential difference between elicited error data and corpus data: Participants are always learning and adapting throughout experiments. This could be seen as either an advantage or a disadvantage, depending on the researcher's intentions.

The SLIP paradigm also sheds light on the role of self-monitoring (attending to whether one's own utterances are being produced as intended) as an error-avoiding mechanism. Stimuli where an exchange would elicit a taboo word ('tool kits' → *'cool tits'*) are less likely to slip than matched neutral items ('tool carts' → *'cool tarts'*), and even correct target productions of the taboo-eliciting item ('tool kits') lead to a large index of physiological arousal due to stress, as measured by galvanic skin response (a measure of skin conductivity; Motley, Camden, & Baars, 1981). This suggests that participants have the ability to avoid making these slips but that it requires effort to do so. A later examination of the exchanges in the SLIP paradigm that are noticed and halted or repaired (e.g., '*c*oo-, uh, tool'), versus those that are fully completed ('*cool tarts*', Nooteboom & Quené, 2008), confirms this finding: self-monitoring is likely an important factor in the SLIP paradigm. Later work following a corpus approach (Hartsuiker, 2006) confirms the plausibility of self-monitoring as a potential driver of other phenomena such as the lexical bias. Finally, various other experimental and computational studies further suggest that self-monitoring may be a generally critical component to speech production (see Nooteboom & Quené, 2017; Nozari & Dell, 2012 for evidence of monitoring in the tongue-twister paradigm, described further below, and see Nozari & Novick, 2017 for a comprehensive review of the role of monitoring in speech production). The insight from elicited errors that monitoring is important in production has therefore had a big influence on the field.

Tongue twister paradigms (e.g., Butterworth & Whittaker, 1980; Shattuck-Hufnagel, 1983, 1992; see Fig. 1C) are the other well-studied paradigm used to elicit phonological errors. These paradigms are rooted in the notion that well-known tongue twisters like the folk rhyme 'She sells sea shells by the sea shore' or many of Dr. Seuss's classic rhymes, such as 'Through three cheese trees three free fleas flew' (from 'Fox in Socks'; Seuss, 1965) are hard to say because of the repetition of similar sounds in the same syllabic position across the utterance, which leads to specific patterns of errors in performance. In tongue twister paradigms, individuals have to produce a set of four one or two syllable words quickly. The sequence of words is difficult because the words have repetition in the onset and/or end of the syllables; these can be words ('pick ton tick pun'; Sevald & Dell, 1994) or non-words ('shif sheev sif seev'; Acheson & MacDonald, 2009). A metronome is typically also used to cue the speaker to use a steady and quick speaking rate. While the task is notably constrained, research shows that errors in tongue twister paradigms seem to result from speech planning, not from effects of the task (Wilshire, 1999). Correspondingly, within this paradigm, similar data obtain as are found in corpora. Like in corpora, most tongue-twister errors are contextual (anticipations, perseverations, or exchanges elicited from somewhere else in the utterance) and appear in the same syllable position as in the target (e.g., 'parrot fad foot peril' → '*f*erret fad foot peril'; Shattuck-Hufnagel, 1992; Acheson & MacDonald, 2009). Like in corpora, tongue-twister errors tend to involve only one sound, typically a consonant, rather than a consonant and vowel (Shattuck-Hufnagel, 1983). Like in corpora, more tongue-twister errors tend to appear at the onsets of sequences than at their ends (Sevald & Dell, 1994). And, like in corpora, the tongue-twister errors that are made are relatively well-formed with respect to the phonotactics (phonological rules) of the language. This is shown most clearly in a variant of this paradigm using non-word sequences that follow a specific grammar where some consonants appear in constrained positions, creating an experimental phonotactic constraint and some follow language-wide phonotactics. An example of this is the tongue-twister 'hes feng neg kem' taken from an experiment where 'h' and 'f' are always onset consonants. Here, /h/ follows the English phonotactic pattern (in English, syllables never end with /h/), while /f/ is an experimental constraint differing from the phonotactics of English (in English, syllables can begin and end with /f/; Dell, Reed, Adams, & Meyer, 2000). This paradigm shows evidence for learned phonotactic constraints impacting the produced pattern of errors: If /f/ always appears in the onset within the experiment, it tends to appear more often in the onset in speakers' errors (Dell et al., 2000; Warker & Dell, 2006). Much like the SLIP paradigm then, the tongue-twister paradigm is useful for examining the guiding principles of sound errors and how these might be influenced by practice or other performance-related factors.

2.3 Moving beyond the phoneme: eliciting other types of errors

While the most prototypical speech error research focuses on phonological errors, we know from corpora that other types of errors occur: speakers also make mistakes in terms of morphosyntax and lexical retrieval. Laboratory-based paradigms have also had success in eliciting several of these error types. In the remainder of this section, I discuss the insights that researchers have drawn from eliciting lexical substitutions, subject-verb agreement errors, blend errors of various types, and the elicitation of errors from patients with language disorders.

Lexical substitutions are common in corpora ('start' → '*stop*'; Dell, 1986), and can also be elicited in the lab (see Fig. 1D). A paradigm used to elicit these errors is the description of complex arrangements of similar objects, such as multicolored circles (Levelt, 1983; Martin, Weisberg, & Saffran, 1989) or pictures of objects with names sharing semantic or phonological features (house, church, barn; school, square, scale; Martin et al., 1989). Lexical substitution errors are quickly noticed when they appear, and often repaired (Levelt, 1983), providing further evidence for self-monitoring as a core part of spoken production. Furthermore, more errors occur for semantically-related sets than phonologically-related sets (Martin et al., 1989), following the tendency in corpora for word errors to be related in meaning to the original target. Similar results obtain in a picture naming test designed to elicit errors with a speeded response deadline, especially when the objects are also visually similar (e.g., giraffe and zebra; Vitkovitch, Humphreys, & Lloyd-Jones, 1993), as well as in a speeded picture-word interference variant of the paradigm where participants have to produce the name of a picture while ignoring a written or auditory presentation of another related word (Starreveld & La Heij, 1999). Finally, a mixed error effect occurs for lexical substitution errors in a picture naming test designed to elicit errors by clustering items into related sets. In this task, participants simply have to produce the names of pictures presented one at a time. The errors made in this task are more likely to be semantically and phonologically similar to the target for both aphasic and non-aphasic participants (Martin, Gagnon, Schwartz, Dell, & Saffran, 1996), providing further evidence for the role of interactivity in error elicitation. The findings from lexical substitution paradigms therefore confirm that words can be substituted for a variety of reasons, with monitoring and interactivity both playing key roles in the system.

A few errors in the number inflections of verbs appear in the Fromkin (1971) corpus, but these are especially easy to elicit in the lab in a preamble-completion paradigm (e.g., Bock & Miller, 1991; Bock & Cutting, 1992; see Fig. 1E), leading to a relatively large literature on agreement error production (see e.g., Brehm, Cho, Smolensky, & Goldrick, 2022for a tabulation of the data from all English preamble completion studies). In this paradigm, a participant sees or hears a sentence fragment such as 'The key to the cabinets' and has to repeat back the fragment and add a completion to it, eliciting a correct target utterance like 'The key to the cabinets was rusty' or an error like 'The key to the cabinets *were* rusty.' Items like 'The key to the cabinets' which have a singular subject noun and a plural local noun (a noun near the verb) elicit more errors compared to items like 'The keys to the cabinet' which have a plural subject noun and singular local noun. This asymmetry has been well-documented (e.g., Bock & Cutting, 1992; Bock & Miller, 1991) and has been attributed to the 'markedness' of plural forms (Eberhard, Cutting, & Bock, 2005) or to language-wide structural frequency patterns (Brehm et al., 2022; Haskell, Thornton, & MacDonald, 2010). Moreover, unlike most of the phenomena mentioned in this chapter, there has been effort to elicit this error type across many languages with varied morpho-syntactic properties and degrees of inflectional richness. This includes Dutch (Antón-Méndez & Hartsuiker, 2010; Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001; Hartsuiker, Schriefers, Bock, & Kikstra, 2003), German (Hartsuiker et al., 2003); French (Franck, Vigliocco, & Nicol, 2002a, 2002b), Hebrew (Deutsch & Dank, 2009, 2011), Italian (Franck, Lassi, Frauenfelder, & Rizzi, 2006; Vigliocco, Butterworth, & Semenza, 1995), Portuguese (Acuña-Fariña, 2018), Russian (Lorimor, Bock, Zalkind, Sheyman, & Beard, 2008), Serbian (Mirković & MacDonald, 2013), and several dialects of Spanish (Acuña-Fariña, 2018; Bock, Carreiras, & Meseguer, 2012; Foote & Bock, 2012; Vigliocco, Butterworth, & Garrett, 1996). The conclusion is that morphosyntactic properties of the language might or might not influence agreement error patterns (see, e.g., Foote & Bock, 2012 vs Bock et al., 2012) and more data in more languages will be the key to success in this domain (so to speak).

Conceptual factors such as the notional (or semantic) number of the referent also play a role in subject-verb agreement errors, much like they do for phonological errors. These include variations in collectivity, distributivity, and notional plurality. Collective nouns ('gang,' 'staff'), which are typically used as singular nouns in American English, are particularly susceptible to plural agreement when used as subjects with plural local nouns ('The cast in the plays *were* performing'; Bock, Nicol, & Cutting, 1999; Haskell & MacDonald, 2003). Similarly, items with distributive referents that refer to multiple tokens of an item or multiple members of a set spread apart in space ('The label on the bottles'; 'The gang on the motorcycles') elicit more plural agreement than items with non-distributive ('The key to the cabinets') or collected ('The gang by the motorcycles') referents

(Eberhard, 1999; Humphreys & Bock, 2005; Vigliocco, 1996; Vigliocco et al., 1995), as do grammatically singular items that are rated to refer to more than one entity (Bock & Middleton, 2011; Brehm & Bock, 2013, 2017; Smith, Franck, & Tabor, 2018).

Much like the corpus data that led Garrett (1975) to sketch out functional and positional levels of sentence planning, clause structure factors also play a role in subject-verb agreement errors. Similar to what is observed in error corpora, subject-verb agreement errors occur most often when the two elements are in the same clause (Bock & Cutting, 1992). There is also evidence that the structural relationship of the two elements matters. Errors occur for items where the local noun is structurally, but not necessarily linearly, between the subject and the verb, and importantly, elicited errors like 'The helicopter for the flights *are* safe' and '*Are* the helicopter for the flights safe' both occur at similar rates (Vigliocco & Nicol, 1998). Syntactic distance also seems to be a better predictor of error rates than linear distance (Franck et al., 2002a, 2002b; see Franck et al., 2006 for an analysis of similar findings within generativist syntax principles). Other studies suggest that rather than syntactic structure, what matters is planning distance, which is intertwined with syntactic distance: As outlined in Garrett (1975), elements need to be planned with some degree of overlap in order to lead to errors. Gillespie and Pearlmutter (2011, 2013) show that these planning factors might matter more than structure does, and Solomon and Pearlmutter (2004) showed that parallelism in planning due to the arrangement of referents leads to an increased error rate (though c.f. Veenstra, Meyer, & Acheson, 2015).

Finally, there is emerging evidence that executive control also impacts agreement production. This parallels data from the SLIP paradigm because executive control is likely one of the factors that allows us to monitor our own speech. Individuals with poor inhibitory control produce more subject-verb agreement errors than those with strong inhibitory control (Veenstra, Antoniou, Katsos, & Kissine, 2018), and increasing processing load hinders both subject-verb agreement production (Hartsuiker & Barkhuysen, 2006) and processing (Vandierendonck, Loncke, Hartsuiker, & Desmet, 2018).

A third class of errors that can be elicited experimentally is blends (see Fig. 1F). These appear at a number of levels in the Fromkin (1971) and Garrett (1975) corpora, as well as in MacKay (1972) and Coppock (2010), including phonological blends, where the blending units are sounds ('Production+Perception' → 'Produ*p*tion'), lexical blends, where the blending units are words ('Mad as a bee+hornet' → 'Mad as a *beet*'), and phrasal blends, where the blending units are larger syntactic constituents ('Sharon H. will give you cookies + Sharon H. will give cookies to you' → 'Sharon H. will give you cookies *to you*'). Lexical blends can be elicited by asking participants to purposefully blend words together as they see fit (Wulff & Gries, 2019), or by asking participants to produce the non-presented member of a memorized pair (e.g., when 'ghastly' appears, produce 'grizzly'; as described in Baars, 1992). Syntactic blends, such as combinations of idioms ('That's the way the cookie crumbles + That's the way the ball bounces' → 'That's the way the *cookie bounces*') can also be elicited in the lab by requiring a speaker to produce one of two similar plans that are presented in close temporal sequence. Blends occur more easily with high syntactic and semantic similarity of the two alternate plans (see Cutting & Bock, 1997, for error elicitation, Konopka & Bock, 2009 for a priming-based paradigm, and Brehm & Goldrick, 2017, for a perception-based analogue). In both cases, presenting two competing and highly similar plans seems to enable the error to happen, affirming patterns observed in corpora and suggesting ideas to be examined experimentally in future research.

Finally, the bulk of the work discussed above focuses on the errors produced by typical, 'normal' individuals. One last sub-field of elicited speech error research focuses on the errors of patients with aphasia or other language disorders. The same logic holds as when studying elicited errors in typical individuals: errors are made because erring units are represented in the mind. Studying the errors made by patients can provide insight in to the representations for language in the typical mind that have gone awry (e.g., Ash et al., 2010; Buchwald & Miozzo, 2012; Caramazza, 1991; Dell et al., 1997; Gordon, 2002; Goldrick & Rapp, 2007; Harvey, Traut, and Middleton (2019); Nickels, 2014; Nickels & Howard, 1994; Martin et al., 1996; Pouplier & Hardcastle, 2005), and can be used to test altered ('lesioned') versions of computational models (Caramazza, 1997; Dell et al., 1997; Foygel & Dell, 2000).

3. Listener-centered approaches: Inferring what was meant

Let us now consider the listener more explicitly. One notable property of the Fromkin corpus (Fromkin, 2022, available online at https://www.mpi.nl/dbmpi/sedb/sperco_form4.pl) is that the observers are extremely confident about the target utterances that were intended by the speakers. I quantified this within the 3661 errors that had a value of 'YES' or 'NO'

Table 2 Errors with unclear to the listener targets in the Fromkin corpus. These are defined as items where the target contains '(?)' or '/.'

	Blends	Other Error Types	Total
Self-Corrected	12	6	18
Not Self-Corrected	35	12	47
Total	47	18	65

within the 'Corrected' column of the corpus, indicating whether the speaker ultimately noticed their error. In this data set, I then coded whether the target was unclear to the listener. To do this, I searched for uncertainty markers within the text of the target utterances: either '/,' indicating multiple target plans, or '(?),' indicating the coder's uncertainty about the target plan. There were only 65 utterances meeting this criterion. These are tabulated in Table 2. This small number means that the intent of over 98% of the utterances in this database was clear to the listener. With the caveat that the listeners in this data set are comprised of expert psycholinguists, the pattern is quite apparent: errors are usually easy to understand. This is the case even if the exact mechanism causing the error is sometimes ambiguous (as discussed by Stemberger, 1992).

Looking within the unclear-to-listener utterances, it also becomes apparent that the majority of these are blends. For these utterances, it is unclear to the listener whether one of the alternate plans was intended, with the other intruding, or whether the speaker could not decide between the two. This means that despite the lack of a clear target, the meaning of the utterance is still clear. This leaves only 18 utterances of 3661 wherein the intended message was truly not clear. This is a vanishingly small proportion. Most speech errors are understood by the listener.

However, a critical divergence occurs when considering mismatches between speaker and listener. Most of the unclear-to-listener utterances also go unnoticed by the speaker. About 62% ($N=47$) of these utterances are also not corrected by the speaker, indicating that the speaker did not notice them; the proportion of unnoticed errors is also higher for blends (35/47, 74%) than for other error types (6/12, 50%). From this we can conclude that speech errors have relatively little consequence for a listener's understanding—but also that when the listener does not understand, the speaker is also quite likely to not notice.

This analysis highlights the importance of considering listener-centered dynamics in understanding speech errors: somehow, listeners are usually skilled at recovering meaning from errors, but listeners and speakers are also not always in alignment about what an utterance means. In the rest of this section, I outline areas of psycholinguistics, sociolinguistics, and conversation analysis that interface with this question and discuss the implications of these findings for speech error research.

3.1 Repairs in the ear and the mind

Error-containing utterances have been studied with respect to their phonetic properties, how listeners tend to interpret them, and the ways that listeners' expectations change what is understood. This work suggests that a listener's understanding is often preserved in the context of speech errors for multiple reasons and that listeners are good at identifying and repairing most of the errors they hear.

First, phonetic research shows that speech errors do not sound exactly like utterances produced as targets. This makes it occasionally possible to tell what is an error by the phonetics of the utterance alone (as shown in Fig. 1G). A /p/ produced in lieu of a /b/ will sound a bit more like a /b/, with a reduced voice onset time. This is true for elicited errors (Goldrick & Blumstein, 2006) and for spontaneously produced errors (Alderete, Baese-Berk, Leung, & Goldrick, 2021). Similarly, a /z/ produced instead of a /s/ differs noticeably in duration, frication amplitude, and voice onset time, making it sound different than a /z/ produced as intended. This difference is occasionally noticeable to trained listeners (Frisch & Wright, 2002). For speech errors that change place of articulation (/t/ → /k/), the observed articulation seems to combine the phonetic gestures (movements of the mouth) that would be used to produce the error and for the target (Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007). Similar gestural distortions are also noticeable in tongue twister completions, regardless of whether the utterance is classified as an error or a correct target (McMillan & Corley, 2010). Moreover, error tokens of /t/, /s/, and /ʃ/ from a tongue twister study are perceptually distinct from correct targets, measured in terms of classification accuracy and reaction time; this even holds for tokens that do not vary from the target categorically (i.e., they are still considered to be examples of /t/, /s/, or /ʃ/; Pouplier & Goldstein, 2005). A final data set comes from comparing the classification of error and correct tokens of the same phonetic categories that were extracted from longer tongue twister productions and either did or did not include speaker self-corrections. In this set of items, listeners find it more difficult to identify the phonetic category of error tokens than correct tokens in terms of accuracy and reaction times, especially for those errors that were not later self-corrected by the speaker (Nooteboom & Quené, 2013). Combined, this research suggests that there are many subtle ways that errors sound uniquely like errors.

Next, recent work falling under the umbrella of noisy channel theory shows that comprehenders often seamlessly convert erroneous or otherwise unexpected utterances into correct productions in their own minds (see Fig. 1H). This has been repeatedly demonstrated for syntactic errors that are created by adding or removing one word in an utterance (see Gibson et al., 2013; Levy, 2008; and Levy, Bicknell, Slattery, & Rayner, 2009). Many studies in this area use a sentence comprehension paradigm where participants read a sentence (i.e. 'The chef baked a cake Lucy') then answer a yes/no comprehension question (i.e. 'Did the chef bake a cake?'). The answer to this question provides information about the sentence interpretation. The answer 'yes' indicates that 'cake' is assigned as the grammatical theme of the sentence, consistent with a non-literal reading of the sentence as 'The chef baked a cake *for* Lucy.' In this paradigm, highly implausible utterances that might be interpreted as containing errors like 'The chef *baked a cake Lucy*' are commonly interpreted such that that missing elements seem to be replaced ('The chef baked a cake *for* Lucy') and added elements seem to be removed ('The chef baked Lucy *for* a cake' → 'The chef baked Lucy a cake'; Gibson et al., 2013). This result is interpreted following a Bayesian belief model: A comprehender infers that an error has occurred and repairs the string if the repair is deemed more likely than the error in the given context, doing so proportionally more often when the error requires only a small edit. Furthermore, errors are deemed to be more or less likely by the comprehender because of context. Increasing rates of irrelevant errors in the experiment by adding filler sentences increases the percept of errors, while increasing the rate of implausible sentences in the experiment decreases the percept of errors (Gibson et al., 2013). Beliefs about which errors are likely also seem to be specific to the expected rates of insertions, exchanges, and deletions produced by a particular speaker in a particular context (Ryskin, Futrell, Kiran, & Gibson, 2018). This shows how comprehenders are skilled at tuning their predictions to the situation. The error repair process is also observable in data from eye-tracking while reading, a paradigm in which the researcher uses a specialized camera to measure a participant's eye movements while they read. Results from this paradigm show that readers seem to maintain uncertainty about possible syntactic structures in unfolding sentences, therefore causing them to go back and re-read earlier words once they have encountered something odd or unexpected (Levy et al., 2009). And, importantly, the repaired structure in sentences is just as psychologically present as the structure of sentences that contained no errors, and is able to influence later productions by syntactic priming (Buxó-Lugo & Slevc, 2022; Cai, Zhao, & Pickering, 2022).

Next, blends of two sentence formulations, which occur in corpora and elicited production experiments, also seem to be surprisingly easy to repair. Like lexical substitution errors, the likelihood of reconstructing blends is proportional to the likelihood of the blend occurring in the first place. Frazier and Clifton Jr (2011, 2015) outline the interpretation of syntactic blend sentences including 'I just like the way the president looks *without* his shirt *off*' (*without* and *off* are two ways of conveying similar information that negate each other), and 'A passerby *rescued* a child from *almost* being run over by a bus' (a blend of formulations including 'rescued' and 'almost hit'). Judgments about the naturalness of the sentence (Frazier & Clifton Jr, 2015) and answers to comprehension questions like 'What did the sentence mean?' (Brehm, Jackson, & Miller, 2019; Frazier & Clifton, 2011, Frazier & Clifton, 2015) correlate, indicating that when the utterance was considered to be an error, it was also more likely to be processed non-literally. In other words: comprehenders are skilled at recovering the meaning of the utterance that is perceived to be most likely to match what the speaker intended. Similar results obtain in a visual world eye-tracking paradigm, where images representing items in the sentence appear on a computer screen and visual attention to these items is tracked over time as an auditory stimulus ('…without her shoes *off*'…) is presented to a listener (Brehm, Jackson, & Miller, 2021). This paradigm further corroborates that the interpretation of blend sentences seems to arise from an 'unblending' process: The two meanings for the '*without* + *off*' utterances seem to be considered to various degrees over time depending on the unfolding sentence.

Morphosyntactic errors are also fairly easily repaired by readers and listeners; however, here, comprehenders (readers/listeners) and speakers do not always form the same interpretation of an utterance. Making a nice parallel to the production literature, subject-verb agreement errors cause relatively little processing slowdown in reading studies (e.g., Lago, Shalom, Sigman, Lau, & Phillips, 2015; Wagers, Lau, & Phillips, 2009). Similarly, items presented in a sentence comprehension paradigm are more often repaired as reflecting an error in verb inflection than in noun inflection

('The key to the cabinet *were* rusty' → The key to the cabinet was rusty,' as measured by responses to the comprehension question 'Was there more than one key?'; Patson & Husband, 2016; Brehm et al., 2019). However, comprehenders also reinterpret plural-verb containing sentences that are fully grammatical ('The key to the cabinets *was* rusty' → 'The keys to the cabinets *were'* rusty'), as shown in a sentence comprehension paradigm (Patson & Husband, 2016; Brehm et al., 2019) and in visual world eye-tracking (Brehm et al., 2021). These lines of research suggest that comprehenders also use expectations about likely errors to infer that errors might be present when they are actually absent. Combined, this means that listeners may not always successfully repair errors to get the interpretation of the utterance that the speaker intended. For agreement errors in particular, speaker and listener come to disagreement about utterance meaning reasonably often.

In addition, the repair process undertaken for subject-verb agreement errors may be somewhat specific and occasionally quite shallow. First, it is clear that when a repair is made, it typically focuses on recasting the number of the head, not the local noun, and any repairs that are made seem to have limited consequences for later parts of a discourse. The lack of attention to the local noun has been shown in a visual world eye-tracking paradigm, where subject-verb agreement violations are associated with fixations to tokens representing the head (one key, two keys) and not the local noun (one cabinet, two cabinets; Brehm et al., 2021). It has also been shown that reading times on verbs embedded in attraction-eliciting sentences ('The boy by the tree *were*') are not associated with the reinterpretation of the sentence subject as indexed by selection of a sentence-completing adjective ('chubby/green'; Schlueter, Parker, & Lau, 2019). The surprising shallowness of repairs in these sentences is also observed in a paradigm where sentences containing attraction eliciting phrases ('The newspaper with the perfume ads') do not facilitate later reuse of a repaired plural version of the referent ('newspapers') in a subsequent sentence (Dempsey, Christianson, & Tanner, 2022). These findings suggest that repairs to subject-verb agreement errors reflect a fairly specific but shallow process: comprehenders seem to maintain uncertainty about the true number of the head noun, but do not commit to these repairs as deeply as they do other morphosyntactic repairs. The nature of this error as a dependency may be the reason for this difference, or possibly, listeners may be aware of the potential mismatches in interpretation in these utterances between speaker and listener.

3.2 Speaker-specific repair processes

For many types of errors, comprehenders' interpretations are also to some extent speaker-specific: utterances are given different meanings depending on who produced them. This provides a critical bridge between psycholinguistics and work on sociolinguistic variation. The field of sociolinguistics shows that individuals of different dialects vary in the morphosyntactic structures they use (Kortmann & Szmrecsanyi, 2004; Labov, 1968; Wolfram & Christian, 1976) and their phonetic realization of specific sounds (e.g., Labov, 1973, 1994). Recent research shows that comprehenders have a demonstrated ability to use these factors in their understanding of what utterances mean. This is clear in a variety of paradigms.

First, speaker-specific repair or accommodation inferences are clearly demonstrated in a variety of priming-based paradigms. As shown in Fig. 1I, these paradigms prime a particular interpretation of an utterance implicitly with speaker accent, by explicitly telling the listener what dialect the speaker has, or with visual cues such as a photograph of the speaker. Accents and dialects have wide-ranging effects on the perception of sounds in these paradigms. Individuals from Michigan and Ontario, Canada typically have a similar accent, producing a raised [aʊ] 'MOUTH' vowel. However, the accent is more stereotypically associated with Canada than Michigan, even for people from Michigan. This leads to differences in perception: when rated by Michiganders, raised [aʊ] vowels attributed to Canadians are perceived as more raised than when they are attributed to other Michiganders (Niedzielski, 1999). Similarly, utterances like /mIs/ for 'mist' that are more likely for Black speakers because they are associated with African-American Vernacular English are more likely to be attributed to Black rather than White speakers (Staum-Casasanto, 2008). For speakers who do not have the NEAR/SQUARE vowel merger associated with youth in New Zealand English ([tʃiə] = 'chair' or 'cheer'), the assignment of vowels to these two categories is mediated by the perceived age of the speaker (cued via photograph; Hay, Warren, & Drager, 2006). These data show how expectations about what a speaker 'should' say can cause listeners to form accurate interpretations when the speaker matches stereotypical expectations, but can also cause listeners to form inaccurate interpretations when the speaker mismatches stereotypical expectations.

Accents and dialects also change the perception of whole utterances or lexical items. This has a dark side: it can cause listeners to hear errors more often for non-standard speakers. This creates potential mismatches between speakers and listeners. Non-native accents increase the percept of speaker errors in sentence comprehension paradigms (Brehm et al., 2019; Gibson et al., 2017), as do accents coming from regional dialects (central Pennsylvania, Brehm et al., 2019).

However, more often than not, listeners seem to be able to account for speaker variations and form expectations accordingly. This typically keeps listeners in line with the intent of speakers. Common regional variations in agreement associated with specific American dialects such as Southern American English, New York City English, and African American Vernacular English ('The turtle *don't*) are perceived as less erroneous than rare ones that are not associated with a specific dialect ('The turtles *doesn't*), receiving less slow-down in a self-paced reading paradigm (Squires, 2014). In an elicited imitation paradigm where a participant has to repeat sentences presented to them, German determiner case inflections are produced to match the register of the speech (casual vs formal). The form 'wegen $des_{genetive}$ Regens' (because of the rain) is considered standard. This always receives an accurate repetition, no matter the register it is presented in, but the variant 'wegen dem_{dative} Regen' receives an accurate repetition in casual but not formal speech (Engel & Hanulíková, 2020). This indicates that the dative form of the determiner is not considered an error in casual speech. Finally, the meanings of words that vary across dialects of English (e.g., 'bonnet,' meaning a hat in American English and the hood of a car in British English) are more likely to be understood with the typical meaning from the accent producing the word (referring to a hat for an American English speaker and the hood of a car for a British English speaker, Cai, 2022; Cai et al., 2017). These studies show that accent can be an extremely reliable cue for a listener to select the appropriate meaning of a word or structure based upon the person speaking.

Similar speaker accommodation phenomena appear as measured by electrical activity associated with neural responses in paradigms such as ERP ('Event related potential'—a time-sensitive method for measuring neural reactions to stimuli; this is shown in Fig. 1J). Sound reduction phenomena in Dutch that eliminate overt gender agreement marking in fast speech show reduced neural responses associated with error processing when the stimulus is presented quickly ('een $spannend_N$ $roman_C$,' a $suspenseful_N$ $novel_C$; Viebahn, Ernestus, & McQueen, 2017). The types of gender agreement errors in Dutch that are expected of second language speakers also show reduced neural responses associated with error processing when presented in a non-native accent ('het_N $cultuur_C$,' the_N $culture_C$; Hanulíková, Van

Alphen, Van Goch, & Weber, 2012 and Lemhöfer, Schriefers, & Indefrey, 2020; 'Thomas was planning to attend the meeting but *she* missed the bus to school;' Grey & van Hell, 2017). Unacceptability can also be triggered by register ('I have a large tattoo on my back,' spoken by a very posh voice; Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008). Finally, these expectations seem to be adaptable over time. Given context about a specific speaker ('Susan works on an ox farm'), atypical exemplars of a category later produced by that speaker – i.e., in response to 'name a farm animal,' Susan replies 'ox' and not 'cow'—show reduced neural responses associated with unacceptability (Ryskin, Ng, Mimnaugh, Brown-Schmidt, & Federmeier, 2020).

These findings combined show that many facets of speaker identity can be used by a listener to make inferences about the meaning of an utterance. While some of these inferences are wrong, causing listeners to make repairs to utterances that were produced as intended, many of them are accurate, causing listeners to accurately recover the intended meaning.

3.3 Making errors in comprehension: slips, Good Enough Processing, and misperceived variation

Early psycholinguistic investigations of listeners and speech errors focused on the analogue of slips of the tongue: so-called 'slips of the ear' (see Fig. 1K). These mismatches between listener and speaker are reported by Meringer and Mayer (1895) in their corpus, and also occur frequently in listener's perceptions of song lyrics, where they are often called 'mondegreens' (Wright, 1954). A classic example of this is the commonly mis-heard 'scuse me while I kiss *this guy'* in Jimi Hendrix song Purple Haze. This perception error follows from the sound and meaning of the utterance. Here, 'the' and 'sky' are heavily reduced and co-articulated, making the original target unclear, and the sounds /g/ and /k/ differ by only one feature, making the misperceived slip a good match to the sound that was heard. In addition, the slip has a well-formed meaning, since it is much more common for a person to kiss a guy than a sky.

Celce-Murcia (1980)'s analysis of the Meringer and Mayer errors highlights the many circumstances in which perception errors occur, including from phonological errors in the perception of consonants or unstressed syllables, and how they are impacted by expectations about what the speaker is likely to say, and differences in dialect between speaker and listener. Browman (1980) and Garnes and Bond (1980) reinforce that perception errors often occur in unstressed syllables, and Browman (1980) notes that they are often flagged in the discourse by a 'what?' or other repair-requesting response from the listener. Cutler (1990) reports that not only do perception errors occur most often in unstressed syllables, but that the slip itself also follows a predictable metrical structure. Specifically, word boundaries in slips move to more probable locations for the language's typical stress pattern (e.g., 'Coke and a Danish' → '*Coconut* Danish'). Finally, Bond (2021) provides a comprehensive review of common perception errors. In addition to the stress phenomena noted above, she also remarks that perception errors seem to frequently involve homophones ('nones' → '*nuns*'). Perception errors are also more frequent for vowels and liquid sounds ('snow pea' → '*Snoopy*') than for consonants, perhaps because the sound categories for vowels and liquids are more 'fuzzy' or malleable. The implication from these analyses is that slips of the ear follow from the dynamics of lexical access: Sometimes the wrong item will be retrieved when listening if the signal to retrieve the correct item is perceptually weak because of acoustics (for most word mis-hearings) or because another lexical item or set of lexical items fits equally well with the string (for homophones and stress shifts).

Another set of recent psycholinguistic investigations on individuals' misunderstanding of speakers falls within the Good-Enough Processing framework (see Fig. 1L). Challenging but fully grammatical 'garden path' sentences like 'While Anna bathed the baby played in the crib' (Christianson et al., 2001) lead to the syntactically unlicensed percept of 'Anna bathed the baby.' This is shown with positive responses ('yes') to comprehension questions like 'Did Anna bathe the baby?.' Similarly, implausible but fully grammatical sentences like 'It was the mouse the cheese bit' (Ferreira, 2003) lead to the syntactically unlicensed percept that the mouse is the agent of the sentence (the actor doing the biting action). Implausible active sentences like 'The fish caught the angler' also serve to prime later passive production, enhancing productions of later utterances like 'The girl was pinched by the boy.' This shows a passive structure was accessed in comprehending the implausible sentence (Christianson, Luke, & Ferreira, 2010). Further evidence from reading studies suggests that an incorrect initial parse can create later difficulty in processing a correct structure (Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013) or meaning of a sentence (Huang & Ferreira, 2021), suggesting that comprehenders may sometimes build the wrong structure of an utterance to start, and then fail to revise it when they need to do so later. However, examination of challenging garden path sentences using an ERP paradigm shows that neural responses consistent with confusion do not predict answers to comprehension questions consistent with reanalysis, suggesting that individuals may just feel remarkably unsure about the meaning of these sentences, rather than accessing an incorrect structure (Qian, Garnsey, & Christianson, 2018). Combined, the most plausible interpretation of this evidence is that comprehenders sometimes fail to form full syntactic representations of sentences: fast, but not entirely accurate, heuristics about what is plausible and expectations from prior sentence comprehension often drive interpretation (see Ferreira & Patson, 2007 and Karimi & Ferreira, 2016 for reviews). This can create large mismatches between what is said by a speaker and what is understood by a listener.

As with repairs in comprehension, errors in comprehension also seem to follow from sociolinguistic factors. This can lead to misalignments between speaker and listener. Celce-Murcia (1980) notes that dialect differences between speaker and listener are predictive of slips of the ear, and Bond (2021) shows that slips of the ear can reverse or create the consonant cluster reductions that are present in some dialects ('I just like it' → 'I *dislike* it'); these perception errors are therefore perhaps in part driven by incorrect expectations about the speaker's accent. Experience with one accent also changes the phonetic expectations for another (Hay, Drager, & Warren, 2010), which may leave a listener with incorrect expectations for a later speaker. Next, in striking contrast to the evidence reviewed in the section above, hearing an utterance in the appropriate dialect also does not always eliminate a neural error response. This seems to be most common with lower-prestige dialects, such as 'The clown, he blowing up balloons at the party,' which is grammatical when spoken by an African-American Vernacular English speaker (Weissler & Brennan, 2020), or 'He might could go,' which is grammatical when spoken by a Southern American English speaker (Zaharchuk, Shevlin, & van Hell, 2021). Finally, agreement variations associated with lower-status dialects of American English ('It don't') are more likely to be attributed to lower-status speakers, but this interpretation is not easily primed with the social identity of the speaker (Squires, 2013). This suggests that listeners may have trouble overriding past experiences, including what the standard dialect of their language predicts. This is the case even when it is clear that the listener needs to do so. The degree to which listeners make correct or erroneous predictions about the intent of their interlocutor is still unknown, as this area of research is still growing, but the potential impact of these miscommunications should not be understated since sociolinguistic attitudes intersect so strongly with power, prestige, and privilege (see e.g., Craft et al., 2020).

3.4 An error can be repaired collaboratively

Finally, a listener's contribution to a conversation is also often quite important. In naturalistic conversation, listeners often give speakers feedback, and this can serve to repair any mismatches in understanding between speaker and listener. Conversation analysis, combining the observation methods used in Fig. 1A and K, gives insight into how speakers and listeners navigate errors in conversation collaboratively. First, speakers notice and repair their own errors in conversation; this is even apparent in the disconnected utterances cataloged by Fromkin (1971) and Garrett (1975). The dynamics of repair are discussed at length in the conversation analysis literature (e.g., Hieke, 1981; Schegloff, 1979; Schegloff, Jefferson, & Sacks, 1977). First, repairs seem to be orderly sequences that follow a set of conversational rules. Repairs can be initiated by the error-uttering speaker or their listener, and repairs and can correct nearly any mistake in the conversation, including errors in phonology, word choice and syntax. Repairs seem to be associated strongly with perceived misunderstandings, and therefore also often occur when one individual is confused despite no obvious speech error in the discourse. Repairs also have specific phonological and structural properties. Self-initiated repairs are often preceded by particular phonological elements like glottal stops, pauses, or cut-off sounds, and occur as either single words or full phrases (Schegloff, 1979). Other-initiated repairs typically occur in the turn following the error, and these could be a simple confusion marker ('huh?') directing the speaker to correct themself, a repeat of information, or a correction offered to the speaker, as in this utterance pair, where the error is italicized and the repair is underlined: 'Martha: Tell me, uh what- d'you need a *hot* sauce? Agnes: t'hhh a <u>Taco</u> sauce. (Schegloff, 1979). In English, listeners seem to be quite eager to offer a correction or replacement to the original speaker (Kendrick, 2015). However, the preferred type of other-initiated repair can differ across languages. Some other common forms of other initiated-repairs in a diverse variety of languages (here all translated into English) include formulaic expression ('sorry?'), a restricted offer repair (A: 'I work polishing glass.' B: 'of cars?') or an alternative question (A: 'ten photographs' B: 'his or someone else's?' A: 'No, no, someone else's; Dingemanse & Enfield, 2015). Repairs also serve a social function: In multi-party conversations, offering a repair of someone else's utterance seem to allow a speaker to join or leave an ongoing conversation (Egbert, 1997). Combined, this means that listeners are remarkably proactive in avoiding errors or misunderstandings in real life conversations. The error documenters in the Fromkin (1971) corpus probably did understand nearly all the errors that they observed.

4. Future Directions: Integrating multiple perspectives

The research reviewed through the first sections of this chapter highlighted the alignment between speaker and listener: speakers make errors in predictable ways, and these can often be repaired by listeners. However, in the final sections, I also noted that listeners can also be remarkably wrong about what the speaker originally intended. This can happen when the error produced can be repaired in more than one way, when the target produced is particularly challenging or otherwise atypical, and when the speaker has a low-prestige dialect. This creates a conversational danger zone which is especially important to understand in order to show how the basic research on speech errors translates to more complex real-life conversations. I highlight three potential future directions for the field here relating to this conversational danger zone.

First, it is important to understand more about the types of situations when listeners fail to match the speaker's intent. This is critical for understanding errors in conversation, but will also shed light on what enters a listener's mind when repairing errors: How much do we use our own experiences in speaking and in listening to infer others' intended meaning? A direct comparison of the same phenomena in an error-elicitation paradigm and in a comprehension paradigm would be an excellent start to answering this question. It would also be useful to contrast errors in dependencies, which can be repaired in more than one way, with similarly-sized errors that there is only one clear way to repair. This will shed light on how listeners chose which repair to make given multiple possible options.

Second, the speaker-specific error processing literature suggests that prestigious dialects may be treated differently than stigmatized dialects. In particular, stronger speaker-specific effects have been shown for register differences and for high-prestige dialects than for lower-prestige dialects like African-American Vernacular English and Southern American English. Much more research is needed on this topic. First, it is necessary to examine a wider variety of stigmatized or lower-prestige language varieties to show whether the apparent connection between prestige and speaker-specific error repair is true. If the same pattern obtains in a larger sample, it then becomes critical to uncover the mechanisms at play: Why does prestige matter for perception? For example, are the listener's attitudes toward the speaker involved in their inaccurate expectations? Can experience help us to form accurate expectations? What interventions can be made for listeners in order to prevent expectation mismatches from occurring ?

Finally, it is important to note that all new words ('dynamical,' 'stan') have the potential to be treated as phonological errors before their meanings are learned; these two examples are items that the author first perceived as typos before later learning their meaning. Experimental research focusing on change in the perception or production of errors over time would be useful to show how we acquire new items. A learning-based perspective that focuses on the conditions under which new items are acquired would also provide insight into speaker-listener mismatches and what gets treated as an error in the first place.

5. Conclusion: What *is* an error anyway?

We now return to the definitions that I raised in the first section of this chapter. As stated in the introduction, the speaker perspective is that an error is an utterance that was not intended. However, the research reviewed here underlines that what was intended is not always obvious or easy to determine. Based upon the research reviewed in this chapter, the listener perspective should be similarly defined: An error is an utterance that does not match the intent of the speaker. However, the work reviewed here also highlights that listeners sometimes have trouble making the correct inferences, and that they make their own errors in perception. This means that a deep understanding of the intention-inferring process is absolutely critical for understanding errors in both production and perception. As such, it is especially important for listeners (including speech error researchers) to consider the properties of the speaker and the speaker's context in defining when an error has occurred. Speaker-listener mismatches are, fortunately uncommon: the evidence suggests that speakers and listeners are remarkably good at uncovering a speaker's intent in real-life scenarios, meaning that while speech errors are common in every-day speech, they do not typically hinder understanding. However, speaker-listener mismatches are especially important to understand in light of issues of prestige and power. Here, as in many cases, systemic biases conspire against those who have the least agency in society.

References

Acheson, D. J., & MacDonald, M. C. (2009). Twisting tongues and memories: Explorations of the relationship between language production and verbal working memory. *Journal of Memory and Language*, *60*(3), 329–350.

Acuña-Fariña, J. C. (2018). The role of morphology in setting production biases in agreement: A cross-linguistic completion study. *Language Sciences*, *66*, 28–41.

Alderete, J., Baese-Berk, M., Leung, K., & Goldrick, M. (2021). Cascading activation in phonological planning and articulation: Evidence from spontaneous speech errors. *Cognition*, *210*, 104577.

Alderete, J., & O'Séaghdha, P. G. (2022). Language generality in phonological encoding: Moving beyond Indo-European languages. *Language and Linguistics Compass*, *16*(7), e12469.

Antón-Méndez, I., & Hartsuiker, R. J. (2010). Morphophonological and conceptual effects on Dutch subject-verb agreement. *Language and Cognitive Processes*, *25*(5), 728–748.

Ash, S., McMillan, C., Gunawardena, D., Avants, B., Morgan, B., Khan, A., et al. (2010). Speech errors in progressive non-fluent aphasia. *Brain and Language*, *113*(1), 13–20.

Baars, B. J. (1992). A dozen competing-plans techniques for inducing predictable slips in speech and action. In *Experimental slips and human error* (pp. 129–150). Boston, MA: Springer.

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, *14*(4), 382–391.

Bock, K., Carreiras, M., & Meseguer, E. (2012). Number meaning and number grammar in English and Spanish. *Journal of Memory and Language*, *66*(1), 17–37.

Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, *31*(1), 99–127.

Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology*, *43*(2), 83–128.

Bock, K., & Ferreira, V. S. (2014). Syntactically speaking. In *The Oxford Handbook of Language Production*.

Bock, K., & Middleton, E. L. (2011). Reaching agreement. *Natural Language & Linguistic Theory*, *29*(4), 1033–1069.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*(1), 45–93.

Bock, K., Nicol, J., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and Language*, *40*(3), 330–346.

Bond, Z. S. (2021). Slips of the ear. In J. S. Pardo, L. C. Nygaard, R. E. Remez, & D. B. Pisoni (Eds.), *The handbook of speech perception, Hoboken*. Wiley-Blackwell: N.J.

Brehm, L., & Bock, K. (2013). What counts in grammatical number agreement? *Cognition*, *128*(2), 149–169.

Brehm, L., & Bock, K. (2017). Referential and lexical forces in number agreement. *Language, Cognition and Neuroscience*, *32*(2), 129–146.

Brehm, L., Cho, P. W., Smolensky, P., & Goldrick, M. A. (2022). PIPS: A parallel planning model of sentence production. *Cognitive Science*, *46*(2), e13079.

Brehm, L., & Goldrick, M. (2017). Distinguishing discrete and gradient category structure in language: Insights from verb-particle constructions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(10), 1537.

Brehm, L., Jackson, C. N., & Miller, K. L. (2019). Speaker-specific processing of anomalous utterances. *Quarterly Journal of Experimental Psychology*, *72*(4), 764–778.

Brehm, L., Jackson, C. N., & Miller, K. L. (2021). Probabilistic online processing of sentence anomalies. *Language, Cognition and Neuroscience*, *36*(8), 959–983.

Browman, C. P. (1980). Perceptual processing: Evidence from slips of the ear. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, Ear. Pen, and Hand*. NY, London: Academic Press.

Buchwald, A., & Miozzo, M. (2012). Phonological and motor errors in individuals with acquired sound production impairment. *Journal of Speech, Language, and Hearing Research*, *55*, S1573–S1586.

Butterworth, B., & Whittaker, S. (1980). Peggy Babcock's relatives. In *Vol. 1. Advances in psychology* (pp. 647–656). North-Holland.

Buxó-Lugo, A., & Slevc, L. R. (2022). *Integration of input and expectations influences syntactic parses, not just sentence interpretation*. (Arxiv Preprint).
Cai, Z. G. (2022). Interlocutor modelling in comprehending speech from interleaved interlocutors of different dialectic backgrounds. *Psychonomic Bulletin & Review*, *29*(3), 1026–1034.
Cai, Z. G., Gilbert, R. A., Davis, M. H., Gaskell, M. G., Farrar, L., Adler, S., et al. (2017). Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology*, *98*, 73–101.
Cai, Z. G., Zhao, N., & Pickering, M. J. (2022). How do people interpret implausible sentences? *Cognition*, *225*, 105101.
Caramazza, A. (1991). Some aspects of language processing revealed through the analysis of acquired aphasia: The lexical system. *Issues in Reading, Writing and Speaking*, 15–44.
Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*(1), 177–208.
Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: Evidence from the tip-of-the-tongue phenomenon. *Cognition*, *64*(3), 309–343.
Celce-Murcia, M. (1980). On Meringer's corpus of 'slips of the ear'. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear. Pen, and hand*. NY, London: Academic Press.
Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407.
Christianson, K., Luke, S. G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 538.
Cohen, A. (1980). Correcting of speech errors in a shadowing task. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear. Pen, and hand*. NY, London: Academic Press.
Coppock, E. (2010). Parallel grammatical encoding in sentence production: Evidence from syntactic blends. *Language and Cognitive Processes*, *25*(1), 38–49.
Craft, J. T., Wright, K. E., Weissler, R. E., & Queen, R. M. (2020). Language and discrimination: Generating meaning, perceiving identities, and discriminating outcomes. *Annual Review of Linguistics*, *6*, 389–407.
Cutler, A. (1981). The reliability of speech error data. *Linguistics*, *19*, 561–582.
Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 105–121). MIT Press.
Cutting, J. C., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition*, *25*(1), 57–71.
Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283.
Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*(1–3), 287–314.
Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1355.
Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 611–629.
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.

Dempsey, J., Christianson, K., & Tanner, D. (2022). Misretrieval but not misrepresentation: A feature misbinding account of post-interpretive effects in number attraction. *Quarterly Journal of Experimental Psychology*, *75*(9), 1727–1745.
Deutsch, A., & Dank, M. (2009). Conflicting cues and competition between notional and grammatical factors in producing number and gender agreement: Evidence from Hebrew. *Journal of Memory and Language*, *60*(1), 112–143.
Deutsch, A., & Dank, M. (2011). Symmetric and asymmetric patterns of attraction errors in producing subject–predicate agreement in Hebrew: An issue of morphological structure. *Language and Cognitive Processes*, *26*(1), 24–46.
Dingemanse, M., & Enfield, N. (2015). Other-initiated repair across languages: Towards a typology of conversational structures. *Open Linguistics*, *1*(1).
Eberhard, K. M. (1999). The accessibility of conceptual number to the processes of subject–verb agreement in English. *Journal of Memory and Language*, *41*(4), 560–578.
Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, *112*(3), 531.
Egbert, M. M. (1997). Some interactional achievements of other-initiated repair in multi-person conversation. *Journal of Pragmatics*, *27*(5), 611–634.
Engel, A., & Hanulíková, A. (2020). Speaking style modulates morphosyntactic expectations in young and older adults: Evidence from a sentence repetition task. *Discourse Processes*, *57*(9), 749–769.
Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203.
Ferreira, F., & Patson, N. D. (2007). The 'good enough'approach to language comprehension. *Language and linguistics compass*, *1*(1–2), 71–83.
Foote, R., & Bock, K. (2012). The role of morphology in subject–verb number agreement: A comparison of Mexican and Dominican Spanish. *Language and Cognitive Processes*, *27*(3), 429–461.
Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language*, *43*(2), 182–216.
Franck, J., Lassi, G., Frauenfelder, U. H., & Rizzi, L. (2006). Agreement and movement: A syntactic analysis of attraction. *Cognition*, *101*(1), 173–216.
Franck, J., Vigliocco, G., & Nicol, J. (2002a). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes*, *17*(4), 371–404.
Franck, J., Vigliocco, G., & Nicol, J. (2002b). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes*, *17*(4), 371–404.
Frazier, L., & Clifton, C., Jr. (2011). Quantifiers undone: Reversing predictable speech errors in comprehension. *Language*, *87*(1), 158.
Frazier, L., & Clifton, C., Jr. (2015). Without his shirt off he saved the child from almost drowning: Interpreting an uncertain input. *Language, Cognition and Neuroscience*, *30*(6), 635–647.
Freud, S. (1958). Psychopathology of everyday life (A. A. Brill, Trans.). New York: New American Library. (Original work published 1901).
Frisch, S. A., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, *30*(2), 139–162.
Fromkin, V.A. Speech Error Database. Retrieved December 1, 2022 from https://www.mpi.nl/dbmpi/sedb/sperco_form4.pl.
Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 27–52.
Garrett, M. F. (1975). The analysis of sentence production. In *Vol. 9. Psychology of learning and motivation* (pp. 133–177). Academic Press.
Garnes, S., & Bond, Z. (1980). A slip of the ear: A snip of the ear? A slip of the year. In *Errors in Linguistic Performance* (pp. 231–239).

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056.
Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., et al. (2017). Don't underestimate the benefits of being misunderstood. *Psychological Science*, *28*(6), 703–712.
Gillespie, M., & Pearlmutter, N. J. (2011). Hierarchy and scope of planning in subject–verb agreement production. *Cognition*, *118*(3), 377–397.
Gillespie, M., & Pearlmutter, N. J. (2013). Against structural constraints in subject–verb agreement production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 515.
Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649–683.
Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, *102*(2), 219–260.
Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, *103*(3), 386–412.
Gordon, J. K. (2002). Phonological neighborhood effects in aphasic speech errors: Spontaneous and structured contexts. *Brain and Language*, *82*(2), 113–145.
Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics*, *42*, 93–108.
Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887.
Hartsuiker, R. J. (2006). Are speech error patterns affected by a monitoring bias? *Language and Cognitive Processes*, *21*(7–8), 856–891.
Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes*, *21*(1–3), 181–204.
Hartsuiker, R. J., Schriefers, H. J., Bock, K., & Kikstra, G. M. (2003). Morphophonological influences on the construction of subject-verb agreement. *Memory & Cognition*, *31*(8), 1316–1326.
Harvey, D. Y., Traut, H. J., & Middleton, E. L. (2019). Semantic interference in speech error production in a randomised continuous naming task: Evidence from aphasia. *Language, cognition and neuroscience*, *34*(1), 69–86.
Haskell, T. R., & MacDonald, M. C. (2003). Conflicting cues and competition in subject–verb agreement. *Journal of Memory and Language*, *48*(4), 760–778.
Haskell, T. R., Thornton, R., & MacDonald, M. C. (2010). Experience and grammatical agreement: Statistical learning shapes number agreement production. *Cognition*, *114*(2), 151–164.
Hay, J., Drager, K., & Warren, P. (2010). Short-term exposure to one dialect affects processing of another. *Language and Speech*, *53*(4), 447–471.
Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458–484.
Hieke, A. E. (1981). A content-processing view of hesitation phenomena. *Language and Speech*, *24*(2), 147–160.
Huang, Y., & Ferreira, F. (2021). What causes lingering misinterpretations of garden-path sentences: Incorrect syntactic representations or fallible memory processes? *Journal of Memory and Language*, *121*, 104288.
Humphreys, K. R., & Bock, K. (2005). Notional number agreement in English. *Psychonomic Bulletin & Review*, *12*(4), 689–695.

Humphreys, K. R., Menzies, H., & Lake, J. K. (2010). Repeated speech errors: Evidence for learning. *Cognition*, *117*(2), 151–165.
Jaeger, T. F., & Norcliffe, E. J. (2009). The cross-linguistic study of sentence production. *Language and Linguistics Compass*, *3*(4), 866–887.
Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, *69*(5), 1013–1040.
Kendrick, K. H. (2015). Other-initiated repair in English. *Open Linguistics*, *1*(1).
Konopka, A. E., & Bock, K. (2009). Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology*, *58*(1), 68–101.
Kortmann, B., & Szmrecsanyi, B. (2004). Global synopsis: Morphological and syntactic variation in English. *A handbook of varieties of English*, *2*, 1142–1202.
Labov, W. (1968). A study of the non-standard English of Negro and Puerto Rican speakers in New York City. Volume I: Phonological and Grammatical Analysis.
Labov, W. (1973). *Sociolinguistic patterns*. University of Pennsylvania press.
Labov, W. (1994). *Principles of linguistic change*. Volume I to III. (Language in Society 20) Oxford: Blackwell.
Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, *82*, 133–149.
Lemhöfer, K., Schriefers, H., & Indefrey, P. (2020). Syntactic processing in L2 depends on perceived reliability of the input: Evidence from P600 responses to correct input. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(10), 1948.
Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104.
Levelt, W. J. (1989). *Speaking: From intention to articulation*. MIT press.
Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38.
Levy, R. (2008, October). A noisy-channel model of human sentence comprehension under uncertain input. In *Proceedings of the 2008 conference on empirical methods in natural language processing* (pp. 234–243).
Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), 21086–21090.
Lorimor, H., Bock, K., Zalkind, E., Sheyman, A., & Beard, R. (2008). Agreement and attraction in Russian. *Language and Cognitive Processes*, *23*(6), 769–799.
MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, *8*(3), 323–350.
MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology*, *3*(2), 210–227.
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.
Martin, N., Weisberg, R. W., & Saffran, E. M. (1989). Variables influencing the occurrence of naming errors: Implications for models of lexical retrieval. *Journal of Memory and Language*, *28*(4), 462–485.
McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition*, *117*(3), 243–260.
Meringer, R. & Mayer, K. (1895) *Versprechen und Verlesen*, Goschenscher-Verlag (Reprinted 1978, with introductory essay by A Cutler and D.A. Fay, Benjamins).
Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition*, *42*(1–3), 181–211.
Mirković, J., & MacDonald, M. C. (2013). When singular and plural are both grammatical: Semantic and morphophonological effects in agreement. *Journal of Memory and Language*, *69*(3), 277–298.

Motley, M. T., & Baars, B. J. (1976a). Laboratory induction of verbal slips: A new method for psycholinguistic research. *Communication Quarterly*, *24*(2), 28–34.
Motley, M. T., & Baars, B. J. (1976b). Semantic bias effects on the outcomes of verbal slips. *Cognition*, *4*(2), 177–187.
Motley, M. T., Camden, C. T., & Baars, B. J. (1981). Toward verifying the assumptions of laboratory-induced slips of the tongue: The output-error and editing issues. *Human Communication Research*, *8*(1), 3–15.
Nickels, L. (2014). *Spoken word production and its breakdown in aphasia*. Psychology Press.
Nickels, L., & Howard, D. (1994). A frequent occurrence? Factors affecting the production of semantic errors in aphasic naming. *Cognitive Neuropsychology*, *11*(3), 289–320.
Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, *18*(1), 62–85.
Nooteboom, S. G. (1969). The tongue slips into patterns. In A. G. Sciarone, A. J. van Essen, & A. A. Van Raad (Eds.), *Leyden studies in linguistics and phonetics* (pp. 114–132). The Hague: Mouton.
Nooteboom, S., & Quené, H. (2008). Self-monitoring and feedback: A new attempt to find the main cause of lexical bias in phonological speech errors. *Journal of Memory and Language*, *58*(3), 837–861.
Nooteboom, S. G., & Quené, H. (2013). Parallels between self-monitoring for speech errors and identification of the misspoken segments. *Journal of Memory and Language*, *69*(3), 417–428.
Nooteboom, S. G., & Quené, H. (2017). Self-monitoring for speech errors: Two-stage detection and repair with and without auditory feedback. *Journal of Memory and Language*, *95*, 19–35.
Nozari, N., & Dell, G. S. (2012). Feature migration in time: Reflection of selective attention on speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1084.
Nozari, N., & Novick, J. (2017). Monitoring and control in language production. *Current Directions in Psychological Science*, *26*(5), 403–410.
Patson, N. D., & Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *Quarterly Journal of Experimental Psychology*, *69*(5), 950–971.
Pouplier, M., & Goldstein, L. (2005). Asymmetries in the perception of speech production errors. *Journal of Phonetics*, *33*(1), 47–75.
Pouplier, M., & Hardcastle, W. (2005). A re-evaluation of the nature of speech errors in normal and disordered speakers. *Phonetica*, *62*(2–4), 227–243.
Qian, Z., Garnsey, S., & Christianson, K. (2018). A comparison of online and offline measures of good-enough processing in garden-path sentences. *Language, Cognition and Neuroscience*, *33*(2), 227–254.
Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*(3), 460.
Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1–3), 107–142.
Ryskin, R., Futrell, R., Kiran, S., & Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition*, *181*, 141–150.
Ryskin, R., Ng, S., Mimnaugh, K., Brown-Schmidt, S., & Federmeier, K. D. (2020). Talker-specific predictions during language processing. *Language, Cognition and Neuroscience*, *35*(6), 797–812.
Schegloff, E. A. (1979). The relevance of repair to syntax-for-conversation. In *Discourse and syntax* (pp. 261–286). Brill.
Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–382.
Schlueter, Z., Parker, D., & Lau, E. (2019). Error-driven retrieval in agreement attraction rarely leads to misinterpretation. *Frontiers in Psychology*, *10*, 1002.

Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language*, *54*(2), 228–264.
Seuss, D. (1965). *Fox in socks*. London: Harper Collins Children's Books.
Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition*, *53*(2), 91–127.
Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial order mechanism in sentence production. In *Sentence processing: Psycholinguistic studies presented to Merril Garrett* Lawrence Erlbaum.
Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In *The production of speech* (pp. 109–136). New York, NY: Springer.
Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, *42*(1–3), 213–259.
Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, *18*(1), 41–55.
Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, *69*(2), 104–120.
Smith, G., Franck, J., & Tabor, W. (2018). A self-organizing approach to subject–verb number agreement. *Cognitive Science*, *42*, 1043–1074.
Smolensky, P., Goldrick, M., & Mathis, D. (2014). Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science*, *38*(6), 1102–1138.
Solomon, E. S., & Pearlmutter, N. J. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology*, *49*(1), 1–46.
Squires, L. (2013). It don't go both ways: Limited bidirectionality in sociolinguistic perception. *Journal of SocioLinguistics*, *17*(2), 200–237.
Squires, L. (2014). Processing, evaluation, knowledge: Testing the perception of English subject–verb agreement variation. *Journal of English Linguistics*, *42*(2), 144–172.
Starreveld, P. A., & La Heij, W. (1999). Word substitution errors in a speeded picture-word task. *The American Journal of Psychology*, *112*(4), 521.
Staum-Casasanto, L. (2008). Does social information influence sentence processing? *Proceedings of the Annual Meeting of the Cognitive Science Society.*, *30*(30), 799–804.
Stemberger, J. P. (1992). The reliability and replicability of naturalistic speech error data: A comparison with experimentally induced errors. In B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition* (pp. 195–215). Plenum Press.
Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591.
Vandierendonck, A., Loncke, M., Hartsuiker, R. J., & Desmet, T. (2018). The role of executive control in resolving grammatical number conflict in sentence comprehension. *Quarterly Journal of Experimental Psychology*, *71*(3), 759–778.
Veenstra, A., Antoniou, K., Katsos, N., & Kissine, M. (2018). Resisting attraction: Individual differences in executive control are associated with subject–verb agreement errors in production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(8), 1242.
Veenstra, A., Meyer, A. S., & Acheson, D. J. (2015). Effects of parallel planning on agreement production. *Acta Psychologica*, *162*, 29–39.
Viebahn, M. C., Ernestus, M., & McQueen, J. M. (2017). Speaking style influences the brain's electrophysiological response to grammatical errors in speech comprehension. *Journal of Cognitive Neuroscience*, *29*(7), 1132–1146.

Vigliocco, G. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes*, *11*(4), 407–442.
Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition*, *61*(3), 261–298.
Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language*, *34*(2), 186–215.
Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin*, *128*(3), 442.
Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition*, *68*(1), B13–B29.
Vitkovitch, M., Humphreys, G. W., & Lloyd-Jones, T. J. (1993). On naming a giraffe a zebra: Picture naming errors across different object categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 243.
Vousden, J. I., Brown, G. D., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology*, *41*(2), 101–175.
Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206–237.
Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 387.
Weissler, R. E., & Brennan, J. R. (2020). How do listeners form grammatical expectations to African American language? *University of Pennsylvania Working Papers Linguistics*, *25*(2), 16.
Wilshire, C. E. (1999). The "tongue twister" paradigm as a technique for studying phonological encoding. *Language and Speech*, *42*(1), 57–82.
Wolfram, W., & Christian, D. (1976). *Appalachian speech*. Arlington, VA: Center for Applied Linguistics.
Wright, S. (1954). The death of lady Mondegreen. *Harper's Magazine.*, *209*(1254), 48–51.
Wulff, S., & Gries, S. T. (2019). Improving on observational blends research: Regression modeling in the study of experimentally-elicited blends. *Lexis. Journal in English Lexicology*, *14*.
Zaharchuk, H. A., Shevlin, A., & van Hell, J. G. (2021). Are our brains more prescriptive than our mouths? Experience with dialectal variation in syntax differentially impacts ERPs and behavior. *Brain and Language*, *218*, 104949.

This page intentionally left blank

CHAPTER TWO

What do we know about the mechanisms of response planning in dialog?

Ruth E. Corps*
Psychology of Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
*Corresponding author: e-mail address: ruth.corps@mpi.nl

Contents

Abstract

During dialog, interlocutors take turns at speaking with little gap or overlap between their contributions. But language production in monolog is comparatively slow. Theories of dialog tend to agree that interlocutors manage these timing demands by planning a response early, before the current speaker reaches the end of their turn. In the first half of this chapter, I review experimental research supporting these theories. But this research also suggests that planning a response early, while simultaneously comprehending, is difficult. Does response planning need to be this difficult during dialog? In other words, is early-planning always necessary? In the second half of this chapter, I discuss research that suggests the answer to this question is no. In particular, corpora of natural conversation demonstrate that speakers do not directly respond to the immediately preceding utterance of their partner—instead, they continue an

Psychology of Learning and Motivation, Volume 78
ISSN 0079-7421
https://doi.org/10.1016/bs.plm.2023.02.002

utterance they produced earlier. This parallel talk likely occurs because speakers are highly incremental and plan only part of their utterance before speaking, leading to pauses, hesitations, and disfluencies. As a result, speakers do not need to engage in extensive advance planning. Thus, laboratory studies do not provide a full picture of language production in dialog, and further research using naturalistic tasks is needed.

1. Introduction

Psycholinguists have developed detailed accounts of the cognitive processes underlying speaking (language production) and listening (language comprehension), and they have traditionally studied these mechanisms separately. In fact, we have sophisticated theories of language production during monolog (i.e., when we speak by ourselves; Section 2). However, the majority of language use occurs in dialog, in which we rarely just speak or listen. Instead, we usually take turns at talking, regularly switching between comprehending our partner and producing our own response. But what do we actually know about the mechanisms of language production, and particularly response planning, in dialog?

Dialog has been of interest to corpus linguists for decades, and particularly since Sacks, Schegloff, and Jefferson's (1978) seminal work on the rules governing interaction. They noted that dialog involves at least two people who take alternating turns at speaking. The content of these turns (i.e., what the speaker wants to say) is not specified in advance, and so speakers have to plan their utterances "on the fly." Importantly, only one speaker tends to talk at a time, and transitions (from one speaker to the next) with no gap or overlap are common but any overlap that does occur is very brief. Thus, turns are coordinated in time.

More recently, corpus studies have confirmed the close timing of turns. For example, Stivers et al. (2009) quantified response times to polar (*yes/no*) questions in ten languages. They found that there was variation in the average gap duration across languages, with some having short average gaps (such as Japanese, with an average gap of 7 milliseconds [ms]) and others having longer average gaps (such as Danish, with an average gap of 469 ms). But despite this variation, most answers (i.e., the peak of the distribution, or the mode) were produced within 200 ms of the question end across all languages. Furthermore, Heldner and Edlund (2010) analyzed gap durations in three different corpora—a Dutch dialog corpus, which consisted of face-to-face and telephone conversations, and English and Swedish Map Task corpora, where speakers worked together to make their way around a map. Although there was a large amount of overlap in all three corpora (40%), the majority of all turn transitions (51–55%) took place within 200 ms, with 70–82% taking place within 500 ms.

The short gaps between turns in conversation contrasts with the much longer latencies in isolated language production. Research has shown that producing a picture name takes between 600 and 1200 ms, depending on factors such as word frequency (e.g., Indefrey & Levelt, 2004), while a complete utterance takes around 1500 ms (e.g., Ferreira, 1991). Thus, if the listener (as the next-speaker) is to achieve a turn gap of 200 ms, then they must begin planning their response before the current speaker reaches the end of their utterance. As a result, the listener must plan while still comprehending the speaker, and comprehension and production processes must overlap (at least momentarily). In fact, there is evidence that responding too slowly is interpreted negatively by the other person in the dialog, and so timely responses are socially desirable (Section 3.1).

This *early-planning* mechanism has been implemented in theories of conversation, which claim that listeners predict what a speaker is likely to say (utterance content) and use these predictions to begin planning a response as soon as possible, even if they are still comprehending the speaker's utterance (e.g., Levinson & Torreira, 2015; see Section 3.2). There is much evidence to support early-planning in dialog (see Sections 3.3 and 3.4). This evidence comes from highly-constrained laboratory tasks that use a variety of techniques, such as question-answering or picture naming, designed to approximate the processes involved in conversation. Much like dialog, these tasks involve planning a response and articulating it at the appropriate moment, so there is little gap or overlap between responses.

This research suggests that early-planning enables interlocutors to closely coordinate their utterances. But this research also suggests that planning a response while simultaneously comprehending is difficult (see Section 3.5). Does language production need to be this difficult during dialog? In other words, does there always need to be this large overlap between comprehension and production processes? Our recent research (Corps, Knudsen, & Meyer, 2022) suggests the answer to this question is no. In a recent corpus analysis, we have shown that speakers do not always respond to each other during dialog (see Section 4.1)—instead, they continue an utterance they produced earlier. In these cases, the listener's response does not depend on the content of the speaker's utterance, and so comprehension and production processes do not always need to extensively overlap.

In addition, studies of monolog suggest that language production is highly incremental, and speakers do not need to plan their full utterance before they actually speak. Theories and studies of dialog have tended to ignore this incrementality. For example, Levinson and Torreira (2015) claim that speakers complete all stages of response planning as early as possible. But incrementality likely makes language production easier than it would be if speakers planned a full sentence before speaking—the cognitive effort of planning is distributed throughout the utterance, rather than concentrated at the start. One consequence of this incrementality is that speakers are often disfluent, producing filled pauses such as *uh* or *um* (see Section 2.2.1). Because these theories of dialog have paid little attention to this incrementality and disfluency, our understanding of language production is incomplete—laboratory studies have focused on idealized situations, in which one speaker plans a full utterance in response to the previous speaker. In Section 5, I report a corpus analysis that investigates the similarity of speech elicited in the laboratory and speech elicited in natural conversations, with the aim of demonstrating that we need to step away from basing theories of language production in dialog on the idealized utterances produced in highly constrained laboratory tasks. In the following sections, I first provide an overview of what we know about language production in monolog before turning to dialog. Note that I limit my discussion to the mechanisms of production. As a result, I do not discuss the extensive literature on priming in dialog, which focuses on what causes speakers to produce one word or syntactic structure over another (see e.g., Garrod & Pickering, 2007, for a review).

2. The mechanisms of language production in monolog

2.1 Producing words

Although producing a word may seem simple, it is no easy feat. Researchers tend to agree that language production is a staged process, typically divided into three steps—deciding what to say (conceptualization), deciding how to say it (formulation), and then finally saying it (articulation; Levelt, 1989). During conceptualization, the speaker decides which message they wish to convey. For example, if the speaker is asked to name a picture of a dog, then they may activate the lexical concept *golden retriever* or *dog*, depending on the context of production. If the speaker names the picture in the context of other dogs, then they will produce *golden retriever*. But if they name in the context of other animals, then they will likely produce *dog* (see Clark, 1997), unless the term *golden retriever* has been used recently (e.g., Brennan & Clark, 1996).

The concept is then formulated, and this process of formulation (or lexicalization) involves two steps. First, activation spreads from the concept to connected abstract lexical representations. For the sake of simplicity, I adopt the terminology of Levelt, Roelofs, and Meyer (1999); see also Kempen & Huijbers (1983) and refer to these representations as *lemmas*, but they have also been referred to as *lexical entries*, *lexical representations*, or simply *words* throughout the literature. Theories tend to differ with respect to how lemmas are characterized. Some researchers claim that lemmas are lexical representations specifying the meaning of a word (or its semantics; e.g., Butterworth, 1989), while others claim that lemmas represent the syntactic features of a word, such as its grammatical class (e.g., whether it is a noun or a verb) or gender (e.g., whether it is gendered or gender-neutral; Levelt et al., 1999). But regardless, this lemma is the interface between the conceptual level and the next stage of formulation—word-form retrieval.

During word-form retrieval (or phonetic encoding), the activated lemma is mapped onto its corresponding word-form, which provides the speaker with information about the word's sound and how it should be produced. Constructing this word-form involves retrieving the word's morphological makeup, its metrical shape, and its segmental makeup (phonological encoding). For example, if the speaker is producing the word *dog*, then they will retrieve the morpheme <dog>. They will then spell out the metrical shape of *dog* (that it is monosyllabic) and its segmental information (/d/ /ɒ/ /g/). These representations spread activation to connected phonemes, which specify the word's syllabary and the articulatory gestures for producing the word (such as the necessary mouth movements). Once this process is complete, the speaker finally articulates the word.

It is worth noting that I have painted a rather simplistic view of word production. Although researchers agree that production involves selecting a word's meaning and its form, this is where the agreement tends to end. Some theories claim that production is strictly serial, so that speakers only activate the word-form of a single lemma (e.g., Levelt et al., 1999). For example, if the speaker wishes to produce the word *dog*, then activation will spread to semantically related lemmas, such as *cat* and *bone*, but only the word-form for the selected lemma (*dog*) is actually activated (e.g., Levelt et al., 1991). Others, however, claim that activation flows freely among meaning, lexical, and sound representations, and so speakers activate the word form of partially activated but unselected lemmas (e.g., Dell, 1986).

For example, the speaker would activate the word-form of *dog*, *cat*, and *bone*, even though they selected the lemma for *dog* (e.g., Peterson & Savoy, 1998).

Additionally, most theories accept the existence of an intermediary stage between conceptualization and word-form access, but others have rejected the existence of lemmas completely. For example, Caramazza (1997); Caramazza & Miozzo (1998); Miozzo & Caramazza (1997) suggested that some characteristics (e.g., verb tense or grammatical category), which are often thought to be activated at the lemma level, can be directly activated from a word's concept, while others (e.g., gender features) can be activated from word form. Thus, Caramazza claims that lemmas are not necessary for production.

But regardless of these disagreements, there is clear evidence for separate meaning and word-form representations in word production. In the classic picture-word interference (PWI) paradigm, participants name pictures while ignoring auditory or written distractor words (e.g., Schriefers, Meyer, & Levelt, 1990). Participants are slower to name a picture (e.g., dog) when the distractor word is semantically related (e.g., *cat*) rather than unrelated. They are also faster to name a picture when the distractor word is phonologically related (e.g., *doll*). Importantly, these effects depend on the time interval between the presentation of the distractor word and the presentation of the picture (the stimulus onset asynchrony, or SOA). In particular, a semantically related distractor word interferes with picture naming when presented 150 ms before the picture (an SOA of −150 ms), while a phonologically related distractor facilitates picture naming when presented at the same time as the picture or 150 ms after (an SOA of 0 or +150 ms). These results suggest that lexical access is staged, with meaning accessed separately from form.

Tip-of-the-tongue (TOT) states also support the separation of meaning and form representations. A TOT state occurs when the speaker cannot recall a particular word (even though they know it), but can recall information about the word. For example, speakers can report information about the word's form, such as its length in syllables or its word onset (e.g., Brown & McNeill, 1966). They can also recall syntactic information, such as the word's grammatical gender (e.g., Vigliocco, Antonini, & Garrett, 1997), its grammatical class (e.g., Iwasaki, Vigliocco, & Garrett, 1998), and whether it is a count or mass noun (e.g., Vigliocco, Vinson, Martin, & Garrett, 1999). These findings suggest that speakers are able to correctly report syntactic and semantic information about the word, even though they cannot retrieve the word's full form for articulation, suggesting that form information is accessed separately from meaning. Thus, we know that speakers produce a word by selecting its meaning separately from its form.

2.2 Incrementality in sentence production

Words are often produced as part of larger sentences, and so speakers have to activate multiple words and order them in an appropriate structure. How many words can speakers activate in parallel? In other words, how far ahead do they plan? Existing theories of sentence production generally assume that speakers do not plan an entire sentence before they begin to speak (e.g., Ferreira & Slevc, 2007). Instead, planning proceeds incrementally—as soon as one piece of the sentence (such as the first word) is processed at one stage of production, it is passed onto the next stage. As a result, the complete sentence does not need to be planned at the conceptual level before it is formulated—later parts of the sentence can be planned while earlier parts are simultaneously formulated. For example, a speaker who wishes to say *Dogs chase cats* could activate the concept for the word *dog*, which triggers retrieval of its corresponding lemma and word-form. This word then takes the first spot in the sentence's syntactic frame, and the speaker can articulate *dog* without necessarily knowing how the sentence will end. Two sources of evidence support incrementality in sentence production—research that has shown that speech is often disfluent (Section 2.2.1), and research manipulating the ease of sentence planning (Section 2.2.2).

2.2.1 Incrementality and disfluencies

Most of the research investigating disfluencies in language production comes from dialog, but these studies are relevant for understanding incrementality during monolog and so I discuss these results here. We know that speakers do not plan their full sentence before they speak because they often produce disfluencies. For example, consider excerpt (1) below from the Santa Barbara corpus of American English (Du Bois et al., 2000), where Lynne is talking about shoeing a horse.

(1) Lynne: But uh what was I gonna say. Oh and it's really tiring though. And it—you know like, you get so—I've only done like, well, at the end of the year, now see I took the second half of the course.

It is clear from (1) that speech can be disfluent in many different ways: Utterances can be incomplete, contain pauses (which may be silent or filled with words like *uh* or *um*), hesitations, repetitions, discourse markers (such as *like*), and utterance restarts (e.g., Fox Tree & Clark, 1997). Many of these disfluencies are present in Lynne's utterance—for example, she produces filled pauses such as *uh*, and her utterance includes incomplete units (e.g., *you get so*) that are abandoned and never resumed. These different types of disfluencies occur at different rates. For example, Eklund and Shriberg (1998) found that 32% of sentences and 5% of words were disfluent in a corpus of American-English telephone conversations, with filled pauses occurring more often (59% of the time) than any other type of disfluency. Similar results were found by Bortfeld, Leon, Bloom, Schober, and Brennan (2001) in a corpus of task-oriented conversations. Furthermore, Branigan, Lickley, and McKelvie (1999) found that 31% of disfluencies in the English Map Task corpus were repetitions, 42% were deletions, 10% were hesitations, and 13% were substitutions.

Much debate has focused on the meaning of these disfluencies in language production (e.g., Fox Tree, 2010; Fox Tree & Schrock, 1999; Fraser, 1999). This debate is primarily of interest to researchers investigating speaking in dialog, since they are largely based on corpora of conversational speech, but I briefly summarize their results here since they are useful for understanding incrementality during sentence planning in monolog. According to the *signal account*, speakers produce disfluencies to signal upcoming difficulty or delay to the listener (e.g., Clark, 1994; Fox Tree & Clark, 1997; Smith & Clark, 1993), perhaps so they can hold the floor or encourage the listener to allocate their attention to forthcoming information. For example, Smith and Clark (1993) found turn gaps of 2230 ms when utterances began without a filler, gaps of 2560 ms when utterances began with *uh*, and a gap of 8830 ms when utterances began with *um*. Additionally, research suggests that comprehenders expect speakers to refer to objects that have not been mentioned before (discourse-new objects) when the speaker produces a disfluency (e.g., *Now put thee uh…*), but objects they have referred to before (discourse-old objects) when they do not produce a disfluency (e.g., *Now put the…*; e.g., Arnold, Fagnano, & Tanenhaus, 2003; Arnold & Tanenhaus, 2011; Arnold, Tanenhaus, Altmann, & Fagnano, 2004). Furthermore, disfluencies can trigger attention to upcoming words (e.g., Bosker, Tjiong, Quené, Sanders, & De Jong, 2015; Collard, Corley, MacGregor, & Donaldson, 2008), making them easier to remember later (e.g., Corley, MacGregor, & Donaldson, 2007). Thus, speakers may intentionally produce disfluencies to signal an upcoming delay or important information to the listener.

Alternatively, these disfluencies could be a *symptom* of difficulty planning to speak (e.g., Levelt, 1989). Although speakers may produce disfluencies to signal discourse-new objects, they may also produce disfluencies before referring to these objects simply because they find them harder to name than discourse-old objects. Consistent with this argument, Schachter, Christenfeld, Ravina, and Bilous (1991) found that speakers hesitated more when they had choice in what they could say, which presumably made planning difficult. Similarly, Hartsuiker and Notebaert (2009) found that participants produced more disfluencies, pauses, and self-corrections when naming pictures with low name agreement (such as *sofa*, which could also be referred to as *couch* or *settee*) than pictures with high name agreement (such as *arm*). They also found that participants made more self-corrections and repetitions when they named gender neuter pictures (which use the infrequent determiner *het* in Dutch) than when they named common gender pictures (which use the more frequent determiner *de*), suggesting speakers produce disfluencies when they experience difficulty during lexical access. However, note that these results could be interpreted in line with the signal account. For example, speakers could be aware that they will have difficulty naming pictures with low name agreement rather than high name agreement and could produce a disfluency to signal this difficulty to the listener. As a result, it is difficult to determine whether speakers produce disfluencies as a signal or a symptom of difficulty during speaking.

Regardless of the meaning of these disfluencies, they demonstrate that speakers do not plan their full sentence before they produce it. If they did, then we would expect disfluencies to rarely occur, unless they were a deliberate signal, and we would expect any that do occur to primarily be located at the beginning of the speaker's utterance, where most of the planning difficulty occurs. But inconsistent with this prediction, Clark and Fox Tree (2002) found that disfluencies are distributed throughout the speaker's utterance. They defined three locations in utterances from the London-Lund corpus of British English, which consists of face-to-face conversations. Corpus analyses often focus on intonation units, which are stretches of speech produced under a continuous intonation contour (e.g., Chafe, 1992). These intonation units can consist of sentences, phrases, parts of phrases, or even single words. Clark and Fox Tree defined three locations in intonation units: (1) at the boundary; (2) after the first word; and (3) later in the utterance. An example of these locations can be found in (2), where commas are used to mark intonation unit boundaries.

(2) and then uh somebody said, . [1] but um—[2] don't you think there's evidence of this, in the twelfth—[3] and thirteenth centuries?

If speakers plan their full sentence before speaking, then disfluencies should primarily occur at location 1, where the speaker begins a new intonation unit, and should not occur at location 2, where the speaker is part way through an intonation unit, and especially not at 3, where they have almost finished the intonation unit. Although speakers produced more *uhs* and *ums* at location 1 (43 per 1000 opportunities), they still occurred at location 2 (27 per 1000 opportunities) and location 3 (13 per 1000 opportunities), suggesting disfluencies are not confined to the start of the speaker's sentence.

Although I have focused on research investigating whether disfluencies are produced as a consequence of difficulties during language production, there is also evidence that disfluencies can serve pragmatic functions, such as signaling new information (e.g., Arnold et al., 2003; Arnold & Tanenhaus, 2011) or discourse structure (e.g., Swerts, 1998). Importantly, however, the occurrence of disfluencies demonstrates that speakers do not plan their full sentences before speaking. Instead, they plan incrementally, and so do not necessarily know how their sentence will end before they start speaking.

2.2.2 Experimental studies on advance planning and incrementality

Although studies investigating the occurrence of disfluencies during production provide evidence that speakers plan their sentences incrementally, these studies were not designed to explicitly test this claim. Experimental studies of sentence planning have investigated the scope of advance planning (i.e., how much of their sentence the speaker plans before speech onset), and provide more direct support for incrementality during sentence production (e.g., Brown-Schmidt & Konopka, 2015; Brown-Schmidt & Tanenhaus, 2006; Griffin & Bock, 2000; Griffin, 2001; Smith & Wheeldon, 1999; see Wheeldon, 2013, for a review). For example, Griffin (2001) conducted an eye-tracking experiment in which participants described objects displayed on-screen using the sentence frame *The A and the B are above the C*. Objects B and C varied in their name agreement—sometimes they had high agreement and only one plausible name (e.g., *apple*), while other times they had low agreement and multiple plausible names (e.g., *sofa* or *couch*). These objects also varied in the frequency of their dominant name—sometimes the dominant name was highly frequent, while other times it was less frequent. Participants spent longer looking at objects B and C when they had low rather than high agreement names. Participants also spent longer looking at these objects if their names were low rather than high frequency. But the agreement and frequency of these objects did not affect how quickly participants named object A, suggesting participants began speaking when they had planned object A's name, before they selected object B and C's names. In other words, speakers did not plan their full sentence before they began articulation.

Other eye-tracking studies also suggest that speakers tend to look at each object as they mention it, only shifting their gaze to the next object prior to articulation of the previous object (e.g., Gleitman, January, Nappa, & Trueswell, 2007; Griffin & Bock, 2000; Griffin & Spieler, 2006; Meyer, Sleiderink, & Levelt, 1998). If participants plan more than one word at a time, then the delay between fixating an object and naming it should be shorter for words occurring later in the sentence. But Griffin and Bock found that this delay was the same for all objects, regardless of their position in the sentence. Furthermore, Meyer et al. had participants name pairs of objects using noun phrase conjunctions, such as *scooter and hair*. Participants shifted their gaze from the current object to the next object only once they had retrieved the word-form of the object they were naming. In other words, they only fixated *hair* once they had retrieved the word-form of *scooter*. Together, these findings suggest that speakers plan only one word before beginning articulation.

However, other research using less predictable sentences suggests speakers can activate more than one word at a time. For example, Smith and Wheeldon (1999) had participants produce sentences about moving objects. Participants were slower to produce sentences beginning with complex noun phrases (e.g., *The dog and the kite move above the house*) rather than a simple noun phrase (e.g., *The dog moves above the kite and the house*), suggesting participants dedicated more resources to planning a later word (*kite*) before the onset of the first word (*dog*) when sentences were more complex. Furthermore, Meyer (1996); see also Wagner, Jescheniak, & Schriefers (2010) had participants name pairs of pictures with either noun phrase conjunctions (e.g., *the arrow and the bag*) or locative sentences (e.g., *the arrow is next to the bag*). While planning their utterance, speakers heard a distractor word, which was semantically related, phonologically related, or unrelated to the first or the second noun. Participants were slower to initiate their sentences when the distractor word was semantically related rather than unrelated to either of the nouns, suggesting that the speaker planned the meaning of both nouns. Participants were also faster to initiate their sentences when the distractor word was phonologically related rather than unrelated to the first but not the second noun, suggesting that participants planned the word-form of only the first noun. These findings suggest that the scope of planning is different for different stages of production.

Thus, although theories of sentence production tend to agree that speakers plan their utterances incrementally, there is disagreement about the scope of this incrementality. Some studies suggest that speakers plan word-by-word (e.g., Griffin, 2001), while others suggest they plan in larger chunks (e.g., Smith & Wheeldon, 1999). One way of reconciling these findings is by assuming that the scope of planning is different for meaning and word-form (e.g., Meyer, 1996). Relatedly, planning is likely flexible, and so the degree of incrementality is under the speaker's control (e.g., Konopka, 2012; Swets, Jacovina, & Gerrig, 2013).

Consistent with this suggestion, there is evidence that the scope of planning is influenced by time pressure. Ferreira and Swets (2002); see also Swets et al. (2013) had participants produce answers to two digit sums (e.g., 9 + 7 = ?) when time pressure was absent (Experiment 1) or present (Experiment 2). In both experiments, initiation times increased as problem difficulty increased. However, problem difficulty influenced utterance duration only in Experiment 2, suggesting that speakers simultaneously planned and articulated when they were encouraged to produce their utterance immediately. When there was no time pressure, participants made use of more extensive advance planning. Similarly, Wagner et al. (2010; Experiment 1) measured planning scope using a PWI task, in which participants produced simple sentences consisting of two nouns (e.g., *the frog is next to the mug*). While producing these sentences, participants heard distractors that were unrelated or semantically related to the first or the second noun. The authors determined whether each participant was a fast or a slow speaker based on their average latencies when they produced the sentence in the presence of an unrelated distractor. Both fast and slow speakers experienced an interference effect for the first noun—they were slower to initiate their sentences when the distractor word was semantically related rather than unrelated. But the interference effect on the second noun was larger for the slow than the fast speakers, suggesting slow speakers had a tendency to plan further in advance than fast speakers.

Planning scope is also sensitive to linguistic factors, such as ease of structural assembly. In their second PWI experiment, Wagner et al. (2010) asked participants to produce simple sentences (e.g., *the frog is next to the mug*) or to switch between producing simple and complex sentences (e.g., *the red frog is next to the red mug*). They found that the additional cognitive load of switching sentence structure eliminated any interference effect for the second noun, regardless of whether speakers were fast or slow. Similarly, Konopka (2012); see also Konopka & Meyer (2014) had participants describe three pictures using a complex noun phrase (e.g., *The axe and the saw are above/below the cup*). On some trials, targets were preceded by primes that elicited the same or a different sentence structure. The results showed that repeating sentence structure extended speakers' planning scope from one to two nouns. Together, these findings suggest speakers reduce their planning scope when structural assembly is difficult. But regardless of how much speakers plan in advance, these studies demonstrate that speakers plan incrementally—they do not need to plan a full sentence before they speak.

3. The mechanisms of language production in dialog

It is clear from Section 2 that we have sophisticated theories of speech production during monolog. But the majority of language use occurs in dialog, where we rarely just speak. Instead, we usually take turns at talking, regularly switching between comprehending our partner and producing our own response. What do we know about the mechanisms of language production, and particularly response planning, in dialog?

3.1 Why is timely language production so important in dialog?

Before discussing the mechanisms of language production, and timely turn-taking, in dialog, it is worth understanding why it is important that the listener responds to the speaker so quickly. Dialog seems difficult—most theories agree that the next-speaker has to juggle comprehension and production processes if they are to achieve turn gaps of 200 ms (e.g., Levinson & Torreira, 2015). The next-speaker could avoid this issue by beginning response planning only once the speaker has reached the end of their turn. So why are short gaps so important? Research suggests they are important for maintaining the flow of conversation, and there is evidence that delayed responses tend to be interpreted negatively by the listener. For example, if you invite someone for dinner then a delayed response may indicate the other person's reluctance. This issue is illustrated in an excerpt from a telephone conversation (3), in which C interprets a pause of 1.86 s as a negative response to his question (Levinson, 1995):

(3)
C: So um I was wondering if you would be in your office on Monday by any chance?
(1.86 s)
C: Probably not.

Experimental studies have investigated the consequences of these delayed responses. For example, Bögels, Kendrick and Levinson (2015), (see also Bögels, Kendrick, & Levinson, 2020) measured Dutch participants' brain activity while they listened to telephone conversations, in which one speaker produced an initiating action (e.g., a request, an offer, or a proposal) and the other speaker produced either an acceptance (*ja* or *yes*) or a rejection (*nee* or *no*). The gap between these two turns was either long (1000 ms) or short (300 ms). Participants displayed a larger N400, which is associated with semantic processing (see Kutas & Federmeier, 2011, for a review), when they encountered a rejection following a short rather than a long gap. This effect suggests that the listener expects an immediate response to be positive, and so they experience processing difficulty when this response is actually negative. Thus, long gaps can indicate that the speaker will produce a rejection, which the listener may interpret negatively.

Research also suggests that gap length affects how listeners view their partner. In one study, Templeton, Chang, Reynolds, Cone LeBeaumont, and Wheatley (2022) investigated whether response times (which are equivalent to turn gaps) provide a useful measure of social connection. Participants held a ten-minute casual conversation with a stranger (Experiment 1) or a friend (Experiment 2) and then rated their social connection. In both experiments, participants felt more connected to their partner and enjoyed the conversation more when their partner responded more quickly. In Experiment 3, participants listened to audio clips in which the gap between the turns was manipulated so it was either short or long. As in the previous experiments, participants thought the interlocutors were more socially connected when they responded more quickly to each other, suggesting overhearers perceive short gaps positively, even if they are not involved in the conversation.

In another study, Koudenburg, Postmes, and Gordijn (2013) had participants interact with each other naturally or with a one second delay between turns in the second half of the conversation. Participants who had a conversation with a delay of one second felt less solidarity with their partner than those who conversed naturally. Furthermore, Roberts and Francis (2013); Roberts, Francis, and Morgan (2006); Roberts, Margutti and Takano (2011) found that listeners' ratings of the speaker's willingness to comply decreased as the length of the gap between turns increased. Together, these findings suggest that long gaps do not only disrupt the flow of conversation—they are also socially undesirable. In the next sections, I discuss the mechanisms that enable interlocutors to avoid long gaps.

3.2 Levinson and Torreira's (2015) theory of language production in dialog

Although other psycholinguistic models of turn-taking exist (e.g., Garrod & Pickering, 2015), I focus my discussion on Levinson and Torreira's (2015) theory because it the most influential model in the literature (see e.g., Bögels & Levinson, 2017; Corps, Gambi, & Pickering, 2018, for a review). They proposed that the production system (supporting speaking) and the comprehension system (supporting listening) are simultaneously engaged in conversation. In particular, the listener (B) focuses on determining the gist of the current speaker's (A) utterance. B can determine the gist by identifying A's speech act (i.e., what type of utterance they are producing, such as a question; e.g., Gisladottir, Chwilla, & Levinson, 2015), or by using the context of A's utterance to predict what she is likely to say (e.g., Altmann & Kamide, 1999).

As soon as B has identified A's speech act or has predicted enough of A's utterance, B begins planning a response. Thus, the content of B's response and the moment he begins planning it both depend heavily on the content of A's utterance. While planning this response, B simultaneously listens to the rest of A's utterance and waits for cues that signal she will soon finish speaking. If B finishes planning before A has reached the end of her utterance, he holds his response in an articulatory buffer (presumably at the phonological level) until he can articulate. Once there is sufficient evidence that the end of the utterance is imminent, B launches his planned response.

This model explains short turn gaps by claiming that next-speakers are highly proactive and begin planning their utterances as soon as the response-relevant information has been provided. Under this theory, turns are coordinated in both content and time because (1) the content of B's utterance depends on the content of A's utterance; and (2) B only initiates articulation once A has finished. In the next section, I review evidence that listeners (as next-speakers) can determine the gist of the speaker's utterance by predicting the content of this utterance. I then discuss evidence that suggests speakers use these predictions to plan a response early, in line with Levinson and Torreira's theory.

3.3 Content prediction during language comprehension

This section provides only a brief review of evidence for prediction during comprehension, since more extensive reviews are readily available elsewhere (e.g., Pickering & Gambi, 2018). The important point is that much research has demonstrated that listeners predict what a speaker is likely to say—that is, the content of the speaker's utterance. For example, participants often expect the same continuation (e.g., *spoon*) when presented with sentence contexts such as *At the dinner party, I wondered why my mother wasn't eating her soup. Then I noticed she didn't have a*.... Importantly, this effect does not only occur in laboratory tasks; in natural conversations, interlocutors sometimes complete each other's utterances (e.g., Howes, Purver, Healey, Mills, & Gregoromichelaki, 2011), suggesting that the listener comprehends the speaker's incoming utterance and predicts what the speaker is likely to say next.

Some research exploring prediction during language comprehension has used the *visual-world paradigm*, in which participants view a visual scene (usually consisting of many objects) while simultaneously listening to sentences. Predictive looking is thought to occur when listeners attend to an object before it is actually mentioned. In one of the first studies using this method, Altmann and Kamide (1999), see also Kamide, Altmann, and Haywood (2003) recorded participants' eye movements while they viewed visual scenes (e.g., a picture of a boy, a cake, a toy car, a toy train set, and a ball) and simultaneously listened to sentences. In one condition, these sentences (e.g., *The boy will eat*…) could apply to only one object in the scene (e.g., the cake), thus making the mention of the cake predictable. In the other condition, the sentences could apply to any of the objects (e.g., *The boy will move*…), making it impossible for the listener to predict how the sentence would continue. When participants heard the verb *eat*, they fixated the cake earlier and for longer than when they heard the verb *move*, suggesting they used the semantics of the verb to predict which of the objects was most likely to be mentioned next.

There is also evidence that listeners predict syntax. In an electroencephalogram (EEG) experiment, Ito, Gambi, Pickering, Fullenbach, and Husband (2020) presented Italian participants with sentences (e.g., *The traffic on the motorway came to a standstill because… [Il traffic in autostrada è rimasto bloccato a causa di…*]) that predicted a particular article and noun combination of a particular syntactic gender (e.g., *an incident [un*$_{\text{masculine}}$ *incidente*$_{\text{masculine}}$]). These sentences continued with the expected article and noun combination, or they continued with an article and noun combination that mismatched the syntactic gender of the expected continuation (e.g., *a flooding [un'*$_{\text{feminine}}$ *inodazione*$_{\text{feminine}}$]). Participants showed a greater negativity around 250 ms after the article when they encountered the unexpected article + noun combination compared to when they encountered the expected article + noun. These findings suggest that listeners can predict the syntactic gender of upcoming words (see also Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Bates, Moreno, & Kutas, 2003).

Finally, there is some evidence that listeners predict word-form (but see DeLong, Urbach, & Kutas, 2017; Ito, Martin, & Nieuwland, 2017; Nieuwland et al., 2018; Urbach, DeLong, Chan, & Kutas, 2020, for interesting discussions of this evidence). In addition to manipulating syntactic gender, Ito et al. (2020) included a condition where the sentences continued with an article matching the gender of the expected article and noun, but mismatching the word-form (e.g., *a collision* [*uno*$_{\text{masculine}}$ *scontro*$_{\text{masculine}}$]). Participants showed a greater negativity around 450 ms after the article when they encountered the form mismatch article compared to when they encountered the expected article, suggesting they predicted the form of upcoming words.

It is worth noting that word-form predictions occurred later (around 450 ms) than syntactic predictions (around 250 ms) in Ito et al.'s (2020) study, consistent with theories that word-form predictions are delayed relative to semantic and syntactic predictions (e.g., Pickering & Gambi, 2018). However, other studies have found that word-form predictions show a similar time-course of activation to semantic predictions. DeLong, Chan, and Kutas (2018) recorded ERPs while participants read highly constraining sentence contexts (e.g., *The woman stashed her wallet in her* ***purse*** *for safety*) which were continued with a highly predictable word (*purse* in this example), an unpredictable word semantically related to the predictable word (*snatcher* rather than *purse*), or an unpredictable word orthographically related to the predictable word (*nurse* rather than *purse*). They found that both semantically related and orthographically related unpredictable words elicited similarly reduced N400s, suggesting word-form predictions show a similar time-course to semantic predictions (but see also Ito, Corley, Pickering, Martin, & Nieuwland, 2016).

In sum, there is evidence that listeners predict what a speaker is likely to say. Once the listener makes this prediction, they can begin the process of planning their response. For example, if the speaker says *Is the boy going to fly his*…, then the listener could predict the meaning of the word *kite* and use this prediction to plan their answer. The next section reviews evidence that supports such early-planning.

3.4 Evidence for early response planning

After having heard or predicted a sufficient part of the speaker's utterance, listeners can begin planning their own response. Studies investigating the time-course of response planning have used a variety of methods, including picture naming and question-answering, which are designed to be highly controlled while still approximating the mechanisms involved in conversation. In particular, participants' responses are generally highly constrained, but they still have to prepare this response and articulate it so they avoid extensive gap or overlap with the previous speaker. Many of these studies support Levinson and Torreira (2015) claim that listeners are highly pro-active and begin planning their response early while still comprehending.

In one of the first studies, Bögels, Magyari, and Levinson (2015) measured EEG correlates during a question-answering task, in which the information (here *007*) needed for response planning was available either early (e.g., *Which character, also called 007, appears in the famous movies*?) or late (e.g., *Which character from the famous movies is also called 007*?). Participants were quicker to answer when the critical information was available early (mean (M) = 640 ms) rather than late (M = 950 ms). EEG correlates showed a larger positivity to the critical word when participants planned a response rather than when they simply listened to the questions. This effect was localized to the middle frontal and precentral gyri, which overlap with brain areas involved in speech production (Indefrey & Levelt, 2004). This effect occurred around 500 ms after the onset of the critical information necessary for planning, suggesting that listeners planned their own response as soon as they could determine the likely answer to the question.

However, follow-up studies suggest that Bögels, Magyari and Levinson (2015) EEG findings could also indicate that participants were monitoring the speaker's utterance to determine when they could initiate articulation. Jongman, Piai, and Meyer (2020) found that the large positivity reported by was also linked to attention to the sequence end in a task where participants had to prepare and maintain an answer until they were given a cue to speak. Furthermore, Bögels, Magyari and Levinson (2015) used general knowledge questions, and so the answers likely had to be retrieved from episodic memory. Although previous research has found that the middle frontal and precentral gyri are associated with language production (Indefrey & Levelt, 2004), other studies report that the middle frontal gyrus may also be involved in episodic memory retrieval (e.g., Cabeza, 2002; Rajah, Languay, & Grady, 2011; Raz et al., 2005). Bögels, Magyari and Levinson (2015) did not observe the same pattern of activation in a control study, in which participants memorized the questions, but their results may still reflect the processes of retrieving the answer from memory for production.

Nevertheless, other studies provide converging evidence for early-planning. Magyari, De Ruiter, and Levinson (2017) used a similar paradigm to Bögels, Magyari and Levinson (2015) and had participants view pictures while answering questions such as *Which animal has a light switch and also a battery?*. In the late condition, both of the animals on-screen had a light switch and another object, and so participants could not plan a response until they heard the final object name. In the early condition, only one of the animals had objects, and so participants could plan a response even before they heard the question. Much like Bögels, Magyari and Levinson (2015), participants answered more quickly in the early ($M=320$ ms) than the late condition ($M=361$ ms), suggesting participants planned a response earlier when they knew the likely answer to the question compared to when they did not. But note that the difference between the two conditions was much smaller than in Bögels, Magyari and Levinson (2015) study, suggesting that the gain in response planning was not particularly large.

Results leading to a similar conclusion were reported by Meyer, Alday, Decuyper, and Knudsen (2018), who had participants answer questions (e.g., *Do you have a green sweater?*) while viewing four objects on-screen (e.g., a cake, a branch, a sweater, and a barrel). In the early condition, all the objects were the same color, and so participants could start planning an answer as soon as they understood the color adjective—for example, they knew as soon as they heard *green* that the answer would be *yes* if the objects were green and *no* if they were a different color. In the late condition, the objects were different colors and so participants could not plan an answer until the speaker produced the object name. Participants answered more quickly in the early ($M=215$ ms) than the late ($M=297$ ms) condition.

Similarly, we tested whether content prediction facilitates response planning in a set of *yes/no* question-answering studies (Corps, Crossley, Gambi, & Pickering, 2018). In one condition, the final words of the question were predictable (e.g., *Are dogs your favorite animal*?) because the majority of participants agreed on this final word as a continuation in a cloze pre-test. In the other condition, the final words were unpredictable (e.g., *Would you like to go to the supermarket*?) and participants provided different continuations in the cloze pre-test—even though some participants completed the question with the word *supermarket*, others responded with different words like *cinema* or *dentist*. We found that participants answered more quickly when the final words of the question were predictable ($M=379$ ms; Experiment 2b) rather than unpredictable ($M=536$ ms), suggesting they predicted the speaker's final word and used this prediction to plan a response. In other words, content prediction facilitated response planning.

Support for early-planning also comes from picture naming studies. Barthel, Sauppe, Levinson, and Meyer (2016); see also Barthel, Meyer, and Levinson (2017) used a task in which German participants completed a confederate's pre-recorded utterances. Participants had to name any on-screen objects that the confederate had not already named, and so they could (in principle) plan their response as soon as the confederate began uttering their last object name (indicated by the use of the word *and*; e.g., *I have a door and a bicycle*). Both eye movements and response latencies suggested that participants planned their response as soon as possible—they were faster to speak when there was a clear lexical cue (i.e., *and*) to the end of the list ($M = 761$ ms) than when there was not ($M = 867$ ms).

In sum, there is good evidence that listeners (as next-speakers) engage in early-planning during laboratory tasks designed to approximate the mechanisms involved in dialog. As a result, the listener plans their response while simultaneously comprehending the current speaker's utterance. In the next section, I discuss the cognitive demands of dual-tasking comprehension and production.

3.5 Early response planning is cognitively demanding

Although there is much experimental evidence to suggest that listeners plan a response early, as claimed by Levinson and Torreira (2015), participants' average response times were always longer than the 200 or 300 ms typically reported in corpus studies (e.g., Stivers et al., 2009). This difference is not particularly interesting—in some studies, participants had to answer general knowledge questions or name pictures, which likely involved memory search processes or object recognition before a response could actually be planned. What is interesting, however, is that the average gain in response times in the early relative to the late condition was much less than the time difference between the occurrence of these two cues. For example, participants in Bögels, Magyari and Levinson (2015) study responded around 300 ms earlier when the critical information necessary for answer planning occurred early rather than late. But the cue that enabled response planning (e.g., *007*) occurred on average 1700 ms earlier in the early than the late condition. Thus, the gain in response time did not match the gain in information, and 1400 ms were "lost."

This inefficiency likely occurs because listeners who plan early must represent both the speaker's utterance (using comprehension mechanisms) and their planned response (using production mechanisms). Both production and comprehension require central attention (see Jongman, 2021, for a review), and so dual-tasking them should be cognitively demanding. As a result, planning early may interfere with simultaneous comprehension (and vice versa). In fact, research suggests that all stages of response planning are cognitively demanding (e.g., Cook & Meyer, 2008; Ferreira & Swets, 2002; Roelofs, 2008; Roelofs & Piai, 2011).

In addition, comprehension and production are two very similar tasks, relying on similar neural circuits (e.g., Menenti, Gierhan, Segaert, & Hagoort, 2011; Silbert, Honey, Simony, Poeppel, & Hasson, 2014). For example, Segaert, Menenti, Weber, Petersson, and Hagoort (2012) found that the same brain areas (the left inferior frontal gyrus, the left middle temporal gyrus, and the bilateral supplementary motor area) were sensitive to syntactic repetition during comprehension and production. Furthermore, the representations for lexical concepts and lemmas are shared between production and comprehension. In the classic picture-word interference (PWI) paradigm, participants name pictures while ignoring simultaneously presented auditory or written distractor words (e.g., Schriefers et al., 1990). These studies have shown that participants are slower to name a picture (e.g., a dog) when the distractor word is semantically related (e.g., *cat*) rather than unrelated, suggesting that there is competition between shared representations of concepts during production (the target) and comprehension (the distractor).

This representational similarity is important because speakers' adjacent utterances are thought to be highly related in conversation, and research suggests that performance on one task suffers more when the other task is more rather than less similar (e.g., Wickens, 2008). For example, Fairs, Bögels, and Meyer (2018) used a psychological refractory period (PRP) paradigm, in which participants completed two separate tasks (Task A and Task B). The authors manipulated the interval between the start of Task B and the start of Task A by varying the stimulus onset asynchrony (SOA), so that participants sometimes completed the tasks in overlap. Participants experienced more interference when performing a picture-naming task alongside a syllable-identification task than when they performed a picture-naming task alongside tone-identification. These results suggest that the phonological representations used during syllable identification were also used during picture-naming, and competition occurred between comprehension and production when participants needed to use them simultaneously.

Thus, planning a response early may interfere with simultaneous comprehension. Research has recently begun to investigate this issue. In one study, Jongman and Meyer (2017) used a picture-naming task, in which half of the participants named the pictures (e.g., apple) while the other half listened to a pre-recorded speaker name the picture (i.e., planning condition was manipulated between-participants). In addition, pictures were preceded by auditory primes, which were either identical to (*apple*), associatively related to (*peel*), or unrelated to the target picture (*nail*). The authors found fastest naming latencies for pictures preceded by an identity prime, intermediate latencies for those preceded by an associatively related prime, and slowest latencies for those preceded by an unrelated prime. This priming pattern was the same regardless of whether or not participants named the non-target picture, suggesting that speech planning did not interfere with comprehension of the prime.

Jongman and Meyer replicated the identity priming effect in a second experiment, in which participants had to decide whether or not to name the picture at the start of each trial (i.e., planning condition was manipulated within participants). However, in this experiment they found an associative priming effect only when participants did not have to name the picture, suggesting that response planning interfered with comprehension. The lack of associative priming in the planning condition was likely related to the difficulty of the task. In Experiment 1, participants' task was predictable and they knew whether they would need to plan a response before picture onset. In Experiment 2, however, participants had to switch between planning and listening, which was likely cognitively demanding. This task-switching is particularly relevant for natural conversation, since the cognitive load is likely to be greater than in Jongman and Meyer's task given that participants often have to plan (and comprehend) longer, more complex utterances.

In another study, Bögels, Casillas, and Levinson (2018) used a similar paradigm as their earlier study (Bögels, Magyari, & Levinson, 2015), but participants viewed pictures on-screen (e.g., a banana and a pineapple) while simultaneously answering questions. Much like the previous study, the critical information (here *curved*) necessary for response planning was available either early (e.g., *Which object is curved and is considered to be a type of fruit*?) or late (e.g., *Which object is considered to be a type of fruit and is curved*?). But in addition, the questions contained either an expected or unexpected word (e.g., *healthy* rather than *fruit* in both examples). The authors found that participants responded later to questions with an unexpected rather than expected word, regardless of when the critical information occurred, suggesting that they still comprehended these words even when they planned their response early. In addition, an N400 effect occurred at the unexpected word in both planning conditions. However, the size of this effect varied as a result of participants' response times: Participants with slower response times showed a larger N400 effect than those with faster response times. Based on these results, the authors concluded that fast responders allocated fewer resources to comprehension (leading to a smaller N400) and more to production (leading to faster response times) when they encountered the information necessary for response planning. In contrast, slow responders allocated more resources to comprehension (leading to a larger N400) and fewer to production (leading to slower response times). Thus, this study provides some preliminary evidence that response planning interferes with comprehension.

Thus far, I have focused on studies that show planning interferes with comprehension. These studies could also demonstrate that comprehension interferes with planning. For example, the slow responders in Bögels et al.'s (2018) study may have been slower than the fast responders because comprehending hindered their response planning. But more direct evidence comes from PWI studies, which have shown that participants are slower to name pictures in the presence of words (even when the words are unrelated) than pseudowords (e.g., Dhooge & Hartsuiker, 2012), noise (e.g., Schriefers et al., 1990), or strings of X's (e.g., Glaser & Glaser, 1989), suggesting that comprehending a distractor word (even when you are told to ignore it) interferes with planning the picture name.

In dialog, however, speakers rarely hear words in isolation—they tend to be produced in sentence context. Recently, He, Meyer, and Brehm (2021) investigated whether unrelated background speech interferes with response planning. Dutch participants named a set of six pictures while they simultaneously ignored speech produced by a Dutch talker (high similarity speech), speech produced by a Chinese talker (moderate similarity speech), or eight-talker babble (low similarity speech). Participants were slower to name the pictures when they had to ignore the Dutch talker compared to the Chinese talker, and pictures in both of these conditions were named slower than in the eight-talker babble condition. These findings indicate that comprehension interferes with planning, but the degree of this interference is affected by the similarity of production and comprehension representations—when these representations are more similar (i.e., the same language), interference is higher than when they are less similar (i.e., a different language).

In sum, it is clear that early-planning is cognitively demanding. Not only is there evidence that planning interferes with comprehension, but comprehension also interferes with planning.

4. Is early-planning really necessary in dialog?

There is clear evidence that speakers plan a response early (Section 3.4), but there is also evidence that planning in this way is difficult (Section 3.5). Does language production need to be this difficult during dialog? In other words, do listeners always need to plan their response while still comprehending the current speaker's utterance? In the next sections, I discuss research that suggests the answer to this question is no, and language production in dialog may be easier than claimed by theories based on laboratory studies (e.g., Levinson & Torreira, 2015). Note that I am not claiming that early-planning never occurs during dialogue. In fact, early-planning is likely to be particularly useful during highly constrained interactions where speakers have clear expectations about what they are likely to say. Rather, I suggest that the need for early-planning may have been overestimated by theories of dialogue.

4.1 Speakers often do not directly respond to each other

Theories of dialog (and the experimental studies testing them) typically think of dialog as like a game of ping pong: The speaker produces an utterance, and the listener uses the content of that utterance to plan an appropriate response. As a result, the content of the speaker's utterance constrains the content of the listener's utterance, and the length of the speaker's utterance constrains the amount of time the listener has for planning. To formulate a response to the previous speaker's utterance, the next speaker must begin planning a response early if they are to achieve turn gaps of 200 ms. Thus, comprehension and production overlap.

In a recent corpus analysis (Corps et al., 2022), however, we observed that many natural dialogs often involve parallel talk, where each speaker develops their turn in parallel with the other speaker over several utterances (or what we refer to as segments, which are stretches of speech produced by one speaker). For example, in (4) from the Santa Barbara corpus of American English (Du Bois et al., 2000), Phil formulates a lunch invitation while Brad talks about a third party (Pat, referred to as *her*). Note that the square brackets indicate overlap. In (5), which is from the German Corpus (GECO; Schweitzer & Lewandowski, 2013), Speaker 31 describes where they live, while Speaker 32 develops a question. Note that the numbers in the square brackets indicate the length of the gap or overlap between speakers.

(4) Phil: ..W- .. w-.. why don't you call me at least a little bit later [maybe,
Brad: [Yeah].
Phil: and] we can [go do that].
Brad: [Can I] do that? Cause I .. she'll be .. Uh ..
Phil: [Ji- .. Jim and I are gonna] have lunch,
Brad: Uh .. I don't want to get her uh ..]

Phil: I don't know if you have plans or not. But we're gonna have lunch later at noon.

(5) 1. Speaker 32: Ja, (Yes) [0.11].
2. Speaker 31: Also, (Well,) [−0.01].
3. Speaker 32: klar (of course) [−0.13].
4. Speaker 31: Kries Böblingen und (district Böblingen and) [−0.2].
5. Speaker 32: Mhm (Uhm) [−0.35].
6. Speaker 31: ahm¨ …das kleine Dorf daneben Ehningen…da (uhm…the small village next to Ehningen…there) [0.08].
7. Speaker 32: Und (And) [−0.13].
8. Speaker 31: wohnen wir (we live) [−0.19].
9. Speaker 32: Du fährst eine dreiviertel Stunde? (you travel three-quarters of an hour?) [−0.12].
10. Speaker 31: Ja (Yes) [−0.02].

For these stretches of parallel talk, the issue of how the listener responds to the speaker's segment does not actually arise because the listener does not respond to the immediately preceding segment at all. Instead, they continue a segment they produced earlier. In these cases, the listener's response does not depend on the content of the speaker's immediately preceding segment—instead, their response depends on the content of a segment they previously produced. As a result, the duration of the speaker's segment does not limit the listener's planning time for their utterance, and so comprehension and production processes do not need to extensively overlap.

We determined the occurrence of such parallel talk by analyzing corpora of conversations in German, Dutch, and American English. The German Corpus (GECO) consisted of 24 face-to-face conversations between two strangers, the Dutch Corpus (Corpus Gesproken Nederlands; CGN) consisted of 18 face-to-face conversations between two friends or family members, and the English corpus (Santa Barbara Corpus of Spoken American English) consisted of 11 face-to-face conversations between two friends or family members. In all corpora, participants were free to talk about anything they liked, and so there was minimal constraint on their utterances. In the German and Dutch corpora, each row in the transcript represented a single word produced by a speaker. In the English corpus, each row in the transcript represented an intonational unit, which is a "stretch of speech uttered under a coherent intonation contour" (Du Bois, Schuetze-Coburn, Paolino, & Cummings, 1992, p. 17). We created segments by collapsing all words or intonation units produced by one speaker in a stretch of speech before a speaker switch (i.e., a same-speaker stretch of speech).

We determined the occurrence of parallel talk by coding whether or not each segment was a continuation of an earlier segment produced by the same speaker. We considered a segment to be a continuation if it contributed to completing an earlier, syntactically incomplete segment. For example, in (5) the segments in lines four, six, and eight were coded as continuations because word meaning and grammatical structure indicated that they belonged to one utterance produced by Speaker 31. If the previous segment was syntactically complete, then we considered the next segment to be a continuation only if the two segments were unambiguously linked by a pronoun or a conjunction.

Although we were primarily interested in segments that were continuations of a previous segment, we also included a number of other categories. For comparison with the continuations, the most important of these are direct responses. These direct responses occurred when one speaker produced an answer to the previous speaker's question (much like those utterances studied in laboratory experiments), expressions of disagreement (e.g., *That's right indeed* or *No, that was before my time*), literal repetitions of parts of the partner's segment (e.g., Speaker A: *...in a boarding school* Speaker B: *In a boarding school!*), segments that referred directly back to the previous speaker's preceding segment, such as with a pronoun (e.g., Speaker A: *I don't have the ambition to speak flawless French one day* Speaker B: *Which actually is almost impossible*), or elaborations and associations (e.g., Speaker A: *My boyfriend's brother had a neighbor who used to cut his lawn meticulously* Speaker B: *With nail scissors*). These direct responses are very similar to the utterances elicited in laboratory studies, where one speaker asks a question and the participant responds.

In the German corpus, we found that 43% of the segments were continuations. These continuations occurred either after the previous speaker had produced a backchannel, such as *uh huh* or *yeah* (19%), or after the previous speaker had produced a segment of their own (24%). In contrast, only 17% of the segments were direct responses. In the Dutch corpus, 48% of the segments were continuations, with 9% produced after the previous speaker produced a backchannel and 39% produced after the previous speaker produced a segment. Only 21% of the segments were direct responses. Finally, in the English corpus 30% of the segments were continuations, either after a backchannel (16%) or after another segment (14%). In this corpus, the proportion of direct responses was 24%—much higher than in the Dutch and German corpora.

Although there were differences across the corpora, our analysis demonstrates that parallel talk regularly occurs in different languages and conversational settings. In cases of such parallel talk, speakers continue a segment they have produced previously, rather than directly responding to the immediately preceding segment produced by the previous speaker. As a result, the listener can plan the content of their utterance independently from the content of the current speaker's utterance. In these cases, the question of how speakers manage to respond to each other's utterances so quickly does not arise—the speakers do not directly respond to each other at all, and the duration of the speaker's segment does not limit the listener's planning time. Thus, language production may be particularly difficult in laboratory tasks because speakers are encouraged to directly respond to each other, and produce pragmatically appropriate utterances (e.g., an answer to a question).

Note that these findings do not suggest that each speaker is holding a separate monolog. Informal inspection of the corpora suggested that successive segments in parallel talk may appear unrelated (i.e., segment two may not appear to be a direct response to segment one), but the turns developed by the two speakers are often related. Speakers usually refer to a common theme, as illustrated in (5), where both speakers talk about Speaker 31's home town. Thus, interlocutors are conversing with each other, but there is not necessarily a close content dependency between their utterances.

4.2 Incrementality and disfluency in dialog

Language production may also be difficult in laboratory tasks because speakers are typically encouraged to produce well-formed utterances, which are syntactically complete and do not contain any disfluencies, such as *uh* or *um*. As a result, participants are encouraged to plan a full utterance before speaking—if they do not, then they risk producing disfluent utterances. Planning in this way may make production difficult, given that there is much evidence for incremental planning in monolog (see Section 2.2). One consequence of this incrementality is that speakers are often disfluent, producing filled pauses such as *uh* or *um* (see Section 2.2.1). Although there is much research showing that utterances are disfluent, this disfluency has been underestimated by theories of dialog, and particularly by Levinson and Torreira (2015), because participants in laboratory tasks are discouraged from producing disfluent utterances. In particular, they have focused on fluent, idealized utterances, with the implicit assumption that disfluencies need to be excluded to study the mechanisms of conversation in their purest form (e.g., Bögels, Magyari, & Levinson, 2015). This point is important because it suggests that speech elicited in laboratory tasks designed to understand the mechanisms of language production in dialog may be very different, and in fact more difficult to produce, from speech as it naturally occurs.

To investigate how much conversational speech deviates from laboratory speech, I conducted further analyses of the Santa Barbara Corpus described in Section 4.1, focusing again on the 11 face-to-face conversations between two people. This corpus has already been used to study disfluency in an analysis by Tottie (2014), but Tottie focused solely on the occurrence of *uh* and *um*. These filled pauses are thought to mark hesitations by the speaker, and could be used to hold the floor while further planning occurs (see Section 2.2.1). I was interested in these filled pauses, but when analyzing the corpus for instances of parallel talk, I noticed that utterances could be disfluent in a number of different ways. For example, speakers often produced discourse markers, such as *well* or *you know*, which are "sequentially dependent elements which bracket units of talk" (Schriffin, 1987). They can be removed from an utterance without altering its meaning or grammaticality (Schourup, 1999). Much like filled pauses, speakers may produce these discourse markers as hesitations, to buy time for further planning. Research suggests that different filled pauses and discourse markers likely have different functions (e.g., Clark & Fox Tree, 2002; Fox Tree & Schrock, 2002; Fuller, 2003). However, my aim was not to determine the different uses of these filled pauses and discourse markers, but rather to illustrate that they occur and contribute to the (dis)fluency of dialog.

Additionally, utterances were often incomplete (6) or contained repetitions (often referred to as self-repairs in the Conversation Analysis literature; e.g., Schegloff, Jefferson, & Sacks, 1977), taking many attempts before successful articulation (7). In these instances, speakers had likely planned part of their utterance, and finished articulating it before they had planned the next part of their utterance. As a result, they abandoned or reformulated their utterance. In other words, incomplete and repeated utterances provide further evidence that planning is incremental. Table 1 provides counts and percentages for the different disfluency categories I considered. I will discuss each of these categories in more detail below, but a full coding criteria (along with examples) can be found at https://osf.io/7aphq/.

(6) Lynne: Cause y- I mean you get so tired.
(7) Lenore: I thought they used the horsehooves in .. for gelatin.

Although previous research has extensively quantified the frequency of discourse markers, repetitions, and filled pauses in corpora (e.g., Crible, 2019; Crible, Degand, & Gilquin, 2017; Crible, Dumont, Grosman, & Notarrigo, 2019; Crible & Pascual, 2020), this work has not considered these findings in the context of theories of dialog, such as Levinson and Torreira (2015). Furthermore, these corpora have often been based on highly restricted tasks, such as describing a route around a map (e.g., Branigan et al., 1999), and have tended to focus on limited disfluency types. Knowing what people say and how they speak in natural dialog is not only critical for determining whether laboratory speech is a good proxy for natural speech, but also for generating theories of speaking in dialog.

Before I discuss the coding criteria I used for identifying disfluent utterances, it is worth noting that previous research has shown that backchannels are common in spontaneous conversation (e.g., Knudsen, Creemers, & Meyer, 2020). The forms and functions of backchannels have been widely discussed from linguistic and psychological perspectives (e.g., Bangerter & Clark, 2003; Clark & Krych, 2004; Tolins & Fox Tree, 2014). They indicate to the present speaker that they should continue talking either by proceeding in their narrative or elaborating it (e.g., Schegloff, 1982, 2000; Tolins & Fox Tree, 2016). These backchannels are unlikely to contribute to disfluency—in fact, they likely contribute to the flow of dialog by allowing the listener to respond without planning a full utterance. However, I still quantified their occurrence because some discourse markers (such as *hmm*) could be produced as backchannels. Table 1 shows that 17% of the

Table 1 Frequencies (*n*) and proportions (%) of backchannels, incomplete segments, repetitions, resumptions, discourse markers, and filled pauses for segments in the Santa Barbara Corpus of Spoken American English.

	Total segments N = 3190	
	N	**%**
Backchannels	533	16.71
Incomplete segments	912	28.59
Interruptions	300	9.40
Repetitions (or self-repairs)	571	17.90
Resumptions (after interruption)	69	2.16
Segments containing at least one filled pause	531	16.65
Segments containing at least one discourse marker	879	27.55
Disfluent segments, containing at least one category	1854	58.12

Note that these categories were not mutually exclusive, and so a segment could belong to more than one category (i.e., it could contain both a filled pause and a discourse marker). The final row in the table shows the number of segments that were disfluent, and contained at least one category. In particular, a segment was disfluent if it was incomplete, interrupted, repeated, resumed, or contained a filled pause or discourse marker, regardless of how many of these phenomena occurred in the segment.

segments were backchannels (calculated as the number of segments containing a backchannel divided by the total number of segments).

Incomplete segments were those that contained an incomplete word or were abandoned by the speaker and were not resumed in any of the surrounding segments (i.e., the whole segment was incomplete). Incomplete segments also included those in which the speaker was interrupted by their partner and so did not finish their utterance. I also considered segments in which the speaker repeated themselves (e.g., *you have to-to graduate*) to be incomplete because the initial portion was incomplete and subsequently repeated. Note that segments could be incomplete in more than one way. For example, it could contain an incomplete word, be resumed, and then subsequently be abandoned by the speaker so the whole segment is incomplete. I did not determine how many times each segment was incomplete—it was considered incomplete if it belonged to any of these categories. In total, 29% of the segments were incomplete, with 9% of them being incomplete because the other speaker interrupted.

When segments were incomplete, speakers often began a new segment by repeating part of their earlier, incomplete segment. To determine how often speakers repeated part of their segment, I identified segments that contained repetitions or that were resumed after an interruption by another speaker. Again, I did not determine how many times each segment was repeated. Rather I considered an utterance to be a repetition if it was repeated at least once. In total, 18% of the segments were repetitions, while 2% were resumptions of an earlier, interrupted segment.

When coding the discourse markers and filled pauses, I considered words (such as *well* or *you know*) and sounds (such as *uh* or *um*) to be discourse markers or filled pauses if they could be removed from the segment without altering the speaker's meaning. For example, *you know* would be considered a discourse marker in a segment such as *And doing it and stuff you know*, but not in a segment such as *Do you know what I mean*?. Table 2 shows the counts and percentages for the individual filled pauses and discourse markers. Segments could contain multiple occurrences of the *same* filled pause or discourse marker. For example, the speaker could produce *uh* multiple times in the same segment. But since I was interested in how many segments contained at least one occurrence of each type of filled pause or discourse marker, Table 1 shows the number of times the speaker produced a particular type of filled pause or discourse marker at least once in a segment. In total, 17% of the segments contained at least one filled pause, and 28% of the segments contained at least one discourse marker.

To determine how often segments were disfluent, I determined how many were incomplete, interrupted, repeated, resumed, or contained a filled pause or a discourse marker. Segments were considered disfluent if they fell into any one of these categories. In total, 58% of the segments were disfluent, and so around only 40% of the segments contained no disfluency and were similar to the idealized utterances elicited in laboratory tasks studying the mechanisms of speaking in dialog.

These findings add to an existing body of research that has shown that spontaneous speech is disfluent (see Section 2.2.1), and suggest that speech planning is incremental. Speakers are likely incremental in this way because planning while comprehending is cognitively demanding (e.g., Oomen & Postma, 2001). Although corpora analyses do not allow us to draw conclusions about the direction of causality, there is some evidence that the fluency of speech is affected when speakers dual-task production and comprehension. For example, Boiteau, Malone, Peters, and Almor (2014) had participants conduct a visuomotor tracking task while simultaneously interacting with a confederate. Participants' tracking performance declined towards the end of the confederate's turn, suggesting they began response planning at this point. Participants' speech rate was also affected by concurrent tracking when they had to plan a response compared to when they just had to listen, but there was no evidence that planning while listening increased the number of disfluencies participants produced. However, the authors considered only *um*s and *uh*s, but it is clear from Tables 1 and 2 that there are many other types of disfluencies.

This incrementality (and disfluencies, by extension) invites parallel talk. Speakers (Speaker A) do not plan their full utterance before they speak, and so they may often pause or hesitate while they plan later parts of their utterance, leading to disfluent speech. This hesitation allows the other speaker (Speaker B) to jump in and articulate their own increment. Speaker A then articulates the rest of their utterance, and so they do not directly respond to the immediately preceding utterance of Speaker B. Thus, incrementality, disfluencies, and parallel talk are closely related to each other.

These findings have important consequences for the way we think about language during dialog. First, they suggest that the utterances we study in the laboratory are very different from the utterances speakers actually produce in natural conversation. This point may seem obvious, but it has important consequences for Levinson and Torreira (2015) theory, which has been used to motivate many studies investigating the mechanisms of speaking during dialog. In particular, Levinson and Torreira (2015) claim that next speakers

Table 2 Frequencies (*n*) and proportions (%) of different types of filled pauses and discourse markets in the Santa Barbara Corpus of Spoken American English.

Filled pause	***N***	**%**
Uh	326	10.22
Oh	134	4.20
Hm	66	2.07
Huh	23	0.72
Ah	11	0.35
Uhuh	4	0.13
Aw	7	0.22
Total filled pauses	571	17.42
Discourse markers		
You know	315	9.88
Well	252	7.90
So	170	5.33
Like	164	5.14
I mean	115	3.61
Kinda	74	2.32
Geez	59	1.85
Man	59	1.85
Oh God	34	1.07
Right	33	1.04
Pretty	28	0.88
See	27	0.85
Really	19	0.60
Now	17	0.53
Sorta	15	0.47
Anyway	13	0.41
Total discourse markers	1394	43.70

Note that these categories were not mutually exclusive, and so a segment could contain more than one filled pause or discourse marker.

must complete all stages of response planning as early as possible (i.e., as soon as they can identify the gist of the current speaker's utterance) if they are to achieve timely turn-taking and respond within 200 ms. But such early-planning may not be necessary in natural conversation—speakers could use disfluencies to hold their turn while planning their utterance, thus minimizing the overlap between production and comprehension processes.

Relatedly, experimental studies investigating production in dialog likely make production harder than it needs to be. First, participants are often encouraged to plan well-formed utterances, and any utterances containing disfluencies are often excluded from analyses. Participants may thus be discouraged from planning incrementally, and may instead plan their complete utterance before they speak in an effort to ensure they produce well-formed utterances. Relatedly, our corpora analyses (Corps et al., 2022) have demonstrated that speakers do not always directly respond to each other—instead, they develop their utterances in parallel and continue an utterance they produced previously. This situation is very different from laboratory tasks, where participants often need to directly respond to the previous speaker and the content of their own utterance depends on the content of the previous speaker's utterance. As a result, speakers likely engage in more extensive advance planning (resulting in a larger overlap between production and comprehension) in laboratory tasks than there needs to be in natural conversation, thus contributing to turn gaps longer than 200 ms.

In sum, it is clear that these theories are missing an important part of natural speech—namely, that speakers are highly disfluent. Thus, these results have important methodological and theoretical consequences, and suggest that we need to study production both in highly controlled laboratory tasks and in natural conversation if we are to build a clear picture of the mechanisms of speaking during dialog (see also De Ruiter & Albert, 2017). In particular, future experimental work could take excerpts from speech corpora and test how disfluencies affect the accuracy of when speakers articulate their responses. Additionally, they could also test how disfluencies affect how participants distribute their attention between response planning and simultaneous comprehension. Finally, research could investigate whether parallel talk is more common in instances where speakers hesitate and produce disfluencies. Testing these hypotheses would provide insight into how comprehension, response planning, and articulation are interwoven during conversation, and would allow researchers to develop theories of language production in natural dialog.

What these findings demonstrate, however, is that we currently do not have a clear picture of speaking in dialog, like we do in monolog, because these studies have tended to focus on highly idealized utterances, often ignoring the fact that production is highly incremental, flexible, and far from perfect.

5. Conclusions

During dialog, interlocutors take turns at speaking with little gap or overlap between their contributions. But language production in monolog is comparatively slow. Theories of dialog tend to agree that interlocutors manage these timing demands by planning a response early, before the current speaker reaches the end of their turn. As a result, there is overlap between production and comprehension processes. Much experimental research supports these theories, but this research also suggests that planning a response early, while simultaneously comprehending, is difficult. Does language production need to be this difficult during dialog? In other words, is early-planning always necessary?

In the second half of this chapter, I discussed research from our lab that suggests the answer to this question is no. In particular, we analyzed corpora of naturally occurring conversations in German, Dutch, and English. We found that speakers often do not directly respond to each other during dialog—instead, they continue an utterance they produced earlier. In these instances of parallel talk, the next speaker's response does not depend on the content of the current speaker's utterance, and so the next speaker's planning time is not constrained by the current speaker's utterance. As a result, comprehension and production do not need to extensively overlap.

This parallel talk likely occurs because speakers are highly incremental. In particular, we also found that speakers are highly disfluent, suggesting they do not plan a full utterance before beginning articulation. This incrementality has not been considered by theories and experimental studies of dialog, which typically focus on idealized utterances. Note that I am not claiming that early-planning never occurs—in fact, it is likely particularly useful in highly constrained interactions (such as question-answering), where speakers do directly respond to each other and must do so in a timely manner. But together, these corpora analyses demonstrate that language production studied in laboratory experiments is very different from how language production actually occurs in natural conversation. Thus, further research using naturalistic tasks is needed to investigate the mechanisms of dialog.

Acknowledgments

I thank Antje Meyer and Martin Pickering for useful comments on earlier versions of this chapter.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research*, *32*, 25–36.

Arnold, J. E., & Tanenhaus, M. K. (2011). Disfluency effects in comprehension: How new information can become accessible. *The Processing and Acquisition of Reference*, 197–217.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science*, *15*, 578–582.

Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, *27*, 195–225.

Barthel, M., Meyer, A. S., & Levinson, S. C. (2017). Next speakers plan their turn early and speak after turn-final "go-signals". *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00393.

Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01858.

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.

Bögels, S., Kendrick, K. H., & Levinson, S. C. (2015). Never say no .. how the brain interprets the pregnant pause in conversation. *PLoS One*, *10*. https://doi.org/10.1371/journal.pone.0145474.

Bögels, S., Kendrick, K. H., & Levinson, S. C. (2020). Conversational expectations get revised as response latencies unfold. *Language, Cognition and Neuroscience*, *35*, 766–779.

Bögels, S., & Levinson, S. C. (2017). The brain behind the response: Insights into turn-taking in conversation from neuroimaging. *Research on Language and Social Interaction*, *50*, 71–89.

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, *5*. https://doi.org/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*, 295–311.

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*, 123–147.

Bosker, H. R., Tjiong, V., Quené, H., Sanders, T., & De Jong, N. H. (2015). Both native and non-native disfluencies trigger listeners' attention. In *Disfluency in Spontaneous Speech: DISS 2015* An ICPhS Satellite Meeting. DISS2015.

Branigan, H., Lickley, R., & McKelvie, D. (1999). Non-linguistic influences on rates of disfluency in spontaneous speech. In *Proceedings of the 14th International Conference of Phonetic Sciences* (pp. 387–390).

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1482–1493.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*, 325–337.

Brown-Schmidt, S., & Konopka, A. E. (2015). Processes of incremental message planning during conversation. *Psychonomic Bulletin & Review*, *22*, 833–843.
Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, *54*, 592–609.
Butterworth, B. (1989). Lexical access in speech production. In W. Marslen-Wilson (Ed.), *Lexical representation and process*. Cambridge, MA: MIT Press.
Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: The HAROLD model. *Psychology and Aging*, *17*, 85–100.
Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177–208.
Caramazza, A., & Miozzo, M. (1998). More is not always better: A response to Roelofs, Meyer, and Levelt. *Cognition*, *69*, 231–241.
Chafe, W. (1992). Information flow in speaking and writing. *The Linguistics of Literacy*, *21*, 17–29.
Clark, H. H. (1994). Managing problems in speaking. *Speech Communication*, *15*, 243–250.
Clark, E. V. (1997). Conceptual perspective and lexical choice in acquisition. *Cognition*, *64*, 1–37.
Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*, 73–111.
Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*, 62–81.
Collard, P., Corley, M., MacGregor, L. J., & Donaldson, D. I. (2008). Attention orienting effects of hesitations in speech: Evidence from ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 696–702.
Cook, A. E., & Meyer, S. (2008). Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 886–899.
Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you say it: Disfluency in speech affects the comprehension process. *Cognition*, *105*, 658–668.
Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it. *Cognition*, *175*, 77–95.
Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse Processes*, *55*, 230–240.
Corps, R. E., Knudsen, B., & Meyer, A. S. (2022). Overrated gaps: Inter-speaker gaps provide limited information about the timing of turns in conversation. *Cognition*, *223*. https://doi.org/10.1016/j.cognition.2022.105037.
Crible, L. (2019). *Discourse markers and (dis)fluency. Forms and functions across languages and registers*. Amsterdam/Philadelphia: John Benjamins.
Crible, L., Degand, L., & Gilquin, G. (2017). The clustering of discourse markers and filled pauses: A corpus-based French-English study of (dis) fluency. *Languages in Contrast*, *17*(1), 69–95.
Crible, L., Dumont, A., Grosman, I., & Notarrigo, I. (2019). Application of an interoperable annotation scheme. *Fluency and Disfluency Across Languages and Language Varieties*, *4*, 17.
Crible, L., & Pascual, E. (2020). Combinations of discourse markers with repairs and repetitions in English, French and Spanish. *Journal of Pragmatics*, *156*, 54–67.
De Ruiter, J. P., & Albert, S. (2017). An appeal for a methodological fusion of conversation analysis and experimental psychology. *Research on Language and Social Interaction*, *50*, 90–107.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.

DeLong, K. A., Chan, W., & Kutas, M. (2018). Similar time courses for word form and meaning preactivation during sentence comprehension. *Psychophysiology*. https://doi.org/10.1111/pstp.13312.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2017). Is there a replication crisis? Perhaps. Is this an example? No: A commentary on Ito, Martin, and Nieuwland (2016). *Language, Cognition and Neuroscience*, *32*, 966–973.

Dhooge, E., & Hartsuiker, R. J. (2012). Lexical selection and verbal self-monitoring: Effects of lexicality, context, and time pressure in picture-word interference. *Journal of Memory and Language*, *66*, 163–176.

Du Bois, Chafe, W. L., Meyer, C., Thompson, S. A., & Martey, N. (2000). Santa Barbara corpus of Spoken American English, Part 1–4. Philadelphia: Linguistic Data Consortium.

Du Bois, J. W., Schuetze-Coburn, S., Paolino, D., & Cummings, S. (1992). *Discourse transcription (Santa Barbara Papers in Linguistics). Vol. 4*. Santa Barbara: University of California, Santa Barbara, Department of Linguistics.

Eklund, R., & Shriberg, E. (1998). Crosslinguistic disfluency modelling: A comparative analysis of Swedish and American English human–human and human–machine dialogues. In *Vol. 6. 5th International Conference on Spoken Language Processing, 30th November–4th December, 1998, Sydney, Australia* (pp. 2627–2630).

Fairs, A., Bögels, S., & Meyer, A. S. (2018). Dual-tasking with simple linguistic tasks: Evidence for serial processing. *Acta Psychologica*, *191*, 131–148.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*, 210–233.

Ferreira, V. S., & Slevc, L. R. (2007). Grammatical encoding. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 453–470). Oxford University Press.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*, 57–84.

Fox Tree, J. E. (2010). Discourse markers across speakers and settings. *Language and Linguistics Compass*, *4*, 269–281.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, *62*, 151–167.

Fox Tree, J. E., & Schrock, J. C. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language*, *40*, 280–295.

Fox Tree, J. E., & Schrock, J. C. (2002). Basic meanings of you know and I mean. *Journal of Pragmatics*, *34*, 727–747.

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, *31*, 931–952.

Fuller, J. M. (2003). Use of the discourse marker like in interviews. *Journal of Sociolinguistics*, *7*, 365–377.

Garrod, S., & Pickering, M. J. (2007). Alignment in dialogue. *The Oxford Handbook of Psycholinguistics*, 443–451.

Garrod, S., & Pickering, M. J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00751.

Gisladottir, R. S., Chwilla, D. J., & Levinson, S. C. (2015). Conversation electrified: ERP correlates of speech act recognition in underspecified utterances. *PLoS One*, *10*. https://doi.org/10.1371/journal.pone.0120068.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. *Journal of Experimental Psychology: General*, *118*, 13–42.

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*, 544–569.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*, B1–B14.
Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274–279.
Griffin, Z. M., & Spieler, D. H. (2006). Observing the what and when of language production for different age groups by monitoring speakers' eye movements. *Brain and Language*, *99*, 272–288.
Hartsuiker, R. J., & Notebaert, L. (2009). Lexical access problems lead to disfluencies in speech. *Experimental Psychology*, *57*, 169–177.
He, J., Meyer, A. S., & Brehm, L. (2021). Concurrent listening affects speech planning and fluency: The roles of representational similarity and capacity limitation. *Language, Cognition and Neuroscience*, *36*, 1258–1280.
Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*, 555–568.
Howes, C., Purver, M., Healey, P. G., Mills, G. J., & Gregoromichelaki, E. (2011). On incrementality in dialogue: Evidence from compound contributions. *Dialogue and Discourse*, *2*, 279–311.
Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144.
Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, *86*, 157–171.
Ito, A., Gambi, C., Pickering, M. J., Fullenbach, K., & Husband, E. M. (2020). Prediction of phonological and gender information: An event-related potential study in Italian. *Neuropsychologia*, *136*. https://doi.org/10.1016/j.neuropsychologia.2019.107291.
Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). Why the a/an prediction effect may be hard to replicate: A rebuttal to DeLong, Urbach, and Kutas (2017). *Language, Cognition and Neuroscience*, *32*, 974–983.
Iwasaki, N., Vigliocco, G., & Garrett, M. F. (1998). Adjectives and adjectival nouns in Japanese: Psychological processes in sentence production. In D. J. Silva (Ed.), *Vol. 8. Japanese/Korean linguistics* (pp. 93–106). Stanford, CA: Center for the Study of Language and Information.
Jongman, S. R. (2021). The attentional demands of combining comprehension and production in conversation. In *Psychology of learning and motivation* (pp. 95–140). Academic Press.
Jongman, S. R., & Meyer, A. S. (2017). To plan or not to plan: Does planning for production remove facilitation from associative priming? *Acta Psychologica*, *181*, 40–50.
Jongman, S. R., Piai, V., & Meyer, A. S. (2020). Planning for language production: The electrophysiological signature of attention to the cue to speak. *Language, Cognition and Neuroscience*, *35*, 915–932.
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*, 133–156.
Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, *14*, 185–209.
Knudsen, B., Creemers, A., & Meyer, A. S. (2020). Forgotten little words: How backchannels and particles may facilitate speech planning in conversation? *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.593671.
Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, *66*, 143–162.
Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology*, *73*, 1–40.

Koudenburg, N., Postmes, T., & Gordijn, E. H. (2013). Conversational flow promotes solidarity. *PLoS One*, *8*. https://doi.org/10.1371/journal.pone.0078363.

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*, 621–647.

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

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38.

Levelt, W. J., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*, 122–142.

Levinson, S. C. (1995). Interactional biases in human thinking. In E. N. Goody (Ed.), *Social intelligence and interaction* (pp. 221–260). Cambridge: Cambridge University Press.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.0073.

Magyari, L., De Ruiter, J. P., & Levinson, S. C. (2017). Temporal preparation for speaking in question-answer sequences. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00211.

Menenti, L., Gierhan, S. M. E., 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*, 1173–1182.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture–word interference experiments. *Journal of Memory and Language*, *35*, 477–496.

Meyer, A. S., Alday, P. M., Decuyper, C., & Knudsen, B. (2018). Working together: Contributions of corpus analyses and experimental psycholinguistics to understanding conversation. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00525.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*, B25–B33.

Miozzo, M., & Caramazza, A. (1997). Retrieval of lexical–syntactic features in tip-of-the tongue states. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1410–1423.

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., et al. (2018). Large-scale replication study reveals a limit on probablistic prediction in language comprehension. *eLife*, *7*. https://doi.org/10.7554/eLife.33469.

Oomen, C. C. E., & Postma, A. (2001). Effects of divided attention on the production of filled pauses and repetitions. *Journal of Speech, Language, and Hearing Research*, *44*, 997–1004. https://doi.org/10.1044/1092-4388(2001/078).

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 539–557.

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

Rajah, M. N., Languay, R., & Grady, C. L. (2011). Agre-related changes in right middle frontal gyrus volume correlate with altered episodic retrieval activity. *The Journal of Neuroscience*, *35*, 17941–17954.

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., et al. (2005). Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. *Cerebral Cortex*, *15*, 1676–1689.

Roberts, F., & Francis, A. L. (2013). Identifying a temporal threshold of tolerance for silent gaps after requests. *The Journal of the Acoustical Society of America*, *133*, EL471–EL477.

Roberts, F., Francis, A. L., & Morgan, M. (2006). The interaction of inter-turn silence with prosodic cues in listener perceptions of "trouble" in conversation. *Speech Communication*, *48*, 1079–1093.
Roberts, F., Margutti, P., & Takano, S. (2011). Judgments concerning the valence of inter-turn silence across speakers of American English, Italian, and Japanese. *Discourse Processes*, *48*, 331–354.
Roelofs, A. (2008). Attention, gaze shifting, and dual-task interference from phonological encoding in spoken word planning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1580–1598.
Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00307.
Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization of turn-taking for conversation. *Langugage*, *50*, 696–735.
Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology*, *60*, 362–367.
Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In D. Tannen (Ed.), *Analyzing text and talk* (pp. 71–93). Georgetown: Georgetown University Press.
Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, *29*, 1–63.
Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*, 361–382.
Schourup, L. (1999). Discourse markers. *Lingua*, *107*, 227–265.
Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*, 86–102.
Schriffin, D. (1987). *Discourse markers. No. 5*. Cambridge: Cambridge University Press.
Schweitzer, A., & Lewandowski, N. (2013). Convergence of articulation rate in spontaneous speech. In *INTERSPEECH 2013, Lyon, France* (pp. 525–529).
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*, 1662–1670.
Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, E4687–E4696.
Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*, 25–38.
Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, *73*, 205–246.
Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*, 10587–10592.
Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, *30*, 485–496.
Swets, B., Jacovina, M. E., & Gerrig, R. J. (2013). Effects of conversational pressures on speech planning. *Discourse Processes*, *50*, 23–51.
Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2022). Fast response times signal social connection in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, *119*. https://doi.org/10.1073/pnas.2116915119.

Tolins, J., & Fox Tree, J. E. (2014). Addressee backchannels steer narrative development. *Journal of Pragmatics*, *70*, 152.

Tolins, J., & Fox Tree, J. E. (2016). Overhearers use addressee backchannels in dialog comprehension. *Cognitive Science*, *40*, 1412–1434.

Tottie, G. (2014). On the use of uh and um in American English. *Functions of Language*, *21*, 6–29.

Urbach, T. P., DeLong, K. A., Chan, W. H., & Kutas, M. (2020). An exploratory data analysis of word form prediction during word-by-word reading. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 20483–20494.

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

Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science*, *8*, 314–317.

Vigliocco, G., Vinson, D. P., Martin, R. C., & Garrett, M. F. (1999). Is "count" and "mass" information available when the noun is not? An investigation of tip of the tongue states and anomia. *Journal of Memory and Language*, *40*, 534–558.

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

Wheeldon, L. (2013). Producing spoken sentences: The scope of incremental planning. In P. Perrier, & P. L. Verlag (Eds.), *Cognitive and physical models of speech production, speech perception, and production-perception integration*.

Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, *346*, 165–168.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*(3), 449–455.

This page intentionally left blank

CHAPTER THREE

Language production under message uncertainty: When, how, and why we speak before we think

Arella E. Gussow*
Department of Psychology, University of Wisconsin-Madison, Madison, WI, United States
*Corresponding author: e-mail address: gussow@wisc.edu

Contents

Abstract

Speakers occasionally begin speaking before they are certain of what to say—that is, before they know what *message* they want to communicate. But theories of language production do not account for these situations, and researchers typically assume that language production begins with a pre-determined message. Given the relative lack of research on production under message uncertainty, here I gather evidence from several neighboring areas to introduce issues of message uncertainty and motivate future research. The arguments raised will point to how and why message uncertainty should be incorporated into language production research: First, message uncertainty might

Psychology of Learning and Motivation, Volume 78
ISSN 0079-7421
https://doi.org/10.1016/bs.plm.2023.02.005

be more common than assumed, and language production research would benefit from identifying the incidence and types of message uncertainty. Second, situations of message uncertainty carry implications for utterance forms and the time course of production, given a highly flexible production system that uses different strategies to mitigate difficulties of planning under uncertainty. Third, uncertainty is ubiquitous in other cognitive domains, and in particular uncertainty in the motor domain, such as beginning an action sequence before the goal is determined, shows many parallels with uncertainty in language. These parallel lines of research could inform and benefit each other, contributing to our understanding of which domain-general cognitive principles are used for planning and acting when there is only partial information about the goal. Finally, investigating situations of message uncertainty would not only address a common everyday context of production, but could also inform theories of language production and cognition more generally.

1. Introduction

Language production is often described as the process of turning thoughts into speech: If you want to communicate a certain thought, or *message*, you must turn it into a verbal utterance that others can understand. An underlying assumption of this idea, however, is that you already have a particular message to communicate before you plan your utterance. But this is not always the case—we don't *always* think before we speak.

In fact, in many situations we are pressured to begin speaking quickly, even if we haven't decided what to say yet. These situations don't necessarily have to be stressful or high-stakes: imagine your partner asks what you'd like to do this weekend, but you're not sure what you're in the mood for. You probably won't just ignore them, as they might think you didn't hear them, or infer you're doing it intentionally. Instead, you might attempt some utterance, even if it is not quite well-formed yet: "um… how about… uh…a hike“, "picni–no, actually let's go for a hike," "picnic–a picnic and then a hike," "well we had a picnic last week already…so maybe a hike?." Or perhaps you'll realize it's going to take you some time to decide, so you'll just say "let me think about it for a moment", then think, and then speak.

It seems quite intuitive that occasionally we begin planning our utterances, or even speaking, before we've settled on the message. As in the previous example, during conversation a speaker might be expected to respond rapidly to an interlocutor's question or comment (*what would you like to do this weekend?*), despite not having decided what to answer. In other cases, a speaker might be describing a rapidly unfolding scene without knowing what is going to occur (*the ball is passed to*…), answering a general knowledge question that requires retrieving information from memory (*what main ingredients go in a daquiri?*), or even debating between multiple message options to start a conversation (*the weather or the virus?*). Despite this intuition, however, language production research has typically assumed that message formulation precedes other stages of utterance planning; that is, that the speaker has a predetermined message to communicate before they begin utterance planning. Much less is known about circumstances in which speakers begin utterance planning *before* they are certain about their message.

In this review, I will consider situations of message uncertainty and how they affect utterance formulation, whether in the choice of utterance forms or in the time course of production. I will begin with a brief description of how language production models view the message processing component. Then I will examine theories of incremental production to understand what planning strategies speakers use when simultaneously planning a message and an utterance—in particular, how much of the utterance is planned in advance, and how the available message information affects which linguistic form is chosen to express it. Next, evidence of production under uncertainty will be considered in context, whether from error and disfluency patterns or from everyday situations likely to induce uncertainty. Finally, I will turn to the motor domain to review evidence of uncertainty in the planning of motor actions, and how it could inform research on message uncertainty in language production.

2. From message to utterance

2.1 The message in language production models

A *message* in language production is a package of information that the speaker intends to communicate (Garrett, 1989; Levelt, 1989). The message is therefore the motivation for speech itself, the reason a speaker begins to formulate their utterance (Garrett, 1989). Given this intent, the speaker must select the precise information needed to be expressed; drawing on several sources of knowledge including the perceptual environment and the conversational context (Guhe, Habel, & Tschander, 2004; Konopka & Brown-Schmidt, 2014). The speaker's goal is then to turn this package of information, or *preverbal message*, into an utterance, i.e. a verbal formulation that can be comprehended by others.

The main challenge in turning a preverbal message into an utterance is that there is no one-to-one mapping between message and utterance form. A given message can often be expressed in several different ways, and it is up to the speaker to decide which words and sentence structures to use (Bock, 1982, 1995). For example, in describing a simple event of a dog chasing a cat, a speaker can choose between the active form "the dog chased the cat" or the passive form "the cat was chased by the dog". Moreover, the speaker can choose more specific lexical forms like "the Labrador chased the Ragdoll," or even call the animals by their given names, "Rebel chased Gigi." Utterance planning can therefore be seen as a series of implicit decision making, where choices are influenced by various factors including perceptual context (Gleitman, January, Nappa, & Trueswell, 2007), shared knowledge between speakers (Heller, Skovbroten, & Tanenhaus, 2009), frequency and priming of words (Bock, 1986; Branigan & McLean, 2016), and several other domain-general cognitive constraints such as memory demands (MacDonald, 2013).

Decisions about the form of an utterance are thought to occur in the utterance "formulator," as described by Levelt in his influential model of language production (Levelt, 1989). The formulator takes the preverbal message as input and passes it through several processing stages for linguistic encoding, or *utterance formulation*. There is some variation across models in how to define or divide these particular stages, but a simplified account would include choosing which words to use (*lexical selection*), arranging them with the appropriate grammatical markings in a sentence structure (*syntactic assembly*), and encoding the particular sounds, or *phonemes*, required for pronunciation (*phonological encoding*). The speaker can then articulate the resulting utterance, using the utterance plan to guide articulation and monitor for errors. If successful, the produced utterance expresses the message intended by the speaker.

Although researchers generally agree that a preverbal message is the required input to the formulator, the particular content and form of this message remains rather vague and difficult to define (Bock, 1996; Konopka & Brown-Schmidt, 2014). Levelt (1989) notes, however, that the preverbal message is the only input to the utterance formulator and therefore must include necessary and sufficient information for the next processing stages (e.g., lexical selection). But it is difficult to determine what exactly is necessary and sufficient, or what would constitute a well-formed message for further processing (Chang, Dell, & Bock, 2006; Konopka & Brown-Schmidt, 2014).

Another challenge in understanding the message input to the formulator is that message formulation itself remains dynamic throughout utterance formulation. That is, message formulation occurs over time such that the message itself can continue to develop or change even after speaking has begun (Brown-Schmidt & Konopka, 2008, 2015; Brown-Schmidt & Tanenhaus, 2006). Clearly, some degree of message planning must be completed before the next stages of production can proceed, since the message constrains which words or sentence structures can be used to express it. However, message inputs to the formulator may vary in their degree of specificity or completeness (Kempen & Hoenkamp, 1987; Konopka & Brown-Schmidt, 2014). In fact, Levelt (1989) describes the process of utterance formulation as beginning with an input of either a message or a message *fragment*—suggesting that even only part of a message is enough to begin with utterance formulation. Again, however, it is unclear what size of a message fragment must be planned in advance for utterance formulation to begin.

2.2 Incremental production

The debate about how much of the message must be prepared in advance of production is one of the earliest in the history of psycholinguistics. Wundt (1900) argued that speakers must complete their message plan before beginning to speak. In describing the event of a dog chasing a cat, then, Wundt would suggest that the entire *gist* of the scene is encoded (Konopka & Brown-Schmidt, 2014)—that there is an event of chasing, that the dog is the chaser (the agent), the cat is the one being chased (the patient), etc—before utterance planning can begin. Paul (1880), however, argued that messages can be planned in smaller fragments, allowing them to be interleaved with production. In that case, the speaker might first encode and begin producing the word for only one of the participants in the event—e.g., the dog; and continue encoding the rest of the scene while production is already underway; i.e., while producing the words "the dog."

Paul's view therefore suggests incremental production: interleaving planning and speaking in order to maintain fluent speech. By extension, incrementality suggests that all intermediate processing stages of language production occur simultaneously on successive segments of the message (Kempen & Hoenkamp, 1987; Levelt, 1989). For example, once the first message segment (*the dog*) has completed grammatical encoding, it can proceed to phonological encoding while the next component (*chased*) begins grammatical encoding. Thus each component of the message is at a different stage of processing at all times.

There is now ample evidence that language production is indeed incremental, and that the degree of incrementality is under some strategic control (Ferreira & Swets, 2002). An important implication of incrementality is that upcoming portions of the utterance are being prepared while production is happening, allowing for online adjustments and interactions between processing stages (Brown-Schmidt & Konopka, 2015; Garrett, 1989; Smith & Wheeldon, 1999). This also means that incomplete messages can be processed for articulation while more information is gathered to complete the message (Dohsaka & Shimazu, 1996; Ferreira & Swets, 2002; Kempen & Hoenkamp, 1987)—as Levelt suggested, a message *fragment* is enough to begin with utterance formulation.

2.3 Incomplete messages

Despite the agreement that message processing can proceed incrementally, most experimental work has not accounted for cases of message uncertainty or incomplete messages in language production. Partly because the message is not easily defined or operationalized, experimental paradigms typically provide a very controlled, complete message that participants need to turn into an utterance: a picture of a scene to be described (Bunger, Papafragou, & Trueswell, 2013; Gleitman et al., 2007; Jaeger, Furth, & Hilliard, 2012; van de Velde, Meyer, & Konopka, 2014), simple questions to be answered (Chia & Kaschak, 2022), or picture naming (Meyer, 1996; Meyer & Schriefers, 1991; Strijkers, Holcomb, & Costa, 2011). This experimental control is useful for studying utterance formulation stages such as lexical choice and grammatical encoding, but it does not account for situations where message formulation itself remains dynamic throughout production, with potential effects on other planning stages during real-time production (Harley, 1984; Konopka & Brown-Schmidt, 2014).

However, a few studies did investigate how message formulation, or *message updating* after speech has begun, affects utterance planning and the resulting utterance forms (Brown-Schmidt & Konopka, 2008, 2015; Brown-Schmidt & Tanenhaus, 2006). In these studies, speakers describe visual displays where certain elements of the message are not immediately apparent, leading speakers to notice key message elements only after production had begun. By using a combination of eye-tracking and speech recording, researchers could track how soon after noticing new message information speakers are able to incorporate it into their utterance.

For example, Brown-Schmidt and Tanenhaus (2006) presented English-speaking participant dyads with identical displays of several images on separate screens. On every trial, one of the objects (e.g., a horse) was highlighted only for one of the participants (the speaker), who then had to name the target object so that the other participant (the listener) could click on it in their own display. On critical trials, the display included a contrast image: the same object as the target but in a different size (e.g., a small horse and a large horse; among several other objects). In those cases, the speaker would need to specify the size modifier in order for the listener to identify the correct target (the *small* horse). By tracking participants' eye movements throughout the trials, the researchers could identify when the speaker fixated on the contrast image; that is, when the speaker noticed that the message was not just horse but small horse, and how that timing affected the speaker's utterance.

Results showed that the utterance form depended on when the first fixation to the contrast image (the large, non-target horse) was, relative to when participants began naming the target (the small, target horse). That is, how soon before (or after) target onset did the participant notice they would need to include size information. Earlier fixations to the contrast image resulted in fluent utterances that incorporated the size information (*the small horse*)—presumably, participants noticed and planned the entire utterance, including the size information, before beginning to speak. Later fixations to the contrast image were associated with post-nominal repairs (after the noun; *the horse…uh the small one*); suggesting participants had to adjust their initial utterance plan while speaking. When fixations to the contrast image were intermediate, speakers were able to incorporate the adjective information pre-nominally, but with disfluencies (*thee uh small horse*).

These results suggest that utterances can be updated to incorporate new message information (the size contrast) even while speech is underway, though that might cause disfluencies at different points in the utterance depending on when the information is received. Moreover, Brown-Schmidt and Konopka (2008) showed that in Spanish, where the adjectives are typically post-nominal (*la mariposa pequeña*), participants could incorporate the adjective information even if they fixated on the contrast image late. This is because in Spanish, speakers can plan the size modifier (*pequeña*, the third word in the noun phrase) while producing the noun (*mariposa*, the second word), providing more time to notice and incorporate new message information while production is ongoing.

Evidently, incremental production is useful when a message changes unexpectedly after speaking has begun and the speaker needs to update their message mid-utterance. However, there are also situations where speakers begin speaking even though they know their message is incomplete. For example, Ferreira and Swets (2002) presented participants with arithmetic problems of varying difficulty. Participants' task was to respond "the answer is…" and the solution (e.g., *the answer is twenty-five*). Results showed that only when speakers were required to begin speaking quickly, their utterance durations were longer for harder problems, suggesting they were computing while speaking. The authors concluded that incrementality is under strategic control: when faced with a deadline participants will begin producing the utterance frame (*the answer is…*) and compute the solution as they speak; but without a deadline they will complete computations before beginning to speak. Interestingly, speech onset latencies were also modulated by problem difficulty, even in trials with a deadline: people took longer to begin speaking when faced with harder problems. This suggests that even when participants were rushed, some part of the message was encoded prior to speech beginning.

Although Ferreira and Swets (2002) did not frame their study as examining message updating per se, arguably their task presents another case of message formulation during production: participants computed the message itself while producing the utterance that expressed that same message. The fixed leading frame (*the answer is*) allowed speakers to begin producing portions of the utterance they were certain about, providing some leverage for computing the uncertain portion of the message they did not know yet (the problem solution). Although this strategic incrementality is different from the eye-tracking studies—where participants gleaned the message information from a visual display without knowing that the message would change – arguably both paradigms present a problem of incomplete messages. Moreover, results from both paradigms support the same conclusion: language production can be incremental even at the message level, providing more flexibility during online production.

2.4 Context-dependent incrementality

Another source of flexibility in production regards the order and the size of the chunks that are incrementally processed from message-to-utterance. Even if the entire message is available to the speaker prior to speaking, planning of a complex message often requires preparing smaller message chunks at a time in order to reduce the cognitive load and begin production sooner. In extensive prior work on incrementality, researchers have asked participants to describe images of complex events with multiple components (agent, patient, verb, theme, etc.) in order to investigate what determines the size of these message-to-utterance chunks and the order in which they are produced (Konopka & Brown-Schmidt, 2014).

The size of the chunks, or the *planning scope*, appears to be under some strategic control, with chunks as small as a single word (Griffin, 2001; Zhao & Yang, 2016) and as large as an entire phrase or clause (Martin, Crowther, Knight, Tamborello, & Yang, 2010). The planning scope may also be hierarchically organized, i.e., not strictly based on the linear order of words in the utterance but also on relations between components such as who did what to whom (Antón-Méndez, 2020; Lee, Brown-Schmidt, & Watson, 2013). In fact, the high level of flexibility in planning scope suggests that the relevant questions are not about the size of the most basic planning unit, but rather what determines the planning unit in that given context (Brown-Schmidt & Konopka, 2015; Konopka, 2012). The "context" can include factors such as time pressures (Ferreira & Swets, 2002), message complexity (Smith & Wheeldon, 1999), properties of the particular language spoken (Jaeger & Norcliffe, 2009), and other task demands.

The second question about incrementality regards the starting point of the utterance; which words or phrases are planned and produced first (MacWhinney, 1977). The starting point might be determined in a bottom-up manner, i.e., based on perceptual or conceptual prominence of message components. For example, when a visual cue directs participants' attention to one of the characters in a depicted event, that character is more likely to be mentioned first (Gleitman et al., 2007; Myachykov & Tomlin, 2008). Similar effects have been reported for other attributes such as animacy (Tanaka, Branigan, McLean, & Pickering, 2011), lexical frequency (Fenk-Oczlon, 1989), or lexical accessibility (Bock, 1986). Alternatively, speakers might select the starting point in a top-down manner: using their higher-level message representation to guide attention to particular components that are useful starting points for utterance formulation (e.g., the agent character). Indeed, when participants describe visual scenes, fixations in the initial phase (0–400 ms) do not show preference for a certain character in the depicted event (Griffin & Bock, 2000), and only later participants begin fixating on the character they will mention first. This suggests that often speakers first encode the gist of the event, i.e., a rudimentary representation of the relationship between characters in the event, and only then choose the starting point for linguistic encoding (Konopka & Brown-Schmidt, 2014).

As with the scope of planning, the factors that determine starting points also seem to be context dependent, with a mixture between top-down and bottom-up strategies (Konopka & Brown-Schmidt, 2014). There is some evidence that low-level attentional cues influence the starting point when the higher-level message plan is not easily available (Kuchinsky & Bock, 2010), such as when the depicted event is not easily codable (e.g., when the action taking place is ambiguous or can be described by several different verbs). This might suggest that top-down message-driven planning is the preferred strategy, perhaps because planning ahead prevents disfluencies and allows the most efficient mapping between message and sentence structure. However, bottom-up attentional cues can support utterance planning when the message information is not easily available, allowing production to begin despite difficulty encoding the message plan. This might result in dispreferred or more demanding sentence forms (e.g., the passive form in English) and might even cause disfluencies or repairs (Brown-Schmidt & Konopka, 2008), but ultimately this strategy provides the utterance formulator with additional flexibility when the message input is difficult to process.

If the choice of planning scope and starting points is highly context-dependent in cases of complex messages, it is likely that similar flexibility would be found for situations of message uncertainty: message uncertainty is another type of production context, with its particular demands that could both affect incremental planning and benefit from it. For example, imagine you see a dog in the park chasing some smaller animal that you cannot identify from afar. Perhaps you will start describing the scene as "the dog is chasing…". Then, while producing those words you realize it is a cat being chased, so you can complete your sentence fluently with "the cat." Alternatively, perhaps you identify the cat first, clearly running away from some animal that is still behind the bushes. Then you might say "the cat is being chased…," and while producing those words you identify the missing component and complete your sentence with "by the dog." That is, despite not having the complete event information, you can begin planning and producing at least part of the utterance. Moreover, the information you *do* have available—the dog agent or the cat patient—could determine which word you produce first in your sentence, and as a consequence, whether you produce a passive or active sentence structure.

In sum, the flexibility of incremental planning suggests that a speaker's planning scope and starting points could be modulated by several factors including the amount of message information available prior to speech, the particular time pressure posed on the speaker, and which particular message information is already available. Incremental planning could allow speakers to begin speaking even before the message is fully settled, and this is likely to affect their utterance forms and time course of production, suggesting an important role for message uncertainty on utterance form decisions.

3. Message uncertainty in real-life contexts

3.1 Uncertainty in production models

The evidence reviewed so far suggests that message uncertainty likely carries implications for utterance planning, whether in the time course of production, the strategies used for planning, or the utterance form itself. However, situations of message uncertainty are treated rather anecdotally in models of language production. Challenges with the message information are mostly discussed at the discourse level and do not permeate into later formulation stages. For example, Levelt (1989) notes that in certain discourse types there is a complex message that requires careful ordering of smaller components; e.g., when a speaker wants to build up a convincing argument. However, the challenges described by Levelt are more about how to organize the message in the best way for the listener to comprehend, and not about uncertainty around the message content, or what challenges an incomplete message might present for the speaker during utterance formulation. Similarly, Bock (1995) mentions uncertainty as a cause of disfluencies and jabberwocky, but suggests that "message uncertainty is more akin to a thinking problem than a talking problem" (p. 183).

Garrett (1989), on the other hand, explicitly acknowledges a certain type of message uncertainty: when the speaker has multiple potential message options but must decide on one to be processed further in the formulator. The notion of multiple representations competing for activation is well established in language production research; with evidence for competition at the lexical (Abdel Rahman & Aristei, 2010; Abdel Rahman & Melinger, 2007), phonological (Cohen-Goldberg, 2012; Sevald & Dell, 1994), and structural (Myachykov, Scheepers, Garrod, Thompson, & Fedorova, 2013) levels. As Garrett notes, it is intuitively plausible to have similar parallelism in messages: if we can assume that a person has more than one train of thought at a time, that opens the possibility that two message representations exist in parallel. But although Garrett acknowledges the option of entertaining multiple messages in parallel, the competing message he discusses is considered an "intruding" message—typically something perceptual that the speaker hears or sees and intrudes the process of formulating the intended message; not uncertainty around the message intended for expression.

One possible reason message uncertainty is not explicitly addressed in previous accounts is that it is viewed as a very particular context of production, while models attempt to provide a simplified overview of processing under standard circumstances. However, message uncertainty might in fact be more common than assumed, and it is arguably difficult to decide what "standard" production circumstances are. Thus the lack of experimental work on message uncertainty is both a cause and an outcome of the scarce treatment of message uncertainty in production models.

To gain a better understanding of types of uncertainty and the planning strategies used to overcome them, the next section will review message uncertainty in context: first by presenting two examples of natural contexts likely to induce uncertainty—conversational turn-taking and live narration; and then by examining error and disfluency patterns that reflect message uncertainty. While not at all comprehensive, these contexts could be used as a starting point for motivating experimental work and incorporating message uncertainty into theories of production.

3.2 Natural contexts of message uncertainty

3.2.1 Turn-taking in conversation

Conversational turn-taking involves rapid exchanges of information between interlocutors, with each turn lasting on average two seconds (but durations are highly variable; Levinson, 2016). Because the interlocutors respond to each other, the content of speaker A's turn will depend on what speaker B said in the prior turn. This means that speakers cannot pre-plan their messages in advance (as might happen when delivering a planned speech), but rather must listen to the interlocuter's turn and rapidly prepare a response that corresponds to it.

Interestingly, the modal gap between turns is only about 200 ms, with little variation cross-linguistically (Stivers et al., 2009). This duration is extremely short given that planning a single word takes about 600 ms when primed (Indefrey & Levelt, 2004) or 1000 ms when not (Bates et al., 2003), while planning a simple event-description sentence takes around 1500 ms (Griffin & Bock, 2000). It is therefore unclear how the 200 ms gap is enough for speaker A to comprehend what speaker B said, think of a response message, process the message for utterance formulation, and launch the response in time.

One common explanation relies on predictive comprehension (Kuperberg & Jaeger, 2016), suggesting that comprehenders can often predict upcoming words or messages based on linguistic and context cues in the conversation. If so, speaker A can predict with some confidence how speaker B will end their utterance and/or what the message is. Speaker A can therefore begin planning the ensuing response even before speaker B has finished their turn (Barthel, Sauppe, Levinson, & Meyer, 2016; Bögels, 2020; Corps, Gambi, & Pickering, 2018; Levinson, 2016; Levinson & Torreira, 2015). For example, Bögels, Magyari, and Levinson (2015) had participants answer general-knowledge questions while their speech and electroencephalography (EEG) responses were recorded. Each question was pre-recorded in one of two conditions: (1) Early; where the key information for answering the question was provided mid-question, e.g., *which character, also called 007, appears in the famous movies?* (2) Late; where the key information only appeared at the end of the question, e.g., *which character from the famous movies is also called 007?*

Results showed that participants were faster to respond in the Early condition compared to the Late condition. Moreover, EEG analyses showed a positive-going wave approximately 500 ms after the onset of the key information in the question ("007"). This positivity, localized to areas which have previously been associated with speech planning, was significantly larger than in a control experiment where participants only listened to the questions but did not respond. Bögels et al. (2015) concluded that participants began planning their response as early as 500 ms after the key information was presented, i.e., as soon as the question (and answer) became predictable—which was already mid-question for the Early condition.

Notably, predictive comprehension is another case of incrementality in language processing: listeners begin creating a representation of the incoming message as soon as possible, and continue building it up as more information becomes available (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003). When this partial representation is constraining enough, listeners can predict how the sentence will unfold with some confidence, allowing them to plan their own response (Levinson, 2016).

Importantly, however, predictions are necessarily uncertain. The listener's accuracy and confidence in their prediction might depend on various factors including the degree of constraint in the sentence, the perceptual context, the discourse context, or even the speaker's familiarity with their interlocutor. Predictions can also vary in their degree of specificity, from the more abstract higher-level message, down to the particular utterance phrasing (Kuperberg & Jaeger, 2016). But if a speaker is still uncertain about what they are responding to, their response message must also be temporarily uncertain.

Because prior turn-taking research focused on stimuli with high message predictability (manipulating only the timing of when the message was revealed), this still leaves open the question of whether—or how—interlocutors plan under message uncertainty. For example, the degree of uncertainty might determine whether speakers plan their response prior to speaking, or prefer alternative strategies to gain processing time (e.g., beginning their turn with filler words such as *um* or *uh*). The degree of uncertainty might also determine whether speakers commit to a plan but are prepared to modify it, or perhaps even maintain multiple rudimentary plans until there is enough information to select one. The dynamics of turn-taking and uncertainty – in both the incoming speech and as a consequence, the message of the response – present a complex context for production, leaving several other options and strategies to be explored in future research. Findings from these studies could have implications at the intersection of turn-taking, incremental planning, and predictive comprehension, while addressing a common everyday context of language production.

3.2.2 Live narration

Narration of live events also poses particular production challenges: the narrator must attend to the ongoing events, interpret what is occurring, transform that into speech, and produce the utterance rapidly enough to keep up with the upcoming events. Rather than uncertainty dependent on language comprehension (as in turn-taking), the uncertainty in narration is dependent on perception of events. In some cases the events might be highly predictable, allowing the narrator to plan their utterance even before the event is completed (e.g., narration of a scripted play), and the narrator only needs to align the utterance with the timing of the event. In other cases, the events might be ambiguous or much less predictable, requiring the speaker to rapidly narrate ongoing events despite some message uncertainty.

Live narration has mostly been studied is the context of sports commentaries. Although different sports have different properties in terms of speed or complexity of plays and scoring, several share the need for real-time rapid narration, in addition to particular requirements of the medium (e.g., TV, radio). However, prior research on linguistic aspects of live commentary is rather scarce, and has focused mostly on register characteristics, audience design (Desmarais & Bruce, 2009, 2010), or turn-taking conventions between commentators (Bowcher, 2003). The cognitive challenges faced by the commentator and how they might be resolved have received much less attention.

However, Aleksander Popov (2019) analyzed utterances from commentators of various sports (cricket, soccer, horse racing, and tennis) and reported a number of effects of message uncertainty on utterance forms. For example, passive sentences are very frequent in televised horse racing commentaries (e.g., *Seabiscuit followed by Kayak II*…). Popov explains that there is often temporary uncertainty around the identity of the horses, and the commentator needs extra time to recognize them (based on color, jersey, headgear, etc.). Using the passive form provides the commentator with a longer lag between naming the two horses compared to what the active form would allow. While the commentator is producing "followed by," they have more time to identify the next horse—taking advantage of incremental planning so that more information can be gathered. Indeed, prior accounts have suggested that the passive is a practical tool for commentators in ball games too: because the action can typically be identified before the player, using the "by" passive allows the commentator to begin their utterance about the action while they continue to identify the player (Balzer-Siber, 2015; Hoyle, 1991). In a cross-game comparison, Popov finds that the use of passives is more frequent in soccer (football) compared to cricket or tennis. Popov argues that this is because ball possession changes rapidly during soccer and more time is needed to identify the player, making the passive a useful form choice for commentators under uncertainty.

In another study by Wanta and Leggett (1988), sports announcers were found to use more clichés when games developed in unexpected ways. The authors note that commentators work under continuous time pressure, and need to report in real time about events that range from fairly expected to completely unexpected. Wanta and Leggett suggested that in the unexpected cases more attention must be directed to processing the game information, and less attention will be available for language production. This might lead the commentators to resort to clichés, which are highly practiced and easily recalled from memory without needing much utterance planning.

Notably, clichés are considered a dispreferred stylistic form that commentators attempt to avoid (Wanta & Leggett, 1988). Similarly, in English the passive form is less frequent and more difficult to process than the active form (Paolazzi, Grillo, & Santi, 2021). Although the evidence is limited, these examples from sports commentators show how message uncertainty can affect utterance forms: producers choose utterance forms that mitigate difficulties associated with the long time course of determining the message, and the producer's needs sometimes even override stylistic or audience design choices.

3.3 Speech patterns of uncertainty

3.3.1 Disfluencies

Another way to identify contexts of message uncertainty is by examining disfluencies and delays in production, which reflect difficulty in planning speech. Difficulties associated with message planning might show a different pattern of disfluencies compared to other difficulties a speaker might encounter. For example, filled pauses (e.g., *um*, *uh*) typically occur at phrase boundaries, where new messages are likely being planned for the next phrase (Bock & Cutting, 1992), while silent pauses are more common within phrases (Maclay & Osgood, 1959). Filled pauses are also more frequent when speakers describe more ambiguous scenes in the Thematic Apperception Test (Siegman & Pope, 1966), perhaps suggesting an association between filled pauses and message planning difficulties.

The exact role of filled pauses (and other disfluencies) in the production process is still unclear, however. One suggestion is that filled pauses are used by speakers to signal that they are not done with their turn yet, and would like to continue holding the ground until their next utterance is ready (Clark & Fox Tree, 2002; Maclay & Osgood, 1959). Interestingly, in a natural environment of university lectures, Schachter, Christenfeld, Ravina, and Bilous (1991) found that the incidence of filled pauses depended on the academic discipline: the more formal and factual the discipline, the fewer filled pauses. This finding is particularly interesting given that there is little chance of interlocutor interference during lectures. Schacter et al. suggested that more factual disciplines constrain the options for message production, and therefore fewer filled pauses are needed. That is, filled pauses might be used when the speaker is having difficulty choosing a message, and could be a marker of message uncertainty.

Similarly, Fraundorf and Watson (2013) hypothesized that fillers (filled pauses) are more common when speakers are engaged in message-level planning, whereas other disfluencies (repeats or silent pauses) are more likely when there is difficulty at the grammatical or phonological levels. To test their hypothesis, Fraundorf and Watson used a story-telling paradigm where participants read passages and retold them in their own words. Results showed that fillers were most likely before articulation of an utterance began (rather than mid-utterance), and in particular at key plot points where participants had to plan a new message component. Silent pauses were also more likely before articulation began, but were less affected by the key plot points than fillers were. Moreover, fillers were not sensitive to several other factors related to grammatical, lexical, and phonological planning (e.g., lexical frequency), but silent pauses were.

Fraundorf and Watson suggested that fillers indicate that the speaker has not yet committed to a new message plan. Moreover, they concluded that their findings support Clark and Fox Tree's (2002) account that speakers use fillers to communicate to their listeners that their utterance planning is being delayed. Under this view, the fillers themselves carry a communicative intention (a message) for the listener. Because fillers require a message-level plan, they are most common when speakers are already engaged in message planning, rather than during articulation when the message is presumably set already.

Together, these findings suggest that disfluencies might be a useful cue for exploring message uncertainty. First, tracking the incidence and distribution of fillers can help identify points of message uncertainty in speech and how common they are. Second, fillers could be investigated as a production strategy that speakers use when faced with uncertainty, allowing them to buy more processing time. Moreover, the type of filler might signal the type of uncertainty – e.g., whether the speaker is debating between several self-generated messages or still retrieving knowledge information to answer a question. Investigating how disfluencies vary with message uncertainty could be informative of which situations tend to cause uncertainty, and the strategies used to mitigate the difficulty.

3.3.2 Errors of message uncertainty

Another way to identify contexts of message uncertainty is by examining speech errors that may derive from an incomplete message plan. For example, Harley (1984) classifies the following as a "high-level intrusion error," occurring at the message level:

(1)
Target Utterance: I want to cut out the elephant on the back of that.
Actual Utterance: I want to cook out the elephant on the back of that.
Relevant Context: the speaker was in the kitchen cooking with some other people. He wanted to make conversation but was unsure whether to talk about cooking or about a picture of an elephant on the back of a box in the kitchen.
(Harley, 1984, p. 200)

In this example, it appears that the speaker's intended message was being processed for formulation when a single word from an alternative message option (the topic of cooking) intruded. This suggests that components of an alternative message might be processed for formulation alongside the intended message, particularly when a speaker is initially uncertain which of the two messages to choose. This parallel processing can result in an intrusion at the output, perhaps reflecting a failure in inhibition. In another type of error, called *blend errors*, the alternative messages become entirely blended into a single utterance:

(2)
The sky is blue.
The sun is shining.
Actual Utterance: The sky is shining.
(Harley, 1984, p. 203)

Blend errors have been extensively studied at the phonological, lexical, and syntactic levels (Coppock, 2010; Dell & Reich, 1981; MacKay, 1972), and appear to result from unresolved competition between multiple options for production. Blends at the message level are rarely discussed, but might similarly reflect competition between intended messages (Harley, 1984): when the speaker is debating between multiple message plans, these messages could begin processing in parallel. If the speaker is late to select a message, the parallel processing can proceed all the way down to articulation, resulting in a blended output.

The semantic and phonological similarity effects often found in message-level errors (e.g., *cut* and *cook* in example (1)) suggest that message planning interacts with later formulation stages, such that high-level processes are sometimes affected by low-level factors (Bock, 1996). For example, phonological or lexical information might be accessed even before message planning is complete, and in turn can affect message planning and utterance formulation. Thus the processing stages of language production are highly interactive, with lower-level processes interacting with higher-level message planning. As Harley argues, errors that originate at the message level prove that language production and errors cannot be studied without considering message planning.

Blend and intrusion errors suggest that competition between multiple potential messages is one type of message uncertainty that speakers face. But several questions remain about these error patterns and their implications for production models. First, what predicts the type of error—between a single word intrusion and a complete blend? For example, the answer might depend on how strong of a competitor the alternative message is, or on how far the competing messages reached in processing before a single message was selected. More generally, what other types of message errors can be identified and how might they reflect the competition between alternatives?

Second, message uncertainty is clearly challenging for production planning, but the production system is known to be flexible and adaptive. When multiple messages are being considered, what kind of strategies are used to select messages for production, inhibit unintended messages, or maintain fluent and rapid production without error? Interestingly, Harley notes that because the order of processing stages is not always fixed but rather depends on the context, important components of the message might be prioritized for utterance planning. These prioritized components might even reach phonological encoding before other processing stages that are typically considered earlier, such as syntactic choice, have occurred. Moreover, if there is some overlap between the competing messages, that overlapping content might be prioritized and planned first (Gussow & MacDonald, 2021), even allowing some delay in selecting the message. Prioritization of more certain components might be one strategy for the language system to deal with uncertainty due to competing messages, but could also lead to error if another message is ultimately selected.

Taken together, the natural contexts reviewed here show examples of where and how message uncertainty might pan out during language production. Sometimes situations of uncertainty can be studied by examining contexts that are likely to induce uncertainty, while at other times it is only based on errors or disfluencies that we can identify the speaker's uncertainty. Examining these contexts can help understand (a) what types of message uncertainty exist, such as choosing between multiple messages or waiting for more event information to unfold; (b) which production strategies are used to mitigate the difficulties of uncertainty, and (c) what effects uncertainty has on utterance forms.

4. Goal uncertainty in action plans

In this final section, I will turn to examine uncertainty in a different cognitive domain: motor action planning. In contrast to the paucity of research on uncertainty in the psycholinguistics literature, uncertainty research in the action domain is quite abundant. But planning parallels between the two domains have been noted before (Anderson & Dell, 2018; Koranda, Bulgarelli, Weiss, & MacDonald, 2020; Lebkuecher et al., 2022; MacDonald, 2016; Rosenbaum, Weber, Hazelett, & Hindorff, 1986), and language production is in fact a type of action—suggesting that research in these areas can be mutually informative.

As in the psycholinguistics literature, motor action researchers view action choices as a series of decision-making: whether about the chosen action goal, the form of movement, or when to initiate the motor plan (Wolpert & Landy, 2012). In the action domain, the goal of the action is the intent, the reason for action—akin to the message in language production, while the motor plan specifies the chosen movements and motor commands to achieve that goal—akin to the utterance plan in language production. The next section will review some of the main questions regarding goal uncertainty in action and its effects on motor plans, while pointing to language parallels that could be similarly investigated.

4.1 Simultaneous perception and action

One main cause of uncertainty in action planning is time constraints. Oftentimes an action needs to be initiated quickly, even before the actor has gathered the full perceptual information needed to complete it. For example, catching a ball in flight must occur before the ball hits the ground and/or passes out of the player's range. Thus a ball player cannot wait until the ball is within their reach in order to plan their catch; they must watch the ball in flight, perceive enough information to predict the ball's trajectory, decide on the optimal time or place to catch it, plan the action, and launch it in time. Time constraints force the ball player to perceive and act simultaneously, and many natural actions contexts require similar overlap between perception and action (Faisal & Wolpert, 2009).

The incremental nature of visual perception suggests a trade-off between the time and processing resources for perception versus action: The more time can be allotted to perception, the more uncertainty can be reduced and the more accurate the action will be (Faisal & Wolpert, 2009). However, spending more time on perception leaves less time for action planning and execution, which increases the risk of missing the opportunity for action (missing the ball) or making an action error. The actor must therefore decide how much uncertainty they are willing to tolerate, or how much information is sufficient for initiating action.

Prior work suggests that people show near optimal performance in motor tasks that require a trade-off between perception and action uncertainties (Battaglia & Schrater, 2007; Faisal & Wolpert, 2009). Participants integrate various sources of information in their decision making—not only from the perceptual environment, but also from prior experience and general knowledge—resulting in statistically optimal decisions about which action to perform and when precisely to execute it. This near-optimal performance has been found for tasks that varied from being rather naturalistic, like virtual reality ball-catching (Faisal & Wolpert, 2009), to tasks that were completely novel, like reaching for invisible targets (Battaglia & Schrater, 2007).

For example, Battaglia and Schrater (2007) had participants move their finger on a haptic workspace from a start button to an invisible target location. Participants were to estimate the invisible target location using dots scattered around it. Dot positions were sampled from a distribution with a mean at the invisible target position and a standard deviation that varied across conditions. The number of dots increased as time elapsed, but once the participant initiated a movement, no further dots appeared. A countdown sand-timer provided only 1200 ms for the trial, introducing a perception-action trade off: waiting for more dots would reduce uncertainty about the precise target location, but would leave less time for action planning and precise execution.

Battaglia and Schrater compared participants' performance to a computed "ideal reacher" who initiates movement at a time that minimizes endpoint deviations from the target location. Results showed that participants' performance was near-optimal, despite not getting any direct feedback about ideal performance. Interestingly, in a ball-catching experiment, Oudejans, Michaels, and Bakker (1997) found that non-experts were faster to initiate movement than experts, but at the cost of accuracy. This might suggest that experts are better at finding the optimal switch point, or that their expertise in planning movements allows them to allot more time to perception without degrading movement accuracy.

There are numerous parallels between the perception-action trade-off required for motor actions and that required for language production under uncertainty. Just like in motor action, speaking is often subject to time pressures, and the speaker must decide how much message information is enough to begin planning or producing their utterance. But a speaker's message often depends on incrementally incoming information, such as comprehension of an interlocutor's utterance or interpretation of an unfolding visual scene.

Prior work on language comprehension has shown that comprehenders integrate and weight cues from various sources of information in order to comprehend incrementally word-by-word, instead of waiting until the full utterance is completed (Altmann & Kamide, 1999; Kamide et al., 2003; MacDonald, 1994; McClelland, Mirman, Bolger, & Khaitan, 2014). This allows for faster and more efficient comprehension, and therefore the language system is willing to risk some error in initial interpretations (Frazier & Rayner, 1982) or predictions of upcoming speech (Clark., 2013). But despite the large body of work on the efficiency and uncertainty of incremental comprehension, the implications for production are not often discussed. This is somewhat surprising given that comprehension often occurs in the context of a conversation that requires a verbal response (or a motor action response)—similar to perception for the purpose of action.

Notably, the turn-taking literature does discuss the prediction-production relationship: as discussed earlier, speaker A can sometimes predict what speaker B's message is going to be, and therefore plan a response even before speaker B has finished speaking (Corps et al., 2018; Levinson, 2016; Levinson & Torreira, 2015). However, these studies did not focus on questions of the degree of uncertainty or the trade-off between predictions and production; i.e., questions about how the strength of the prediction might affect the timing or form of utterances.

Quantifications of the degree of sentence predictability do exist, however, including cloze probabilities (Taylor, 1953) or sentence constraint (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Schwanenflugel & Shoben, 1985; Staub, Grant, Astheimer, & Cohen, 2015). A natural next step might be to relate predictability measures and utterance planning—asking how much certainty in comprehension the producer needs in order to begin production planning, or how the degree of uncertainty affects their utterance forms. This likely varies between producers and between contexts, as it does in motor action where the optimal switch point is highly variable (Faisal & Wolpert, 2009). In fact, coders have been found to disagree about when exactly in an incoming question the listener can identify the answer and begin planning a response (Bögels, 2020). In addition, other communicative cues such as facial expressions, gestures, or interruptions might also be used to estimate when listeners are confident enough to begin their own message planning. Producers might even prefer certain communicative cues depending on the level of uncertainty—e.g., a facial expression might be more ambiguous than a linguistic utterance, making it a less risky response in situations of uncertainty.

Although perception and action are more temporally separate in language than they are in motor action—given the limited ability to speak and listen at the same time, and social conventions of turn-taking—at the very least there is overlap between language comprehension and production *planning* (Bögels et al., 2015; Levinson, 2016), and producers have to balance between them effectively, just as in motor action. Moreover, the cognitive toll and interference between perceiving and planning simultaneously has been discussed in both the language (Jongman & Meyer, 2017) and the motor domains (Liu, Chua, & Enns, 2008). Both literatures also have theories of shared versus separate systems, for perception and action (Creem-Regehr & Kunz, 2010) or comprehension and production (Pickering & Garrod, 2013). Thus integrating research in these domains could provide additional insight into production and action under uncertainty, in particular for the language domain where the research is relatively scarce.

4.2 Intermediate movements

A major debate in the motor uncertainty literature regards the underlying cause of *intermediate*, or *averaged* movements: when faced with two competing goals, movement toward the target goal often shows properties of the movement that would be required for the competing goal. For example, movement trajectories might be initially directed in between two opposite goal locations (Chapman et al., 2010) and hand orientation might be intermediate between pronation and supination (Gallivan, Barton, Chapman, Wolpert, & Randall Flanagan, 2015). A commonly used paradigm to investigate these movements is the "go-before-you-know" paradigm: participants are presented with multiple potential reach targets, and the goal target is only revealed *after* participants initiate their movement. Thus movement towards the targets necessarily begins when there is still goal uncertainty.

One interpretation of intermediate movements is that they represent an average of the competing movement plans: when participants are pressured to begin an action immediately, multiple movement plans are computed in parallel and the resulting movement represents their average (Chapman et al., 2010; Stewart, Baugh, Gallivan, & Flanagan, 2013). The goal action can often be completed successfully because the target is disambiguated mid-trial and participants adjust their movement, but the movement still deviates from the most direct route to the target. This deviation can even be viewed as error (Hening, Favilla, & Ghez, 1988), the result of a planning system taxed by multiple potential target options and time pressures.

However, a more refined account is that intermediate movements reflect a *co-optimized* motor plan (Haith, Huberdeau, & Krakauer, 2015; Wong & Haith, 2017). Under this account, the motor system computes a single action plan that is most optimal for later movement corrections—considering the various potential targets, motor costs, efficiency, timing, and other task demands. Then, once the target is disambiguated, the movement can be adjusted online to reach the goal. The argument is not that people engage in explicit strategizing per se, but rather that an implicit property of the motor system is to plan optimally under uncertainty.

There is still much debate about whether intermediate actions reflect the competition of multiple parallel motor plans or a single optimized plan (Alhussein & Smith, 2021; Enachescu, Schrater, Schaal, & Christopoulos, 2021; Gallivan, Chapman, Wolpert, & Flanagan, 2018; Wong & Haith, 2017), though the latter seems to be better supported (Alhussein & Smith, 2021). These are not necessarily mutually exclusive, however. Given a highly flexible motor planning system, the strategy for action under goal uncertainty might depend on the particular action context or paradigm. This includes which particular aspects of the action are uncertain (e.g., the spatial location of the target versus the required grasp), whether the target is disambiguated before or after the movement begins, and even what individual differences exist in performance strategies (Wong & Haith, 2017).

In the language domain, the feasibility of maintaining multiple utterance plans at once is unclear and likely depends on the particular stage of planning. It might be possible to maintain multiple messages (Garrett, 1989), but maintaining multiple phonological plans for multiword utterances would likely be too taxing on memory and very error-prone (Dell, Burger, & Svec, 1997; Wilshire, 1999)—perhaps as reflected in blend errors discussed earlier (Harley, 1984). However, some version of plan optimization in the face of uncertainty seems plausible. For example, speakers might choose to produce more certain components of their utterance first, allowing them to begin production sooner while also buying time to gather information about uncertain components of the utterance (Gussow & MacDonald, 2021). Indeed, some motor researchers have also suggested that intermediate movements are used to buy time until uncertainty can be reduced (Enachescu et al., 2021). Production strategies might also depend on the source of uncertainty, with differences between imposed message uncertainty (e.g., describing an unfolding scene) vs free deliberation between message options. This has already been attested in the motor domain, where movement variability is higher when participants have a free choice between targets compared to a predetermined target (Krüger & Hermsdörfer, 2019). These findings suggest that the particular source of uncertainty affects movement strategy and variability, and it seems likely that similar findings would emerge in language production.

4.3 Neural correlates of goal uncertainty in motor planning

Because motor actions can be studied on non-human primates, the neural correlates of action planning are better understood than those of language planning. This is particularly true for situations of goal uncertainty, which have been extensively studied using paradigms that present monkeys with multiple potential targets and then cue one target that the monkeys need to reach for. In contrast to the 'go-before-you-know' paradigm, in this case decision making and motor preparation precede the actual movement. By using single-cell recordings, researchers can glean rather specific information about neuronal activity that is tuned to particular target locations.

Results suggest that multiple potential reach targets are represented simultaneously in the brains of macaque monkeys, and activity in the dorsal premotor cortex is modulated by the locations of potential targets (Cisek & Kalaska, 2005; Pastor-Bernier & Cisek, 2011). Once the goal target is disambiguated, its associated neural signal increases while the signal of the competing target decreases (Cisek & Kalaska, 2005). In fact, neural representations are very dynamic throughout the decision making process, changing based on both the degree of uncertainty and the approaching response time (Bastian, Schöner, & Riehle, 2003). The research also suggests that motor decision making occurs within the same neural substrates that execute the action; that is, it does not necessarily implicate separate processes for decision making versus implementation (Cisek, 2006; Pastor-Bernier & Cisek, 2011).

In human participants, magnetoencephalography (MEG) studies show modulation of oscillatory activity during motor decision making, mainly implicating the beta band in planning under uncertainty. Specifically, decreases in beta-band power are observed in preparation for movement, but this effect is attenuated in situations of uncertainty: Modulation depends on the number of potential targets (Tzagarakis, Ince, Leuthold, & Pellizzer, 2010; Tzagarakis, West, & Pellizzer, 2015), their proximity to each other (Grent, Oostenveld, Medendorp, & Praamstra, 2015), and even on hand choice for executing the action (van Helvert, Oostwoud Wijdenes, Geerligs, & Medendorp, 2021). Together, these findings suggest that some movement preparation begins even before the target is disambiguated, and takes into account properties of the various potential targets. Moreover, neural correlates of action planning reflect key components of uncertainty, including the degree of uncertainty for each target (e.g., depending on the number of potential targets) and the degree of similarity between target options (e.g., spatial proximity).

Although the evidence from language production research is limited, perhaps some insight into early planning under message uncertainty can be gained from studies where participants produce a word after a semantically-constraining context (Blackford, Holcomb, Grainger, & Kuperberg, 2012; de Zubicaray, McMahon, Eastburn, & Pringle, 2006; Piai, Klaus, & Rossetto, 2020; Piai, Roelofs, Jensen, Schoffelen, & Bonnefond, 2014; Piai, Rommers, & Knight, 2018). For example, Piai et al. (2020) had participants name a target picture (e.g., *cow*) to complete a sentence that appeared word-by-word on screen. Results showed that neural oscillations in the alpha-beta band decreased in power when the sentence context was semantically constraining (e.g., *the farmer milked the* __) compared to non-constraining (e.g., *the child drew a* __). This reduction was interpreted as an index of lexical-semantic retrieval, with the constraining context allowing for some early preparation of the likely target—even though the target was still uncertain. Moreover, right before viewing the target picture, participants heard a distractor word that was either semantically related (e.g., *goat*) or unrelated (e.g., *bean*) to the target. When the distractor was semantically related to the target, the alpha-beta power reduction began later than when the distractor was unrelated. Piai et al. (2020) speculated that semantic competition between the expected target and the related distractor caused this delay, as the competition interfered with word retrieval processes. Notably, naming latencies were not affected by the type of distractor, so the effects of this semantic competition were only evident in the neural oscillations.

Because Piai et al.'s (2020) focus was on lexical-semantic processes and not message uncertainty, key components of message uncertainty were not systematically manipulated in their study—including the degree of sentence constraint as a continuous measure, target similarity, or the particular message component under uncertainty (e.g., the thematic role). Moreover, the imposed distractors indeed presented more of a "distraction," or intrusion, rather than deliberation between message options; and participants only named a single word rather than producing the entire message or sentence themselves. But semantic competition appears highly intertwined with message competition in this case, and similar paradigms might be used for a more systematic study of production under message uncertainty and its neural correlates. Stimuli norming in Piai et al. (2020) even included measures of both cloze probability and semantic similarity between targets and distractors; two measures that could be used for testing questions of message uncertainty in future work—again similar to the motor uncertainty work, where neural activity is sensitive to the number of potential targets (resembling the degree of sentence constraint) and their proximity to each other (a measure of similarity).

Admittedly, the research on neural correlates of message uncertainty will be necessarily exploratory at first and more complex than in the motor work, given the richness of messages and semantic information. But investigations can begin with simple paradigms, focusing on neural correlates already associated with relevant aspects of production planning: whether particular oscillation patterns such as alpha-beta decreases associated with lexical-semantic retrieval (Piai et al., 2014, 2020), ERP markers of early planning (Bögels, 2020; Bögels et al., 2015), or even spatially, within brain regions and networks associated with language production (Friederici, 2011; Indefrey, 2011). The initial overarching goal would be to examine how typical neural markers of language production are modulated by the type or degree of message uncertainty, and at a later stage, even by the speaker's production strategy and resulting speech patterns.

5. Conclusions

This chapter introduced *production under message uncertainty*, suggesting that sometimes a speaker begins utterance planning before they are certain of the message content they want to communicate. Given the relative lack of research on message uncertainty in production, the goal was to gather evidence from several neighboring areas in order to describe incidences and consequences of message uncertainty and to motivate future research. Conclusions from this review make clear that topics of message uncertainty could and should be incorporated into language production research. First, message uncertainty might be more common than assumed, and an initial step would be to identify the incidence and types of message uncertainty—e.g., whether a speaker is debating between several messages options, waiting for an event to unfold, or still retrieving knowledge to answer a question. It might be possible to identify situations of message uncertainty by examining contexts likely to induce uncertainty, or by tracking speech errors and disfluencies that reflect uncertainty. Second, situations of message uncertainty carry implications for utterance forms and the time course of production.

Existing models of language production typically assume that the message is settled before utterance planning begins, but this assumption obscures how real-time message formulation can affect utterance planning and the resulting utterance – given a highly flexible production system that uses various strategies to mitigate difficulties of planning under uncertainty. Investigating situations of message uncertainty would therefore not only address a common everyday context of production, but could also inform theories of language production more generally. Finally, uncertainty is ubiquitous in other cognitive domains, and in particular goal uncertainty in the motor domain shows many parallels with message uncertainty in language. These parallel lines of research can therefore inform and benefit each other, contributing to our understanding of which domain-general cognitive principles are used for planning and acting under goal uncertainty.

Acknowledgments

This review was initially written as part of my preliminary exams for PhD candidacy. I would like to thank Maryellen MacDonald for continuous guidance and helpful discussion throughout the development of these ideas. Preparation of the article was supported by the Graduate School, part of the Office of Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison, with funding from the Wisconsin Alumni Research Foundation and the UW-Madison.

References

Abdel Rahman, R., & Aristei, S. (2010). Now you see it … and now again: Semantic interference reflects lexical competition in speech production with and without articulation. *Psychonomic Bulletin & Review*, *17*(5), 657–661. https://doi.org/10.3758/PBR.17.5.657.

Abdel Rahman, R., & Melinger, A. (2007). When bees hamper the production of honey: Lexical interference from associates in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 604.

Alhussein, L., & Smith, M. A. (2021). Motor planning under uncertainty. *ELife*, *10*, e67019. https://doi.org/10.7554/eLife.67019.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1.

Anderson, N. D., & Dell, G. S. (2018). The role of consolidation in learning context-dependent phonotactic patterns in speech and digital sequence production. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(14), 3617–3622. https://doi.org/10.1073/pnas.1721107115.

Antón-Méndez, I. (2020). The role of verbs in sentence production. *Frontiers in Psychology*, *11*, 189. https://doi.org/10.3389/fpsyg.2020.00189.

Balzer-Siber, M. (2015). *Functional and stylistic features of sports announcer talk: A discourse analysis of the register of major league soccer television broadcasts [Doctoral dissertation]*. East Tennessee State University.

Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. *Frontiers in Psychology*, 7. https://doi.org/10.3389/fpsyg.2016.01858.
Bastian, A., Schöner, G., & Riehle, A. (2003). Preshaping and continuous evolution of motor cortical representations during movement preparation. *European Journal of Neuroscience*, *18*(7), 2047–2058. https://doi.org/10.1046/j.1460-9568.2003.02906.x.
Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., et al. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344–380. https://doi.org/10.3758/BF03196494.
Battaglia, P. W., & Schrater, P. R. (2007). Humans trade off viewing time and movement duration to improve visuomotor accuracy in a fast reaching task. *Journal of Neuroscience*, *27*(26), 6984–6994.
Blackford, T., Holcomb, P. J., Grainger, J., & Kuperberg, G. R. (2012). A funny thing happened on the way to articulation: N400 attenuation despite behavioral interference in picture naming. *Cognition*, *123*(1), 84–99. https://doi.org/10.1016/j.cognition.2011.12.007.
Bock, K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, *89*(1), 1–47. https://doi.org/10.1037/0033-295X.89.1.1.
Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(4), 575–586. https://doi.org/10.1037/0278-7393.12.4.575.
Bock, K. (1995). Sentence production: From mind to mouth. In *Speech, language, and communication* (pp. 181–216). Academic Press. https://doi.org/10.1016/B978-012497770-9/50008-X. Quote Reprinted with permission from Elsevier.
Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review*, *3*(4), 395–421. https://doi.org/10.3758/BF03214545.
Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, *31*(1), 99–127. https://doi.org/10.1016/0749-596X(92)90007-K.
Bögels, S. (2020). Neural correlates of turn-taking in the wild: Response planning starts early in free interviews. *Cognition*, *203*, 104347. https://doi.org/10.1016/j.cognition.2020.104347.
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. https://doi.org/10.1038/srep12881.
Bowcher, W. (2003). Speaker contributions in radio sports commentary. *Text & Talk*, *23*(4), 445–476. https://doi.org/10.1515/text.2003.018.
Branigan, H. P., & McLean, J. F. (2016). What children learn from adults' utterances: An ephemeral lexical boost and persistent syntactic priming in adult–child dialogue. *Journal of Memory and Language*, *91*, 141–157. https://doi.org/10.1016/j.jml.2016.02.002.
Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, *109*(2), 274–280. https://doi.org/10.1016/j.cognition.2008.07.011.
Brown-Schmidt, S., & Konopka, A. E. (2015). Processes of incremental message planning during conversation. *Psychonomic Bulletin & Review*, *22*(3), 833–843. https://doi.org/10.3758/s13423-014-0714-2.
Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, *54*(4), 592–609. https://doi.org/10.1016/j.jml.2005.12.008.
Bunger, A., Papafragou, A., & Trueswell, J. C. (2013). Event structure influences language production: Evidence from structural priming in motion event description. *Journal of Memory and Language*, *69*(3), 299–323. https://doi.org/10.1016/j.jml.2013.04.002.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234.
Chapman, C. S., Gallivan, J. P., Wood, D. K., Milne, J. L., Culham, J. C., & Goodale, M. A. (2010). Reaching for the unknown: Multiple target encoding and real-time decision-making in a rapid reach task. *Cognition*, *116*(2), 168–176. https://doi.org/10.1016/j.cognition.2010.04.008.
Chia, K., & Kaschak, M. P. (2022). Structural priming in question-answer dialogues. *Psychonomic Bulletin & Review*, *29*(1), 262–267. https://doi.org/10.3758/s13423-021-01976-z.
Cisek, P. (2006). Integrated neural processes for defining potential actions and deciding between them: A computational model. *Journal of Neuroscience*, *26*(38), 9761–9770. https://doi.org/10.1523/JNEUROSCI.5605-05.2006.
Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron*, *45*(5), 801–814. https://doi.org/10.1016/j.neuron.2005.01.027.
Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3.
Clark. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181–204. https://doi.org/10.1017/S0140525X12000477.
Cohen-Goldberg, A. M. (2012). Phonological competition within the word: Evidence from the phoneme similarity effect in spoken production. *Journal of Memory and Language*, *67*(1), 184–198. https://doi.org/10.1016/j.jml.2012.03.007.
Coppock, E. (2010). Parallel grammatical encoding in sentence production: Evidence from syntactic blends. *Language and Cognitive Processes*, *25*(1), 38–49. https://doi.org/10.1080/01690960902840261.
Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse Processes*, *55*(2), 230–240. https://doi.org/10.1080/0163853X.2017.1330031.
Creem-Regehr, S. H., & Kunz, B. R. (2010). Perception and action. *WIREs Cognitive Science*, *1*(6), 800–810. https://doi.org/10.1002/wcs.82.
de Zubicaray, G., McMahon, K., Eastburn, M., & Pringle, A. (2006). Top-down influences on lexical selection during spoken word production: A 4T fMRI investigation of refractory effects in picture naming. *Human Brain Mapping*, *27*(11), 864–873. https://doi.org/10.1002/hbm.20227.
Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*(1), 123–147. https://doi.org/10.1037/0033-295X.104.1.123.
Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 611–629. https://doi.org/10.1016/S0022-5371(81)90202-4.
Desmarais, F., & Bruce, T. (2009). The power of the local in sports broadcasting: A cross-cultural analysis of rugby commentary. *International Journal of Sport Communication*, *2*(2), 129–145. https://doi.org/10.1123/ijsc.2.2.129.
Desmarais, F., & Bruce, T. (2010). The power of stereotypes: Anchoring images through language in live sports broadcasts. *Journal of Language and Social Psychology*, *29*(3), 338–362. https://doi.org/10.1177/0261927X10368836.
Dohsaka, K., & Shimazu, A. (1996). A computational model of incremental utterance production in task-oriented dialogues. In *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*COLING. https://aclanthology.org/C96-1052.
Enachescu, V., Schrater, P., Schaal, S., & Christopoulos, V. (2021). Action planning and control under uncertainty emerge through a desirability-driven competition between parallel encoding motor plans. *PLoS Computational Biology*, *17*(10), e1009429.

Faisal, A. A., & Wolpert, D. M. (2009). Near optimal combination of sensory and motor uncertainty in time during a naturalistic perception-action task. *Journal of Neurophysiology*, *101*(4), 1901–1912. https://doi.org/10.1152/jn.90974.2008.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. https://doi.org/10.1016/j.brainres.2006.06.101.

Fenk-Oczlon, G. (1989). Word frequency and word order in freezes. *Linguistics*, *27*(3), 517–556. https://doi.org/10.1515/ling.1989.27.3.517.

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

Fraundorf, S. H., & Watson, D. G. (2013). Alice's adventures in um-derland: Psycholinguistic sources of variation in disfluency production. *Language and Cognitive Processes*, *29*(9), 1083–1096. https://doi.org/10.1080/01690965.2013.832785.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1.

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, *91*(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011.

Gallivan, J. P., Barton, K. S., Chapman, C. S., Wolpert, D. M., & Randall Flanagan, J. (2015). Action plan co-optimization reveals the parallel encoding of competing reach movements. *Nature Communications*, *6*(1), 7428. https://doi.org/10.1038/ncomms8428.

Gallivan, J. P., Chapman, C. S., Wolpert, D. M., & Flanagan, J. R. (2018). Decision-making in sensorimotor control. *Nature Reviews Neuroscience*, *19*(9). Article 9. https://doi.org/10.1038/s41583-018-0045-9.

Garrett, M. F. (1989). Processes in language production. *Linguistics: The Cambridge Survey*, *3*, 69–96.

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544–569. https://doi.org/10.1016/j.jml.2007.01.007.

Grent, T., Oostenveld, R., Medendorp, W. P., & Praamstra, P. (2015). Separating visual and motor components of motor cortex activation for multiple reach targets: A visuomotor adaptation study. *Journal of Neuroscience*, *35*(45), 15135–15144.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*(1), B1–B14. https://doi.org/10.1016/S0010-0277(01)00138-X.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279. https://doi.org/10.1111/1467-9280.00255.

Guhe, M., Habel, C., & Tschander, L. (2004). Incremental generation of interconnected preverbal messages. *Trends in Linguistics Studies and Monographs*, *157*, 7–52.

Gussow, A. E., & MacDonald, M. C. (2021). Language production under uncertainty: Advance planning and incrementality [Paper presentation]. In *34th Annual CUNY conference on human sentence processing*.

Haith, A. M., Huberdeau, D. M., & Krakauer, J. W. (2015). Hedging your bets: Intermediate movements as optimal behavior in the context of an incomplete decision. *PLoS Computational Biology*, *11*(3), e1004171.

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science*, *8*(3), 191–219. https://doi.org/10.1016/S0364-0213(84)80001-4. Examples used with permission of John Wiley & Sons - Books; permission conveyed through Copyright Clearance Center, Inc.

Heller, D., Skovbroten, K., & Tanenhaus, M. K. (2009). Experimental evidence for speakers sensitivity to common vs. privileged ground in the production of names. *Proceedings of the Pre-Cognitive Science*.

Hening, W., Favilla, M., & Ghez, C. (1988). Trajectory control in targeted force impulses. V. Gradual specification of response amplitude. *Experimental Brain Research*, *71*(1), 116–128. https://doi.org/10.1007/BF00247527.

Hoyle, S. M. (1991). Children's competence in the specialized register of sportscasting. *Journal of Child Language*, *18*(2), 435–450. https://doi.org/10.1017/S0305000900011144.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2*, 255. https://doi.org/10.3389/fpsyg.2011.00255.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001.

Jaeger, F. T. F., Furth, K., & Hilliard, C. (2012). Incremental phonological encoding during unscripted sentence production. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00481.

Jaeger, T. F., & Norcliffe, E. J. (2009). The cross-linguistic study of sentence production. *Language and Linguistics Compass*, *3*(4), 866–887. https://doi.org/10.1111/j.1749-818X.2009.00147.x.

Jongman, S. R., & Meyer, A. S. (2017). To plan or not to plan: Does planning for production remove facilitation from associative priming? *Acta Psychologica*, *181*, 40–50. https://doi.org/10.1016/j.actpsy.2017.10.003.

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.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and english. *Journal of Psycholinguistic Research*, *32*(1), 37–55. https://doi.org/10.1023/A:1021933015362.

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

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, *66*(1), 143–162. https://doi.org/10.1016/j.jml.2011.08.003.

Konopka, A. E., & Brown-Schmidt, S. (2014). Message encoding. In *The Oxford handbook of language production* (pp. 3–20). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199735471.013.010.

Koranda, M. J., Bulgarelli, F., Weiss, D. J., & MacDonald, M. C. (2020). Is language production planning emergent from action planning? A preliminary investigation. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01193.

Krüger, M., & Hermsdörfer, J. (2019). Target uncertainty during motor decision-making: The time course of movement variability reveals the effect of different sources of uncertainty on the control of reaching movements. *Frontiers in Psychology*, *10*, 41. https://doi.org/10.3389/fpsyg.2019.00041.

Kuchinsky, S. E., & Bock, K. (2010). *Paper Presented at the 23rd Meeting of the CUNY Human Sentence Processing Conference*.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299.

Lebkuecher, A. L., Schwob, N., Kabasa, M., Gussow, A. E., MacDonald, M. C., & Weiss, D. J. (2022). Hysteresis in motor and language production. *Quarterly Journal of Experimental Psychology*, *17470218221094568*. https://doi.org/10.1177/17470218221094568.

Lee, E.-K., Brown-Schmidt, S., & Watson, D. G. (2013). Ways of looking ahead: Hierarchical planning in language production. *Cognition*, *129*(3), 544–562. https://doi.org/10.1016/j.cognition.2013.08.007.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
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.
Liu, G., Chua, R., & Enns, J. T. (2008). Attention for perception and action: Task interference for action planning, but not for online control. *Experimental Brain Research*, *185*(4), 709–717. https://doi.org/10.1007/s00221-007-1196-5.
MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*(2), 157–201. https://doi.org/10.1080/01690969408402115.
MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00226.
MacDonald, M. C. (2016). Speak, act, remember: The language-production basis of serial order and maintenance in verbal memory. *Current Directions in Psychological Science*, *25*(1), 47–53. https://doi.org/10.1177/0963721415620776.
MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology*, *3*(2), 210–227. https://doi.org/10.1016/0010-0285(72)90004-7.
Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, *15*(1), 19–44.
MacWhinney, B. (1977). Starting points. *Language*, 152–168.
Martin, R. C., Crowther, J. E., Knight, M., Tamborello, F. P., & Yang, C.-L. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition*, *116*(2), 177–192. https://doi.org/10.1016/j.cognition.2010.04.010.
McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive activation and mutual constraint satisfaction in perception and cognition. *Cognitive Science*, *38*(6), 1139–1189.
Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture–word interference experiments. *Journal of Memory and Language*, *35*(4), 477–496. https://doi.org/10.1006/jmla.1996.0026.
Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1146.
Myachykov, A., Scheepers, C., Garrod, S., Thompson, D., & Fedorova, O. (2013). Syntactic flexibility and competition in sentence production: The case of English and Russian. *Quarterly Journal of Experimental Psychology*, *66*(8), 1601–1619. https://doi.org/10.1080/17470218.2012.754910.
Myachykov, A., & Tomlin, R. (2008). Perceptual priming and structural choice in Russian sentence production. *Journal of Cognitive Science*, *9*, 31–48. https://doi.org/10.17791/jcs.2008.9.1.31.
Oudejans, R. R., Michaels, C. F., & Bakker, F. C. (1997). The effects of baseball experience on movement initiation in catching fly balls. *Journal of Sports Sciences*, *15*(6), 587–595.
Paolazzi, C. L., Grillo, N., & Santi, A. (2021). *The source of passive sentence difficulty: Task effects and predicate semantics, not argument order* (pp. 359–393). Brill. https://doi.org/10.1163/9789004433427_012.
Pastor-Bernier, A., & Cisek, P. (2011). Neural correlates of biased competition in premotor cortex. *Journal of Neuroscience*, *31*(19), 7083–7088. https://doi.org/10.1523/JNEUROSCI.5681-10.2011.
Paul, H. (1880). *Prinzipien der Sprachgeschichte [Principles of the history of language]*. Tübingen, Germany: Max Niemeyer.
Piai, V., Klaus, J., & Rossetto, E. (2020). The lexical nature of alpha-beta oscillations in context-driven word production. *Journal of Neurolinguistics*, *55*, 100905. https://doi.org/10.1016/j.jneuroling.2020.100905.

Piai, V., Roelofs, A., Jensen, O., Schoffelen, J.-M., & Bonnefond, M. (2014). Distinct patterns of brain activity characterise lexical activation and competition in spoken word production. *PLoS One*, *9*(2), e88674. https://doi.org/10.1371/journal.pone.0088674.

Piai, V., Rommers, J., & Knight, R. T. (2018). Lesion evidence for a critical role of left posterior but not frontal areas in alpha–beta power decreases during context-driven word production. *European Journal of Neuroscience*, *48*(7), 2622–2629. https://doi.org/10.1111/ejn.13695.

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

Popov, A. (2019). *Live sports commentating: A comparative study of cricket, football, horse racing and tennis commentaries* (PhD Thesis). University of Zurich.

Rosenbaum, D. A., Weber, R. J., Hazelett, W. M., & Hindorff, V. (1986). The parameter remapping effect in human performance: Evidence from tongue twisters and finger fumblers. *Journal of Memory and Language*, *25*(6), 710–725. https://doi.org/10.1016/0749-596X(86)90045-8.

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology*, *60*(3), 362.

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.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition*, *53*(2), 91–127. https://doi.org/10.1016/0010-0277(94)90067-1.

Siegman, A. W., & Pope, B. (1966). Ambiguity and verbal fluency in the TAT. *Journal of Consulting Psychology*, *30*(3), 239–245. https://doi.org/10.1037/h0023374.

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

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

Stewart, B. M., Baugh, L. A., Gallivan, J. P., & Flanagan, J. R. (2013). Simultaneous encoding of the direction and orientation of potential targets during reach planning: Evidence of multiple competing reach plans. *Journal of Neurophysiology*, *110*(4), 807–816. https://doi.org/10.1152/jn.00131.2013.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (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.

Strijkers, K., Holcomb, P. J., & Costa, A. (2011). Conscious intention to speak proactively facilitates lexical access during overt object naming. *Journal of Memory and Language*, *65*(4), 345–362. https://doi.org/10.1016/j.jml.2011.06.002.

Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language*, *65*(3), 318–330. https://doi.org/10.1016/j.jml.2011.04.009.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly*, *30*(4), 415–433.

Tzagarakis, C., Ince, N. F., Leuthold, A. C., & Pellizzer, G. (2010). Beta-band activity during motor planning reflects response uncertainty. *Journal of Neuroscience*, *30*(34), 11270–11277.

Tzagarakis, C., West, S., & Pellizzer, G. (2015). Brain oscillatory activity during motor preparation: Effect of directional uncertainty on beta, but not alpha, frequency band. *Frontiers in Neuroscience*, *9*, 246.

van de Velde, M., Meyer, A. S., & Konopka, A. E. (2014). Message formulation and structural assembly: Describing "easy" and "hard" events with preferred and dispreferred syntactic structures. *Journal of Memory and Language*, *71*(1), 124–144. https://doi.org/10.1016/j.jml.2013.11.001.

van Helvert, M. J. L., Oostwoud Wijdenes, L., Geerligs, L., & Medendorp, W. P. (2021). Cortical beta-band power modulates with uncertainty in effector selection during motor planning. *Journal of Neurophysiology*, *126*(6), 1891–1902. https://doi.org/10.1152/jn.00198.2021.

Wanta, W., & Leggett, D. (1988). "Hitting paydirt": Capacity theory and sports announcers' use of cliches. *Journal of Communication*, *38*(4), 82–89. https://doi.org/10.1111/j.1460-2466.1988.tb02071.x.

Wilshire, C. E. (1999). The "Tongue Twister" paradigm as a technique for studying phonological encoding. *Language and Speech*, *42*(1), 57–82. https://doi.org/10.1177/00238309990420010301.

Wolpert, D. M., & Landy, M. S. (2012). Motor control is decision-making. *Current Opinion in Neurobiology*, *22*(6), 996–1003. https://doi.org/10.1016/j.conb.2012.05.003.

Wong, A. L., & Haith, A. M. (2017). Motor planning flexibly optimizes performance under uncertainty about task goals. *Nature Communications*, *8*, 14624. https://doi.org/10.1038/ncomms14624.

Wundt, W. (1900). *Die Sprache [Language](Vol. 1)*. Leipzig: Kroner-Engelmann.

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

This page intentionally left blank

CHAPTER FOUR

Speaking in dialects: How dialect words are represented and selected for production

Eleanor Heggdal Lønes, Yuki Kamide, and Alissa Melinger*
Division of Psychology, University of Dundee, Dundee, United Kingdom
*Corresponding author: e-mail address: a.melinger@dundee.ac.uk

Contents

Abstract

In the current review, we explore the processes and representations that underpin dialectal word production. Most monolingual speakers have a diverse set of repertoires at their disposal, including dialectal and register variants, that they may deploy in the appropriate social context. This linguistic flexibility must be supported by processes that allow speakers to tailor their utterances not only to convey the semantic content of their message but the social and pragmatic content as well. Despite the ubiquitousness

Psychology of Learning and Motivation, Volume 78
ISSN 0079-7421
https://doi.org/10.1016/bs.plm.2023.02.003

of these different repertoires in everyday language use, very little research has been undertaken to understand how multiple linguistic varieties are represented, organized and ultimately produced by "monolingual" speakers.

In this article, we provide an overview of the extant literature and theories on dialect-level language processing, considering dialectal language production across multiple levels of representation with a particular focus on word production. The review covers a number of psycholinguistic methodologies including picture-word interference and language-switching paradigms, as well as sociolinguistic observations such as dialect leveling and style shifting. We discuss both monolingual and bilingual models of word production with the aim of determining whether bidialectalism should rightly be seen as a special case of bilingualism, or whether this conceptualization is inaccurate. We also review evidence regarding strategic and automatic dialectal alignment processes. We conclude the review by proposing a new model of bidialectal production based on our findings and we suggest how to fill the gaps identified in the existing literature and theory.

1. Introduction

Through language, we are empowered to share our ideas, emotions, and knowledge with others. It is an incredibly powerful social and cognitive tool that humans have evolved to master. It is also an incredibly complex process, involving the translation of abstract, non-linguistic thoughts into a coordinated series of muscle movements for the production of spoken, written, or signed language. This translation process involves mapping conceptual information onto grammatically specified lexical representations and retrieving associated phonological/orthographic representations which are then assembled into structured syllables, phrases, and sentences. Over recent decades, the field of language production has expanded immensely, with great debates focussing on the number and nature of the stages of processing in production, the dynamics of activation flow across the system, and the nature of the decision processes that underpin selection at each stage, among others.

Most language production research has focused on monolingual speakers. However, even monolingual speakers have a diverse set of repertoires at their disposal, including dialectal and register variants that speakers may deploy in the appropriate social context. This means there must be a process that allows speakers to tailor their utterances not only to convey the semantic content of their message but the social and pragmatic content as well. Despite the ubiquitousness of these different repertoires in everyday language use, very little research has been undertaken to understand how multiple linguistic varieties are represented, organized and ultimately produced by "monolingual" speakers.

In the current review, we consider dialectal language production across levels of representation, with a focus on word production. This article provides an overview of relevant literature and theoretical considerations with regard to language processing and production at the dialect level. The review will cover monolingual and bilingual models of production, with a view to clarifying whether bidialectalism can be regarded as a special case of bilingualism, or whether such a view is insufficient. A number of relevant psycholinguistic methodologies are covered in the review, including picture-word interference and task-switching paradigms, as are sociolinguistic observations such as dialect leveling and style shifting. The article concludes by summarizing the field's current state of play and suggesting how to fill the gaps identified in the existing literature and theory.

1.1 Defining dialects

One barrier to understanding how existing production models can account for dialectal language production is the absence of an accepted definition of what constitutes a dialect. Dialects are often defined as regional varieties that differ from the standard form of language by pronunciation, vocabulary and grammar (Trudgill, 2000). The limitation of this definition is that it provides no guidance on where the boundary between a language and a dialect should be drawn, as two distinct languages would clearly vary on these three dimensions as well. While size, prestige and mutual intelligibility are also factors which have been posited for distinguishing between languages and dialects (Haugen, 1966; Hudson, 1996; Wei, 2000), these criteria are problematic and can lead to artificial distinctions and inconsistent classification (Melinger, 2018). Further complicating the situation is the fact that dialect use frequently co-varies with other characteristics such as education level and minority status (e.g., Crinson & Williamson, 2004; Washington & Craig, 1998). In fact, some authors contend that "language" and "dialect" are socially and politically constructed, and therefore cannot be objectively distinguished at all (Hudson, 1996; Wei, 2000). Hazen (2001) goes further, calling into question whether bidialectalism, akin to bilingualism, can truly exist. And yet, it is well known that speakers can produce multiple repertoires, including regional varieties (Giles & Coupland, 1991; Labov, 1998). Many people across the world, including the UK, speak more than one dialect, such as a standard, "official" variety and a regional variety. These "bidialectal speakers" choose words, grammatical constructions, and even pronunciations that are appropriate to their current social situation. In many respects, this is similar to bilingual speakers, who also select the language that is appropriate to the context with surprisingly few intrusions from the language not-in-use. However, to what extent bidialectalism truly mirrors bilingualism has rarely been investigated.

Indeed, while much linguistic research has looked at distributional patterns of dialectal usage (e.g., Trudgill & Hannah, 2008 for English varieties) and some psychological research has investigated dialect acquisition and processing (e.g., Antoniou, Grohmann, Kambanaros, & Katsos, 2016; Floccia, Goslin, Girard, & Konopczynski, 2006; Kirk, Kempe, Scott-Brown, Philipp, & Declerck, 2018; Martin, Garcia, Potter, Melinger, & Costa, 2016; Ross & Melinger, 2017; Sumner & Samuel, 2009; Vangsnes, Söderlund, & Blekesaune, 2017; Woutersen, Cox, Weltens, & De Bot, 1994), still little is known about how dialects are represented and produced by a speaker. It is unclear whether dialects are represented and processed like separate languages or as a subset of rules and representations within the larger linguistic system. Labov (1998) argues for the latter, claiming that dialects are co-dependent rather than independent. Dialect speakers often retain or mix features from their different varieties in the same utterance, implying a lack of separation. Hazen (2001) notes there is little empirical evidence to suggest that a bidialectal speaker can keep their dialectal system completely separate, as (some) bilingual speakers can. Hence, they argue that bidialectal speakers and bilingual speakers are distinct in the organization of their linguistic systems.

Since the time these claims were made, some experimental work has been conducted to test them. Before examining the evidence for how speakers represent and produce dialectal variants, we will first introduce the broad principles of, and key observations for, monolingual and bilingual production, to provide a framework for considering dialectal processing.

1.2 Models of word production

Models of word production incorporate three main stages of processing, namely, the conceptualization, formulation, and articulation stages. Speaking begins with *conceptualization*, the formation of a *message*, which captures what the speaker intends to say. The message is a conceptual, pre-verbal, representation, meaning it does not include language-specific features. Instead, the message must be organized to be compatible with linguistic encoding (i.e., *thinking for speaking*, Slobin, 1987, 1996) and the linear nature of spoken language (Levelt, 1993). Importantly, the organization that results from the "thinking for speaking" process may differ for different languages (Kita & Özyürek, 2003; McNeill, 2000). The conceptualization stage is also where a speaker decides how much detail to include in their utterance, applying pragmatic principles such as Gricean maxims (Grice, 1975) and considering audience design (e.g., Gann & Barr, 2014; Horton & Gerrig, 2002).

Messages are translated into linguistic representations and structures specific to the language in use. This is the *formulation* process. First, the concept is mapped onto a lemma, which is the grammatical representation of the word. Lemmas specify the grammatical properties of the word, such as gender for nouns or transitivity for verbs. The mapping from a concept to a lemma is not one-to-one. Instead, the intent to produce even a simple one-word utterance leads to the activation of a cohort of semantically similar and related lexical candidates, one of which will best match the speaker's intended meaning. One of these active candidates must be selected, either through a competitive selection process (Abdel Rahman & Melinger, 2009, 2019) or a non-competitive process (Dell, 1986; Mahon, Costa, Peterson, Vargas, & Caramazza, 2007). The selected lemma in turn activates its morpho-phonological form, a lexeme, which must be retrieved from long-term memory and assembled into phonological syllables (see Roelofs, 1997 for a detailed description of these latter stages of processing). Once the phonological form is assembled and the component syllables are retrieved from the syllabary (Cholin, 2008), appropriate abstract motor programs can be retrieved (Browman & Goldstein, 1992) and passed to the articulatory system (Goldstein & Fowler, 2003) for *articulation*.

The process of producing even a single word is sensitive to a host of factors that have influenced the development of production models. Word naming latencies are sensitive to endogenous variables such as lexical frequency (Jescheniak & Levelt, 1994; Oldfield & Wingfield, 1965), word length (Meyer, Roelofs, & Levelt, 2003), semantic concreteness (Lupker, 1979), semantic density (Rabovsky, Schad, & Rahman, 2016), and name agreement (Vitkovitch & Tyrrell, 1995), to name but a few. They are also sensitive to exogenous influences, such as the context in which a word is produced. Generally speaking, producing a word in the context of other semantically similar words tends to slow down naming times and increase error rates (Belke, Meyer, & Damian, 2005; Glaser & Düngelhoff, 1984; Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Schnur, Schwartz, Brecher, & Hodgson, 2006), with some authors finding increased interference with increased semantic similarity (Rose, Aristei, Melinger, & Abdel Rahman, 2019). In contrast, producing a word in the context of phonologically similar words facilitates naming (Glaser & Düngelhoff, 1984; Lupker, 1979; Schriefers, Meyer, & Levelt, 1990, but see Breining, Nozari, & Rapp, 2016). The magnitude of the facilitation effect is also correlated with the amount of segmental overlap (Abdel Rahman & Melinger, 2008). Together with evidence from speech errors (e.g., Dell, 1986), these patterns of results have contributed to the arguments for distinct grammatical (lemma) and morpho-phonological (lexeme) representations as well as provided evidence for the activation of the semantically related cohort and for shared phonemic content across phonologically similar words.

1.3 Bilingual models of word production

Bilingual models build upon the foundation of monolingual models, including all the same stages of processing and types of representations. However, bilingual models additionally need to explain how the speaker's two languages are related and their use controlled, allowing the speaker to select only the intended linguistic system without intrusions or catastrophic interference from the language not-in-use. Most models of bilingual word production assume that there are distinct and separate lexical entries for words in both languages and that these words are organized into linguistic systems, for example, using language membership tags (Abutalebi & Green, 2008; Green, 1998). Research into bilingual language production has revealed that when bilinguals speak, both language systems become activated in parallel and exert an influence on processing (Costa, Caramazza, & Sebastian-Galles, 2000; De Bot, 1992; Green, 1986; Hermans, Bongaerts, De Bot, & Schreuder, 1998; Poulisse, 1997; Poulisse & Bongaerts, 1994). The two active systems interact in a variety of ways. Co-active lemmas from both languages appear to compete for selection (Hermans et al., 1998, but see below for a discussion about translation equivalents) and spread activation to the phonological level. In the case of cognates, words in the two variants that share both sound and meaning (e.g., English "hair" and the German equivalent "Haar"), this activation can converge on shared phonological representations, facilitating the retrieval of the target representation (Costa et al., 2000). In order to prevent blending or catastrophic interference between two language systems, it has been suggested that bilinguals inhibit the non-target language, using the language membership tags as a mechanism that can operate over the whole system (e.g., Green, 1986). This systemic inhibition has been cited as the source of language switching costs, which can be asymmetric in language learners (Meuter & Allport, 1999) or symmetrical in proficient bilinguals (Costa, Santesteban, & Ivanova, 2006). These, and other, distinctive patterns of behavior paint a broad picture of bilingual language organization and production. They have also served as a guide for the development of bidialectal research.

As certain behavioral characteristic of bilingualism are well documented, some researchers are now starting to apply bilingualism methodology to dialect speakers. In an effort to better understand how dialects are processed, the similarities and differences between bilinguals and bidialectals are being compared. This effort will not only improve our understanding of how dialects are processed, but also how monolingual, bilingual and bidialectal speakers should be defined and categorized. Having laid out the basic principles that underpin word production and some unique hallmarks of bilingual processing, we can now turn to consider the evidence that informs our understanding of bidialectal language production.

When considering the empirical evidence for dialect production, we will focus on the question of how bidialectal speakers select words from one of their two dialects, lexical selection, but will also examine the limited evidence related to other stages of processing. Our discussion will also touch on issues of lexical and conceptual organization and dialect control. Our aim is not to argue for any particular model of language production with this review. Also, for simplicity's sake, we may make processing assumptions based on the most common models used in the literature to discuss the issues. Despite these choices, we remain mostly agnostic as to the specific mechanisms or architectures and instead aim to focus on similarities and differences between monolingual, bilingual and bidialectal language production (mainly word production) with the aim of filling in this substantial gap in our understanding. We will also consider whether the reviewed findings help identify processing characteristics of language-hood or dialect-hood, following the arguments laid out in Melinger (2018; see also Kirk et al., 2018).

2. Are bidialectal speakers like bilingual speakers?

Bidialectal speakers have typically been classed as monolinguals for the purposes of language research, and they would likely self-report as monolingual too. However, many of the empirical questions asked of the bilingual lexicon apply to the bidialectal lexicon as well. Crucially, just as bilingual research has asked how bilingual speakers can effectively operate in one linguistic system without intrusions or interference from the other system, the same question can be asked of bidialectal speakers. Similarly, we can ask whether observations from bilinguals are also found for bidialectal speakers.

2.1 Evidence from repetition priming

The earliest study we are aware of that applied this approach to the investigation of bidialectal speakers was Woutersen et al. (1994) using a repetition priming study. Repetition priming is obtained when the same lexical representation is retrieved a second time, resulting in faster processing. Based on the observation that bilinguals show no interlingual repetition priming for non-cognate words (e.g., processing of "bike" is not facilitated by the prior processing of the Dutch translation equivalent "fiets"; Monsell, Matthews, & Miller, 1992), Woutersen et al. (1994) asked whether Dutch participants from the Maastricht region of the Netherlands would show inter-dialect repetition priming. They included both cognate and non-cognate stimuli in a repetition priming study using a lexical decision paradigm. They classified their participants either as dominant Dialect speakers or dominant Standard Dutch speakers. Their results showed that the Dialect speakers produced no repetition priming effect while the Standard Dutch speakers showed repetition priming for both cognates and non-cognates. Woustersen et al. interpreted this result as evidence that the Dialect speakers had separate and independent lexica for their two dialects, given the similarity in results to what has been observed in bilinguals. In contrast, the observation of repetition priming in the Standard Dutch group suggests the Maastricht dialect was subordinate to the Standard variant for this group. A similar pattern of results was recently reported for bidialectal speakers of Hakka (D1) and Mandarin Chinese (D2). In a lexical decision task, Chen and Zhou (2022) also found no repetition priming from D1 to D2 or from D2 to D1, and similarly interpreted the absence of priming as evidence for independent lexica.

2.2 Evidence from translation equivalent distractor words

Melinger (2018) used the same rationale to investigate lexical selection processes in Standard Scottish English (SSE, hereafter; D1)—Scottish (D2) bidialectal speakers. She capitalized on a unique pattern of results observed in bilingual speakers. Costa and Caramazza (1999), Costa, Miozzo, and Caramazza (1999) observed that translation equivalent words (e.g., *dog* - *perro*) facilitate production when most models of word production would have expected them to interfere with production. Specifically, using the PWI paradigm, they observed that Spanish (L1)—English (L2) bilingual participants were faster to produce the target "dog" when the distractor was the translation equivalent ("perro" superimposed on a picture of a dog) than when the distractor was an unrelated word from the non-target language ("mesa" (table)). In contrast, semantic interference was observed when the distractor was semantically related to the target picture regardless of the language of the distractor, with both "cat" and "gato" slowing participants' naming times for "dog" equally (relative to unrelated words from the respective languages). Since the original observation reported by Costa and colleagues, this *translation equivalent facilitation effect* has been replicated across other language pairs and by other research teams (Dylman & Barry, 2018; Hermans, 2004, see also Hall, 2011, for a meta-analysis).

Citing this result as a hallmark of bilingual lexical selection, and contrasting it with what is predicted by models of monolingual production, Melinger (2018, 2021) asked whether dialectal alternatives presented as distractor words would similarly speed picture naming times. In experiments that paralleled the design used by Costa and colleagues (1999), Melinger (2018, 2021) failed to reproduce a dialect translation facilitation effect across several studies. Instead, she repeatedly found a dialect translation interference effect. Interference was observed for SSE—Scottish bidialectal speakers presented with Scottish distractors while naming British English targets (e.g., distractor "breeks" for the target "trousers"), for British speakers presented with American distractors while naming British English targets (e.g., distractor "elevator" for the target "lift"), and more recently for Italian-Neapolitan bidialectal speakers naming pictures in Italian (distractor "pastenacchia" for the target "carota" which is carrot in Italian; Melinger, 2023a). While the *dialect equivalent translation interference effect* has proven robust when participants are naming pictures in their dominant dialect, interference is not observed when participants name the picture in their dispreferred dialect. The effect is also weaker with visual, rather than auditory, distractors. Semantic interference effects were also reliably observed, both from within and across dialects, when naming in both the dominant and non-dominant variants (Melinger, 2018).

Based on the differences observed for between-language and between-dialect translation equivalents, Melinger (2018, 2021) argued that the dialects investigated in her studies, namely Scots English, American English, and Neapolitan, compared to the standard for her sample, namely SSE, British English, and Italian, respectively, were not represented as distinct linguistic systems (i.e., not as separate languages). If they had been, she argued, translation equivalent facilitation should have been observed. Instead, she argued for the inclusion of sociolinguistic features, such as dialect or register status, as part of the feature distribution at the conceptual level, allowing dialectal equivalents to compete for selection with the target word like other semantically similar, but not identical, words.

Melinger (2018, 2021) limits her interpretation to the variant pairs she studies; she does not make the strong claim that no dialects are represented or processed like languages. Indeed, her argument rests on the claim that there is no reliable way to define a dialect in opposition to a language and that experimental manipulations, like the one developed by Costa and Caramazza (1999), Costa, Miozzo, and Caramazza (1999) and utilized in her studies, should be used to assess variants on a case-by-case basis. Importantly, she argues that these experimental techniques offer a new method for discriminating between languages and dialects.

Interestingly, Melinger (2018, 2021)'s premise assumed that within-language synonymous distractors (e.g., target "couch," distractor "sofa") would produce a sizable interference effect, as this is what most models of monolingual word production would predict. However, when that prediction was tested directly by Dylman and Barry (2018), that is not what they found. Participants named pictures in the context of near-synonymous written distractor words, (e.g., distractor "hound," target "dog"). They either named the pictures with the preferred label (e.g., "dog") or with the dispreferred label (e.g., "hound"). While times to produce the preferred picture name were not impacted by the presentation of the synonymous dispreferred picture label, times to produce the dispreferred label were facilitated by the presentation of the preferred label, relative to the unrelated condition. Replicating this effect with auditory distractor words, Melinger (2021) also found no facilitation effect, but also no interference effect, from synonymous distractor words when speakers were producing the preferred picture label. The fact that both bilingual translation equivalent and monolingual synonymous distractor words facilitated picture naming while dialectal translation equivalent distractor words slowed picture naming suggests there is something different in the representation of dialectal alternatives.

2.3 Evidence from dialect switching

Another hallmark of bilingual language processing is the ability to switch between languages with symmetrical switching costs. Meuter and Allport (1999) asked language learners to name pictures in their two languages, classifying each trial either as a non-switch trial, meaning naming was in the same language as the previous trial, or a switch trial, meaning naming was in a different language to the previous trial. They observed that language learners produced asymmetrical switching costs, with a larger cost for switching into their first language from their second language than when switching into their second language from their first language. According to the Inhibitory Control Model (Abutalebi & Green, 2008; Green, 1998), this asymmetric switch cost is due to the degree of cognitive control required to prevent interference from non-target language items while naming in the target language. The level of control required to successfully inhibit the dominant language when naming in the non-dominant language is far greater than the level of control required to inhibit the non-dominant language when naming in the dominant language. Therefore, once the dominant language has been successfully inhibited, the time required to subsequently overcome this inhibition when switching back to naming in the dominant language leads to a greater switch cost.

In contrast to the asymmetric switching costs observed for late learners of a language, Costa and Santesteban (2004) found that highly proficient bilinguals produced symmetrical switching costs, supporting the notion that inhibition in the switch task is proportionate to the activation level of the two systems. Kirk et al. (2018) and Kirk, Declerck, Kemp, and Kempe (2022) applied the same language switching paradigm to investigate how bidialectal speakers choose between dialectal alternatives. Specifically, they asked whether bidialectal speakers used the same control mechanisms to select between linguistic varieties as bilingual speakers. They reasoned that, if bidialectal speakers produce symmetrical switching costs when switching between their dialects, analogous to the pattern observed for highly proficient bilingual speakers, then it would suggest that bidialectal speakers utilize the same control mechanism as bilingual speakers. At the same time, if asymmetrical switch costs are observed, these should provide insights as to which dialect is dominant for a given population. When proficient speakers of two linguistic varieties were tested, Kirk et al. (2018) found symmetrical switching costs. But, when non-proficient speakers were tested, an asymmetry was observed (Kirk et al., 2022). Specifically, Kirk et al. (2022)

conducted a language switching task in which bidialectal participants responded in Orcadian (a variety spoken in Orkney, Scotland) or SSE. Their results revealed a switch cost asymmetry, with higher costs observed when participants switched back into Orcadian, suggesting that Orcadian, and not SSE, was the dominant variety for this population.

Interestingly, half of Kirk et al.'s (2018, 2022) stimuli were cognates, meaning the translation equivalents in the two dialects shared phonological content (e.g., "house"—"hoose," "eyes"—"ezz"). Studies from bilingualism have found that cognates are named faster than non-cognates (Costa et al., 2000). Specifically, in a picture naming task in which only one language was used, pictures with phonologically similar labels in the speaker's two languages were named more quickly than pictures with phonologically distinct labels. Crucially, when the same set of pictures was named by a monolingual speaker, no difference was observed, which confirms that the naming difference was not due to other spurious differences between the picture sets. This cognate facilitation effect provides evidence for the co-activation of the language not in use and evidence that activation from the two lemmas cascades down and co-activates the phonological representation. In the case of cognates, these phonological representations share content, and hence the phonological encoding is facilitated.[a] Like bilingual observations of cognate facilitation, Kirk and colleagues also found that pictures with related names across the two dialects were named faster than pictures with unrelated names. This observation is important because it suggests that even dialectal words that are phonologically related to the standard variant still must have their own lexical representation; the dialectal variant cannot be produced by selecting the standard form and performing some phonological operations.

All of the above studies use a shared rationale—they all took established findings from the bilingual literature and applied those designs and methods to the bidialectal context with the aim of determining whether a particular linguistic variant was represented and processed like an independent language or not. We now move away from the parallels with the bilingualism literature to examine how speakers decide between two dialectal variants in any given communicative encounter.

[a] An alternative interpretation can localize the facilitation effect at the lemma selection stage, if the phonological representations feed activation back to the lemma, strengthening the target and speeding selection (Bernolet, Hartsuiker, & Pickering, 2012; Wang, Cai, Wang, Branigan, & Pickering, 2020).

3. Choosing between dialects

All of the above studies investigated lexical selection processes in situations where the dialect to be used on any given trial was determined by the experimenter. However, in most real-world situations, the choice to use one dialectal form or another is determined by the speaker, based on the social context they are in. Such free choice situations are likely common in bidialectal contexts, as interlocutors from the same region will likely be familiar with the same varieties, making both variants viable options. Having discussed how language selection mechanisms operate in a bidialectal context, we can now begin to examine how linguistic choices are made in cross-dialectal communication. Specifically, we are interested in whether two speakers from different dialect backgrounds will continue to produce disparate linguistic characteristics as their interaction progresses, or whether they will converge on a common form. As an observed phenomenon, convergence is the process whereby a talker adjusts their linguistic output to match their interlocutor's characteristics more closely. This process is also known as alignment or accommodation. As these different terms are used to describe the same phenomenon in different literatures, this article shall use them interchangeably and consistently with the original research.

To be sure, alignment is not restricted to bidialectal communication. When any two people converse, their linguistic features can differ on a number of characteristics, including speech rate, accent, word choice and syntactic structure, among others. Over time, conversation partners can become more *aligned* across multiple linguistic levels including phonological (e.g., Pardo, 2013), lexical (e.g., Bortfeld & Brennan, 1997), syntactic (e.g., Branigan, Pickering, & Cleland, 2000) and prosodic (e.g., Levitan, Beňuš, Gravano, & Hirschberg, 2015).

Pickering and Garrod (2004) argue that alignment is an automatic process driven by a general priming mechanism, whereby recently processed features or representations become more accessible for subsequent processing. Such a mechanism should be observed across levels of representation and should not be sensitive to social factors such as one's emotional disposition toward the interlocutor. Other mechanisms that could also result in alignment are audience design (e.g., Bell, 1984), conceptual pacts (Brennan & Clark, 1996), and even strategic social motivations (e.g., Giles & Powesland, 1975). Here, the speaker might purposefully and strategically align with an interlocutor to appear more friendly, more accommodating, or more similar. Strategic alignment need not be conscious, but it can be. Like other types of alignment, it is acknowledged that speakers can align on dialectal features, but the underlying mechanism is unclear. In this section, we will explore the evidence for both kinds of mechanisms.

3.1 Alignment mechanisms across multiple levels of representation

The lexical choices one makes in an interaction are salient features of the linguistic repertoire in use. Selecting the wrong word for a given audience can result in communicative breakdown in extreme cases. Alignment on morphosyntactic and phonological features are arguably more subtle; speakers may not even be consciously aware of some of the differences in these levels between dialects. To compare alignment behaviors across levels of representation, Smith and Durham (2011) examined the style shifting behaviors in a population where the prevalence of the dialect was declining, namely in the Shetland Islands of Scotland. They observed a split in the usage of Shetland forms by younger speakers. While some younger speakers produced dialectal lexical, morphosyntactic and phonetic variants most of the time and at the same rate as older speakers, others used the standard variants almost exclusively. In a follow-up study (Smith & Durham, 2012), those original speakers who preserved the Shetland forms (dialect speakers) were interviewed by an SSE speaker while those speakers who exclusively used standard forms (standard speakers) were interviewed by a local community member. Because token counts for individual words were low in the interviews, the authors focused on specific frequent exemplars at each level of representation. At the lexical level, they examined the occurrence of the common word *ken/know*. At the morphosyntactic level, they focussed on productions of *be* in the perfect tense where SSE typically uses *have* (e.g., "I'*m* no *been* in Imelda's" instead of "I *have* not *been* in Imelda's"). At the phonological level, they studied the lexically-conditioned [ʌʉ]—[u:] alternation (e.g., house—hoose). At the phonetic level, productions of th-stopping (i.e., pronouncing [d] instead of /ð/ in word initial and medial positions; e.g., a mixed sentence "So I think */ð/ey* (they) [standard] just let him have */d/e* (the) [dialect] drum shop") were measured.

For lexical, morphosyntactic and phonological levels,[b] the dialect speakers aligned to the interviewer's use of SSE, producing more instances

[a] Smith and Durham (2012) argue that the "hoose"/"house" alternation is phonological in nature. However, Kirk et al. (2018) presented evidence that these sorts of cognates have separate lemma representations. Therefore, it is possible that this example actually reflects another example of lexical alignment.

of *know* in lieu of their preferred *ken,* more instances of "house" than their preferred "hoose," and more forms of "have" than their preferred use of "be" compared to their baseline rates. In contrast, the standard speakers did not align to the interviewer's use of the local dialect at the lexical, morphosyntactic or phonological levels; they persisted in their use of the standard forms. Their rates of the local forms did not vary from their baseline, suggesting that the standard speakers only had access to the standard forms. Interestingly, the phonetic variable was slightly different, as all the participants produced instances of both dialect and standard pronunciations. Here, individual differences were greater than group level differences: While some dialect speakers aligned their phonetic productions to the standard-speaking interviewer's, others did not. Likewise, some standard speakers aligned to the local interviewer's phonetic properties, but others did not. Overall, though the local phonetic characteristics were retained in all speakers, only half the speakers retained the wider dialect features.

Smith and colleagues' (2011, 2012) findings support their earlier claim of a divide in the younger generation, with some being bidialectal and others being monolectal. Those younger speakers who were bidialectal produced dialectal forms at a similar rate as the older generations and were able to style-shift across levels of representation when speaking to a SSE speaker. The monolectal speakers, in contrast, only produced standard lexical, phonological, and morphosyntactic forms. These speakers were raised in an environment where the dialect was the dominant form, so they surely were passively familiar with the dialect, but they were either unable or unwilling to produce it. Interestingly, all speakers produced some dialectal phonetic variants. Smith and Durham (2011, 2012) note that style-switching is related to social awareness of linguistic variables within the community, and the phonetic variables are less salient, therefore, less likely to be deliberately dropped. This interpretation implies a social, strategic motivation for aligning. In terms of production models, Smith and Durham (2011, 2012) concluded that bidialectalism is an extreme case of style-shifting rather than a special case of bilingualism. Following Hazen (2001), the evidence of dialect mixing, in the absence of evidence of dialect separation, suggests the two linguistic systems are not independent. Instead, they argue that developing linguistic and social factors trigger corresponding shifts along a dialect-standard continuum embedded within a single linguistic system.

Smith and Durham's (2011, 2012) study illustrated that when Shetland-SSE bidialectal speakers style shift, they do not shift between pure versions of either variant, but rather they move along a continuum, speaking more or less Shetland or more or less SSE, mixing features of both grammars. Phonetic properties also patterned differently from features at other levels of representation, with the authors alluding to strategic mechanisms for salient properties and automatic mechanisms for properties that speakers are less aware of. Hence, we see that different alignment mechanisms can potentially be isolated. In the next section, we examine alignment processes at different levels of representation separately in an effort to understand how speakers choose between dialectal variants at different levels of representation.

3.2 Lexical alignment

Lexical and morphosyntactic alignment has been observed for Spanish dialects. In Spanish, the Salvadoran and Mexican dialects have distinct pronominal systems. While Mexican Spanish eschews the formal forms of Castilian Spanish (the standard variant spoken in Spain) in favor of the "tuteo" system (for 2nd person singular pronouns), Salvadoran Spanish preserves the Castilian "voseo" system. (Note, in Spanish, pronoun choice entails verb conjugation as well.) Hernández (2002) observed that Salvadoran Spanish dialect speakers who had moved to the United States produced the Mexican dialect "tuteo" pronouns rather than their native "voseo" pronouns when interacting with a Mexican interviewer. Likewise, alignment was observed for lexically distinct verbs such as the Mexican Spanish variant "traer" and the Salvadoran Spanish variant "andar," which both mean "to carry something on one's person." Hernández (2002) found that Salvadoran speakers produced "traer" more when speaking with a Mexican interviewer than when speaking with fellow Salvadoran speakers. Similar alignment behavior has been observed in Scottish-English speakers when interacting with individuals on either side of the Scotland/England border (Watt, Llamas, & Johnson, 2010). While these studies clearly demonstrate that speakers can adjust their lexical choices and morphosyntax to be more dialectally similar to their interlocutor, it is difficult to attribute a processing mechanism in naturalistic settings; They cannot distinguish between an automatic priming mechanism, a strategic social alignment mechanism (Giles & Powesland, 1975) or a desire to establish rapport (Tannen, 1989). Interestingly, Kühne, Rosenthal-von der Pütten, and Krämer (2013) observed dialectal lexical alignment when German participants interacted with a computer avatar speaking either High German or the Rhine-Ruhr dialect. The fact that dialectal alignment occurs even with computerized entities may initially suggest an automatic process is at play here. However, multiple studies have shown that humans make social overtures toward computer avatars, even when fully aware they are interacting with a machine (e.g., Von Der Pütten, Krämer, Gratch, & Kang, 2010). Therefore, it remains difficult to identify the alignment mechanism in these naturalistic studies.

3.2.1 Knowledge about community membership

Lexical alignment can also be investigated experimentally to more effectively disentangle potential explanatory mechanisms, allowing more precision in the interpretation. One such investigation was conducted by Tobar-Henríquez, Rabagliati, and Branigan (2021), who examined whether participants' lexical choice was affected by community level factors (i.e., whether the interlocutor was from the same or different linguistic community) and how their interpersonal experiences within the experimental session contributed to the acquisition of community-level knowledge. In this context, "community-level" refers to linguistic differences that are based on nationality and can be considered a proxy mechanism for identifying which dialect of Spanish the confederate would speak, while "interpersonal" refers to the history of interactions a participant has with a specific interlocutor over the course of the experiment (also called common ground, or lexical entrainment). To give an example of a community-level difference, Castilian Spanish speakers prefer to use the word "patata" when referring to a potato, while Mexican Spanish speakers prefer to use "papa," however both terms are acceptable in both dialects. Across two blocks of trials, participants, who were either native Castilian Spanish speakers (Exp 1) or native Mexican Spanish speakers (Exp 2), completed an online study where they worked with a confederate (in reality, the confederate's responses were scripted) to take turns matching and naming objects. In the first block, participants were either paired with a confederate from their own linguistic community or from the outgroup linguistic community. All participants were presented with the same picture labels in matching trials. These picture labels were always the dispreferred label for the participant. In naming trials, participants' rates of use of the dispreferred labels were comparable across in-group and out-group conditions, suggesting a simple lexical-level priming effect that was insensitive to community preferences.

In the second block, participants were paired with a new in-group or out-group community member. In this block, the confederates did not name any of the experimental items, so the question the authors were interested in was whether participants were more likely to maintain the dispreferred label from the first block if it was community-appropriate for their new partner. If participants are sensitive to their partner's community membership (here viewed as dialect membership) and attribute the partner's choices to their community's linguistic preferences, then dispreferred labels might be preserved more often when speaking to the outgroup member than the ingroup member. Indeed, Tobar-Henríquez et al. (2021) found that, when participants had entrained on an outgroup partner in the first block, they maintained more dispreferred labels when speaking to an outgroup partner in the second block than an ingroup partner. This finding demonstrates that speakers can notice their partner's community membership, use this information to extrapolate beliefs about community-level preferences, and design utterances based on these preferences. Crucially, they do this when they believe their partner is a typical representative of the outgroup's preferences, and not when they are an idiosyncratic exemplar of the ingroup.

Interestingly, the participant–confederate interactions in this study were in the visual, written, modality, not the auditory, spoken, modality. The manipulation of community membership was achieved only by introducing each communication partner at the start of the experiment as being from a specific country. There were no other cues, such as accent, to ground or reinforce the dialect or community membership. As such, this study demonstrates that speakers can generate extremely abstract knowledge of community linguistic preferences based on minimal explicitly supplied information.

3.2.2 Abstract priming across the dialect lexicon

Heggdal Lønes, Kamide, and Melinger (2023) investigated whether lexical choice would be affected by the interlocutor's perceived dialect background, using Scottish dialect as the target dialect and contrasting it with standard British English. Crucially, unlike the Spanish items in Tobar-Henríquez et al.'s (2021) study, the Scottish and English lexical items here would not be equally acceptable for both Scottish and English community members. While Scottish dialect users can be assumed to have familiarity with both the Scottish dialect items and the standard English items, standard English speakers are unlikely to have any knowledge of the

Scottish items at all. In the first of two studies, an online picture matching-and-naming experiment tested Scottish dialect speakers who alternated between speaking to an English and a Scottish confederate on a trial-by-trial basis. The confederates' nationalities were introduced at the start of the study as part of the experimental cover story. On matching trials, the confederate's word choices were presented to participants in the written rather than the auditory modality, much like Tobar-Henríquez et al.'s (2021) study. On naming trials, both the dialect and the standard word options were presented visually for participants to choose via mouse click, alongside two unrelated distractors. Trials were equally divided between the Scottish and the English confederate, with trial order being randomized so that either partner could appear on any trial. The English confederate only produced standard English picture labels, e.g., "slippers," while the Scottish confederate produced uniquely Scottish picture labels for those trials where the dialects diverge, e.g., "baffies" (80% of the Scottish confederate's trials were uniquely Scottish items, while the remaining 20% were standard English filler items for realism). The aim here was twofold: first, to find out whether participants' lexical choice would change throughout the experimental session, with a general increase in Scottish words due to priming of abstract dialect-level features, and second, if so, whether the increase in Scottish words was reserved for exchanges with the Scottish partner. Stimuli were counterbalanced across lists so that individual participants never named an item they had already matched. If participants increased their usage of Scottish words on trials in which they were paired with the Scottish interlocutor, this would be evidence of dialect community membership affecting lexical choice. Alternatively, if participants increased their proportion of Scottish words generally, irrespective of whether they were directing a Scottish or English partner on a specific trial, this would be evidence of another mechanism, such as abstract dialect priming based on exposure to other Scottish words earlier in the session. The results supported the latter prediction but not the former; While participants did increase the proportion of Scottish words they produced in the second half of the session (related to the first half), they did not do so exclusively on trials where they were addressing the Scottish partner, as shown in Fig. 1. Therefore, we concluded that participants were not reacting to community membership. Instead, we interpret this specific priming effect to be evidence of some sort of dialectal lexical feature that links all Scottish-specific lexical items together, such that the priming can abstract over the lexicon.

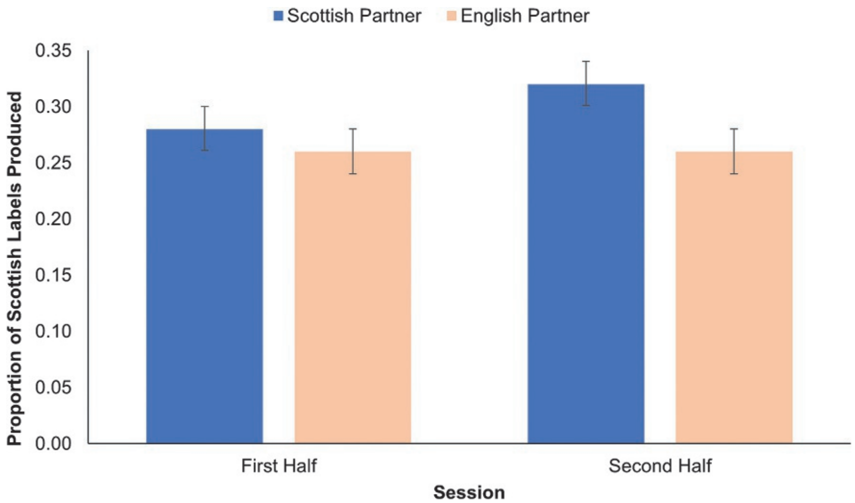

Fig. 1 Proportion of Scottish labels produced as a function of partner dialect in each half of the session.

To support the claim of abstract dialectical priming, Heggdal Lønes et al. conducted a second experiment employing a blocked variant of the picture matching-and-naming task. In this variant, participants first take part in a matching block where the confederate is the director. In the second block, the roles swap, with the participant naming pictures, ostensibly to direct the confederate in the matching task. As in the previous study, the experiment was conducted online using the written modality rather than spoken responses, hence no other dialect cues were available to influence the participants. Different items were again used in the two blocks, to test for abstract level priming across the dialect. In the matching block, Scottish participants worked with either an English confederate who only produced standard English picture labels, e.g., "trousers," or a Scottish confederate, who produced uniquely Scottish picture labels for those trials where the dialects diverged, e.g., "breeks." In the subsequent naming block, participants were all paired with a new Scottish partner. This design tests for automatic dialect-level priming, which is blind to the preferences of the current interlocutor, by maintaining identical social conditions during the naming block after manipulating dialect exposure in the matching block. Results showed that Scottish participants produced more Scottish picture labels after processing *different* Scottish words in the previous block, supporting abstract priming over the dialect lexicon, as shown in Fig. 2. Unlike

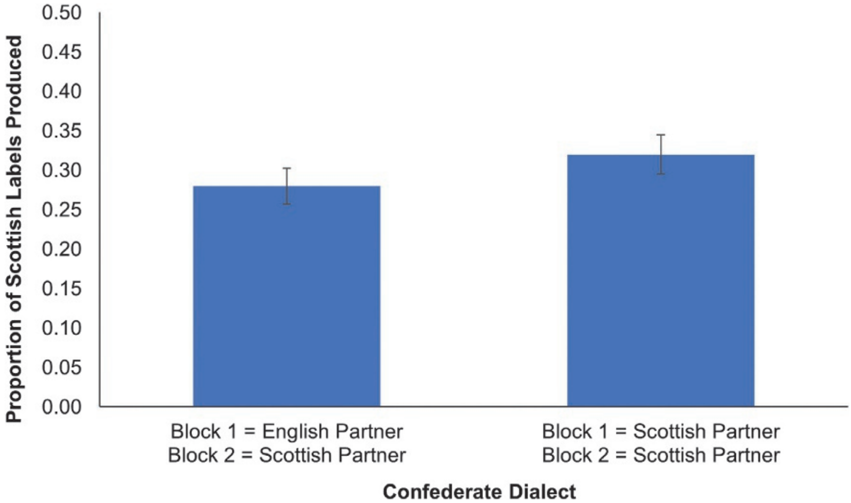

Fig. 2 Proportion of Scottish labels produced as a function of the stimuli dialect in Block 1.

Tobar-Henríquez et al. (2021), this experiment did not aim to test whether speakers designed their utterances with the preferences of their current partner in mind; rather, our results present evidence of a priming effect which occurs independent of the current partner.

Finding evidence of abstract priming (Heggdal Lønes et al., 2023) adds weight to the notion that some representational features bind all Scottish dialectal items together. This could be a dialect membership tag, akin to those proposed for bilingual language control (Green, 1998), or a conceptual feature shared by all dialectal words (La Heij, 2005). It should be noted that, as in Tobar-Henríquez et al.'s (2021) experiment, our experiments did not include any auditory or accentual cues to the confederates' nationality or dialect membership. Although prior work has shown that dialectal alignment can operate over form-level representations (Smith & Durham, 2012), the lack of any auditory cues in these experiments means the observed priming effects cannot be attributed to co-occurrence of features at the phonological or phonetic level. While Tobar-Henriquez et al. attributed their effect to participants' knowledge about the dialect membership and linguistic preferences of both their first and their second partner, this cannot be the same for our second study. In our study, the second partner's dialect membership did not differ between conditions, yet the rate of Scottish words that participants addressed toward the partner did. Clearly, the differences observed in block 2 must therefore have been driven by a difference implemented in block 1, so the question is which mechanism was responsible. Unlike in Tobar-Henriquez et al.'s study, our task partners' lexical choices were consistent with their stated community membership, so an explanation that relies on prototypicality of the partner's community membership would not apply. Likewise, the increased rate of Scottish word production cannot be due to lexical priming, as participants named different items than they matched. Hence, the activation and subsequent production of Scottish words by the participant can only be caused by abstract, dialect-level, priming from the dialect words encountered in block 1 to other dialect words encountered in block 2, thereby increasing their likelihood of being produced. To explain this in terms of knowledge, it would be the participant's knowledge about the dialect membership of words in their own lexicon rather than their knowledge of the interlocutor's dialect community membership which is responsible for the effect. In other words, the dialect words produced by the partner share some sort of dialect membership feature in the participant's lexicon, and the increased activation of these items spreads across all other items in the lexicon which share the same feature.

3.2.3 Dialect equivalents compete for activation

Enhancing the availability of dialect words can also lead to changes in the temporal dynamics of production. Using a blocked matching-naming task, Melinger (2023b) presented British participants with pictures of objects that have non-cognate American and British labels (e.g., "elevator"—"lift," "flashlight"—"torch"). The confederate was either American or British and in the first block they produced picture labels consistent with their dialect and participants identified the named picture from an array of options (Task 1). Crucially, and different from other studies using a similar paradigm, the study was conducted in person with spoken responses; hence, participants could rely on accentual cues to identify the nationality of the confederate as well as "overhearing" an exchange between the confederate and the experimenter where the confederate disclosed where they were from. In this way, the manipulation of dialect membership was stronger but less overt in this study compared to other work.

In the second block, the participant took on the director role, naming objects for the confederate (Task 2). Participants were paired with a new confederate who could also be either American or British, creating a 2×2 factorial manipulation of partner dialect in the 2 blocks. We coded the likelihood of the participant to produce the dispreferred American picture label based on (a) which dialect they were exposed to in the matching task and (b) which confederate they were paired with in the directing block. Since the same pictures were used in both blocks, we predicted that exposure to the American labels in the matching task should strengthen the activation level of the American labels, increasing the likelihood of the participant producing those American labels in the naming task, irrespective of the nationality of the confederate. Based on a theory of dialectal lexical selection that assumes selection by competition (Melinger, 2018), we further predicted that, even when the British lexical item was ultimately selected, the naming times should be slower due to competition with the primed American item.

Fig. 3 shows that participants produced more British picture labels when their matching task partner had produced British labels. Conversely, British labels were less likely, and in turn American labels more likely, when the matching task partner had produced American labels. The community membership of the confederate in the naming block had a negligible impact on the lexical choices of participants, suggesting that choice was driven by automatic priming from task 1 rather than the dialectal preferences of the task 2 partner, i.e., no evidence for audience design. Fig. 4 shows that, even when participants produced their preferred British labels in the

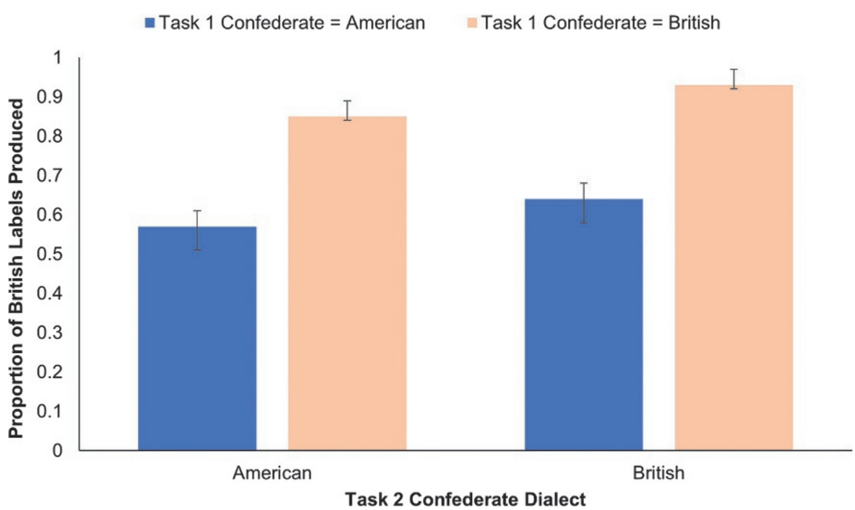

Fig. 3 Proportion of British responses as a function of the Task 1 and Task 2 confederates' dialect membership.

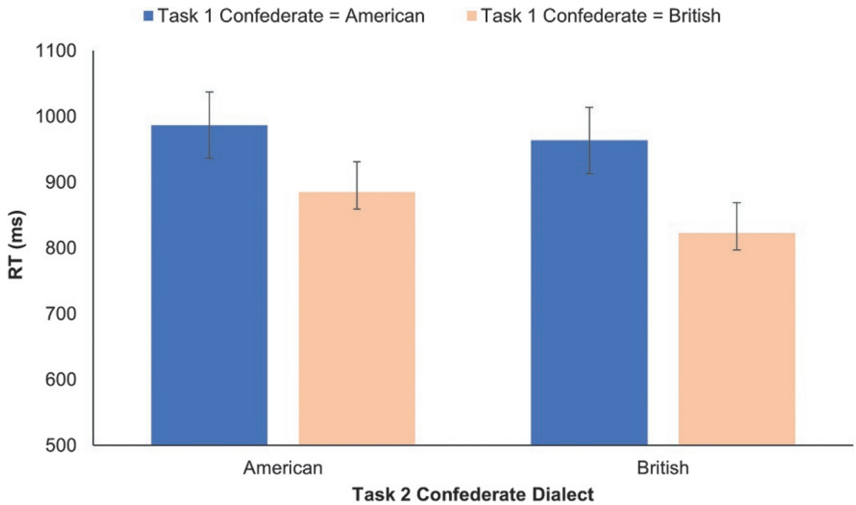

Fig. 4 Mean reaction time to produce British picture labels as a function of Task 1 and Task 2 confederates' dialect membership.

naming task, the time to initiate naming was sensitive to whether participants had recently processed the American alternative picture name, with longer naming times after previous exposure to the American label. There was also a modest influence of the dialect preferences of the current partner, with slightly faster naming times for British lexical items when paired with a British confederate in the naming block. While the former influence of the task 1 partner may reflect a repetition priming effect, the latter influence of the task 2 partner provides further evidence that dialectal alternatives compete for selection. In this context, preparing an utterance for an American partner increases the availability of the American picture labels, even when those labels have not been presented in the context of the experiment. This result represents a combination of previous observations. Specifically, it reflects the participants' ability to extract community membership linguistic behavior (Tobar-Henríquez et al., 2021), even when it was not modeled in the experiment. This may reflect the fact that participants may simulate the productions of their task partners, so the American labels could be internally activated through this simulation process (Kuhlen & Abdel Rahman, 2017).

3.3 Phonological alignment

When two dialects include distinct, non-cognate lexical items, it seems reasonable to assume that both variants must have their own distinct lexical representations, potentially linked to shared semantic representation.

The evidence from Kirk et al. (2018) further demonstrates that even cross-dialectal cognates have distinct lemma representations. The question addressed above is primarily how speakers select between these alternative lexical representations. When we consider form-based representations, the picture is more complex, as word forms are composed of segmental information that can be shared between dialects. Some dialectal variants can be regularly derived from a common abstract phonological representation. For example, the variant of American English spoken in New York City is characterized by r-less productions in word final position (e.g., "bak[ə]" which contrasts with the r-full standard American variant "bak[ɚ]"). This alternation is predictable and rule-governed, meaning the two dialects could share a common phonological representation which is then modified during phonological or phonetic encoding. For purposes of recognition, one's production experience can influence perceptual input representations. Sumner and Samuel (2009) demonstrated that active speakers of the New York dialect established separate phonological representations for the dialectal and standard forms while those who were familiar with the New York dialect but were not active speakers established a single phonological representation and both variants are mapped onto it (see also Sumner & Samuel, 2005). Similarly, speakers who do not perceive a phonological distinction have more difficulty producing that distinction. For bilinguals, Pallier, Bosch, and Sebastián-Gallés (1997, see also Bosch, Costa, & Sebastián-Gallés, 2000; Sebastián-Gallés & Soto-Faraco, 1999) showed that Spanish dominant Spanish-Catalan bilinguals had difficulty learning the Catalan vowel contrast /e/—/ε/, as Spanish does not have an /ε/ phoneme. Similar difficulty discriminating between vowels has been observed for speakers of dialects that have merged vowels (e.g., *pin*—*pen* in certain US English dialects; Labov, Karen, & Miller, 1991). What these observations demonstrate is a link between production and perception; Experience with a dialect, specifically actively speaking a dialect, can influence which phonological representations are established, both for the production and the recognition systems. Note, this is somewhat in contrast to what has been found for lexical selection. For example, Kirk et al. (2018) found that both active and passive users of the Dundonian dialect produced equivalent symmetrical switching costs.

3.3.1 Speakers align phonologically to the dialect of the current interlocutor

Despite its less salient nature, phonological alignment has been reported in observational and experimental studies. In naturalistic settings, phonological alignment has been shown to be driven by the immediate context and characteristics of the interlocutor. For example, Hernández (2009) investigated the productions of word-final velar nasals, a characteristic of Salvadoran Spanish, by Salvadoran Spanish speakers living in the United States where the non-velar alternative typical of Mexican Spanish predominates. Speakers were interviewed either by a fellow Salvadoran or Mexican dialect speaker. Hernández (2009) analyzed the interviewees' speech for presence of word-final nasal velarization. The results showed that Salvadorans produced a non-velarized nasal in word final position more often than their preferred velarized pronunciation when the interviewer was Mexican. Salvadoran dialect speakers who were interviewed by a fellow Salvadoran dialect speaker retained the word-final nasal velarization characteristic of the Salvadoran dialect. All speakers had lived in the same area of the United States and had extensive exposure to the Mexican variant. Hence, the results show that merely living in an area where the Mexican pronunciation predominated and had higher prestige did not lead to a blanket switch to Mexican dialect productions in Salvadoran dialect speakers. In fact, the rate of word-final nasal velarization produced when interacting with a fellow Salvadoran was the same as the rate produced by a control group of Salvadorans who still lived in El Salvador and had not had extensive exposure to the non-velarized variant. Rather, the shift in nasal velarization was only observed when speakers were engaging with a speaker of the Mexican dialect. Hence, the change of dialect characteristics produced by interviewees was the product of alignment to their current interlocutor.

3.3.2 Phonological alignment is mediated by social attitudes

Social attitudes toward a community group can also impact alignment behavior. An English language study by Babel (2010) revealed that New Zealand speakers converged phonetically on some Australian vowels (e.g. "dress," pronounced as [dɹi˕s] in New Zealand English and as [de̞s] in Australian English) when interacting with an Australian talker. The degree of convergence observed was affected by the New Zealanders' pre-existing feelings toward Australians, with those who had more positive attitudes demonstrating greater convergence. Importantly, the effect of pre-existing attitudes was linked to the community group and not the individual interlocutor, which again demonstrates that speakers can extract social and linguistic information about a speaker based on their accent, consistent with claims by Tobar-Henríquez et al. (2021). Moreover, the fact that the degree of convergence observed differs based on the speaker's social evaluation of their interlocutor speaks against alignment being a purely automatic process. Instead, such a pattern is indicative of a more strategic process, as suggested by Giles and Powesland (1975).

It is also interesting to consider how the effects of long-term exposure to another dialect (such as immersion due to moving to a new area) differ from the short-term effects observed in laboratories and interviews. There is evidence that perception processes become more flexible when subjected to more variable input, and seeing as perception and production are linked, it might be the case that production of the new dialect gets better as the amount of input increases, perhaps parallel in shape but lagging behind. As mentioned earlier, Salvadoran Spanish speakers living alongside Mexicans in the United States produced dialectal features which aligned with those of their current interlocutor. While this effect was observed for the whole group, it is important to note that the participants comprised both first-generation immigrants and their children. While length of time spent in contact with the Mexican variety was not measured, age of arrival was. Older arrivals had a higher relative frequency of the Salvadoran pronunciation than younger arrivals. The explanation given for the difference was that younger arrivals were more motivated to align with their peer groups and the wider social milieu, which favored the Mexican dialect.

Along these lines, a study by Campbell-Kibler, Walker, Elward, and Carmichael (2014) investigated the underlying forces that influence long-term phonetic accommodation. Using speech recordings from a corpus of Ohio State University students, the pronunciations of native Ohioans from different dialect areas were analyzed both as a function of time spent at university and as a function of self-reported social contact with native Ohioans from other dialect areas. The results found no effect of time spent at university but did find a tentative effect of social network. Similar to Hernández (2009), Campbell-Kibler et al. (2014) speculated that social identity, which undergoes important development in adolescence, may be the main driving force behind individuals' linguistic changes. An important point about this study is that no interlocutor was present during the recordings; accommodation in this study refers to shifts in accent resulting from contact with their social network rather than socially-driven alignment to a specific interlocutor. Universities are places where young people from different backgrounds mingle, so it is not surprising that social relationships play the greatest role here.

3.3.3 The salience of a contrast affects the degree of phonological alignment

A study by Walker and Campbell-Kibler (2015) investigated phonological convergence across English dialects under a more socially impoverished scenario. In a word shadowing task, the speech of participants from New Zealand and the US Midlands was recorded, first while they read from a target word list, and subsequently while they shadowed pre-recorded blocks of the same target words read by model speakers from four dialectal backgrounds (New Zealand, US Midlands, Australia, US Inland North). Convergence was calculated by subtracting the phonetic distance between baseline speech and shadowed speech from the phonetic distance between shadowed speech and model speech. Results showed that convergence was greater between more dissimilar dialects. Specifically, New Zealand participants displayed more convergence than US Midland participants overall, and converged more strongly with the two US models than the Australian model, while US Midland participants converged more strongly on the Australian model than the US Inland North model. Interestingly, both groups of participants demonstrated greater convergence on the most prototypical phonetic variables of their dialect (i.e., US Midlanders on rhoticity, New Zealanders on vowels), suggesting that metalinguistic awareness is one of the factors driving alignment. Note that speakers in this study converged even when aware that their supposed interlocutor could not hear them, suggesting that the presence of a pre-recorded speaker is a sufficient cue to trigger alignment. Overall, the study provides evidence that perceptual processes can lead to convergence in the absence of social motivation to align with a partner.

4. Discussion

The aim of this review was to examine the literature on dialectal word production to understand how speakers choose between their dialectal varieties and whether bidialectal word production is analogous to bilingual word production. To answer these questions, we reviewed evidence from psycholinguistic and sociolinguistic traditions, across lexical, morphosyntactic, phonological and phonetic levels of processing. The conclusions from these diverse domains converge on many points. In this discussion, we'll draw out what we see as the novel and important observations of this review. We'll conclude with a proposed model of bidialectal lexical organization.

4.1 Are dialects represented and processed like languages?

For the linguistic varieties that have been examined to date, the evidence suggests there are meaningful differences in how bilingual and bidialectal speakers represent and select words from their different repertoires. Findings from the PWI paradigm revealed that bilinguals show translation-equivalent facilitation (Costa & Caramazza, 1999; Costa et al., 1999) while bidialectals show dialect-equivalent interference (Melinger, 2018, 2021, 2023a). Melinger (2018) interpreted this as an indication that the dialects *she investigated* are represented differently to languages, arguing for co-dependent organizations for these dialects, in contrast to languages, which are represented independently (Labov, 1998). This conclusion is consistent with the evidence from Shetland dialect speakers. Following Hazen (2001), Smith and Durham (2012) reasoned that if bidialectal speakers have independent grammatical systems for their two dialects, then they should align across the levels of representation in similar ways. Their results instead showed that the four levels of representation, namely lexical, morphosyntactic, phonological and phonetic, followed different patterns, both for their Shetland dialect speakers and their standard speakers. In essence, no speaker style shifted completely into a purely Shetland or purely SSE style.

The observation that monolinguals produce synonym facilitation in the PWI paradigm (Dylman & Barry, 2018; Melinger, 2021) further suggested that there may be something unique about bidialectal speakers; they are not "just" monolinguals. Melinger (2021) argued that dialectal equivalents must mark their membership in different varieties, but they do so in a different way to bilinguals. Following a proposal from La Heij (2005), she proposed that dialect membership could be represented at the conceptual, message, level as a meaning feature of the words.

In contrast, in an interlingual repetition priming task, Woutersen et al. (1994; see also Chen & Zhou, 2022) found that dialect speakers did not show repetition priming between translation equivalents, similar to bilinguals. This was taken as evidence that the lexica for those speakers must be independent. While this study focussed on word recognition processes, a study by Kirk et al. (2018) also found similarities between bidialectal and bilingual processing in a word production task. Proficient bidialectal speakers produced symmetrical switching costs, while dialect learners produced asymmetrical switch costs—the same pattern as observed in bilinguals (Costa & Santesteban, 2004). Despite this common pattern, Kirk et al. (2018) do not argue in favor of dialect tags, which are the analog to language tags proposed to control selection in bilinguals (Green, 1998). Instead, they also advocate for a conceptual feature that denotes dialect status, arguing that it would be representationally untenable to have membership tags for every linguistic variety that a person can utilize. Note that some recent studies have also reported similarities between monolingual and bilingual switching (Declerck, Ivanova, Grainger, & Duñabeitia, 2020; Finkbeiner, Almeida, Janssen, & Caramazza, 2006; Ivanova & Hernandez, 2021), suggesting that the "bilingual" control mechanism may be better conceptualized as a general, top-down language control mechanism.

Hence, the jury on this key question is still out. Both the translation equivalent interference effect and the symmetric switching costs, initially identified as hallmarks of bilingual processing, have since been shown to be less defining than initially thought. In terms of the conditions that Hazen (2001) identified for bidialectalism to be analogous to bilingualism, we did not find any report that evidenced bidialectal's ability to switch between dialects in a pure fashion. Interestingly, despite the differences in the reported findings, most authors concur that bidialectal speakers have a single linguistic system and within that system, both dialects are represented.

4.2 How do dialect variants maintain separation within a single linguistic system?

If two dialect variants are shared within a single linguistic system, how then are dialect-specific representations distinguished? Relevant evidence is provided by Kirk et al. (2022, 2018) who obtained cognate facilitation effects for phonologically similar words in the two tested varieties. Cognate facilitation effects are taken as evidence for distinct lemmas and related phonological representations. This is an interesting finding, and it begs the question as to whether all cognates, including those distinguished by a regular phonological process, have distinct lemma representations. It also suggests that the "phonological" alternation examined by Smith and Durham (2012), namely the lexically-constrained alternation between "house" and "hoose," may in fact reflect a lexical selection choice rather than a phonological process.

We also reviewed examples where dialects differed on their morphosyntactic variables (Smith & Durham, 2012; Hernández, 2002). Grammatical features, such as lexical category, gender, number, transitivity, etc., are associated with lemma representations (Kempen & Hoenkamp, 1987; Levelt, Roelofs, & Meyer, 1999). Therefore, in addition to marking dialect membership as a feature of a words' meaning, dialect membership can also be represented through shared links with dialect-specific grammatical features. In other words, considering the Shetland dialect's *be-have* alternation, SSE verbs would link to a grammatical feature that specifies the *have* auxiliary in perfect constructions while Shetland verbs would link to a grammatical feature that specifies for the *be* auxiliary. These shared grammatical features can also serve as an organizing lexical principle, bundling words from the same dialect together.

The abstract dialect priming reported by Heggdal Lønes et al. (2023) supports this architecture. Our study shows that Scottish speakers increased their use of Scottish-specific items after processing different Scottish-specific items, providing the first demonstration that alignment is not reduced to item-specific priming but operates over the whole lexicon. For priming to occur at an abstract level, there must be some shared representation such that activation from one node can spread to other representations sharing that feature. Such priming has been reported over abstract semantic features, such as semantic category membership (Belke et al., 2005; Cirillo, Runnqvist, Strijkers, Nguyen, & Baus, 2022) and grammatical features (Melinger & Koenig, 2007). With the support of this finding, we argue that, while the dialectal variants do not occupy separate linguistic systems, they are nevertheless functionally separated via their connections with different conceptual and grammatical features.

4.3 What mechanisms are responsible for alignment at each level of representation?

We have presented evidence that speakers can align to the dialect of their interlocutor across phonetic, phonological, morphosyntactic, and lexical levels of representation. Moreover, these alignment processes can be strategic and/or automatic.

At the phonological and phonetic levels, the evidence we reviewed supports a social, strategic mechanism underlying alignment. Babel (2010) observed that alignment was sensitive to social motivation. Hernández (2009) found that younger Salvadoran speakers aligned to the local Mexican variant more than older speakers, ostensibly driven by a desire to align to their social peer group. Interestingly, phonetic alignment in Shetland seemed immune to strategic pressures and Smith and Durham (2012) argued this was because the th-stopping alternation fell outside the speakers' meta-linguistic awareness. They reasoned that speakers could only make strategic choices for those differences they are aware of. Instead, alignment on this phonetic variable showed evidence of an automatic mechanism, as all speakers produced both variants in their speech. The role of meta-linguistic awareness, or salience, may also explain why Walker and Campbell-Kibler (2015) found greater phonetic alignment between more distinct variants.

In contrast, at the lexical and morphosyntactic levels, we found little evidence that alignment was modulated by social factors. Indeed, none of the studies examining alignment at these levels reported any direct effects of social attitudes toward the interlocutors. Of course, an absence of evidence is not evidence and it is almost certainly true that strategic mechanisms can also operate over lexical and morphosyntactic variables. Indeed, Smith and Durham (2012) interpret their findings as evidence that alignment on salient linguistic variables is socially motivated (see also Labov, 2001). However, such a conclusion would be strengthened, for example, by evidence that preexisting social attitudes modulate lexical alignment rates (i.e., evidence of convergence and divergence across these levels of representation; Giles & Powesland, 1975).

We also found evidence for an automatic alignment process over lexical representations. Heggdal Lønes et al. (2023) showed abstract priming of dialect words. In two separate experiments, participants who processed Scottish dialect words subsequently produced more, different, Scottish words in comparison to their own baseline rate.Interestingly, in our first experiment, the increased tendency to produce Scottish words was not restricted to trials where the speaker was addressing a Scottish partner. Instead, the alignment was insensitive to the community membership of the partner, implying an absence of strategic, socially-motivated dialect production. In our second experiment, lexical choice was driven by exposure to Scottish words in the matching task. In the naming task, all participants were paired with an identical Scottish partner. Therefore, the lexical alignment observed here cannot be attributed to changing social motivations, as there were none. Instead, the difference can only be due to the priming manipulation conducted in the prior block. As the words produced by participants were different to those processed in the prior block, increased rates of Scottish words must be due to abstract, rather than lexical, priming. In the absence of a strategic explanation, the abstract primingmust be an automatic process. We argue that encountering the Scottish words in the prior block caused a spread of activation to other Scottish words in the participant's lexicon, leading them to produce different Scottish words later. For activation to spread in this manner, Scottish items in the lexicon must all be connected by an abstract dialect membership feature.

It should be emphasized that these findings are not exhaustive, and future studies may reveal a nuanced interplay of different mechanisms driving alignment at each representational level, depending on the circumstances or populations. Indeed, while none of the studies we reviewed found strong evidence for automatic phonological or phonetic priming, to our knowledge, no studies specifically investigating this have been carried out. Similarly, while there is no direct evidence for strategic lexical alignment, interview-based studies (Smith & Durham, 2012) do implicate it, so further experiments should be conducted. Finally, while our review focussed only on word production processes, similar questions can and should be asked about sentence-level processes (Weatherholtz, Campbell-Kibler, & Jaeger, 2014).

4.4 Proposing a model for bidialectal production

Fig. 5 provides a schematic model that captures both the conceptual and grammatical features we propose function to distinguish words from different dialects and serve as the mechanism that allows the observed behavioral patterns—namely, interference, dialect switching, cognate facilitation, automatic and strategic alignment—to operate. In this figure, we represent a conceptual feature activation pattern for the concept *kirk*. This activation pattern includes a dialect membership feature, denoted by the diamond. This dialect feature would also form part of the activation pattern for other dialect words, such as "hoose." At the lemma level, we have represented four lemmas that would likely be activated, to varying degrees, by this conceptual activation pattern. Each of these lemmas is associated with a set of grammatical features, some of which may vary according to the dialect. Here, we have represented the Scottish dialect specification for a singular distal demonstrative, "yon,", as well as the SSE variant, "that" (Smith & Durham, 2011). Correspondingly, the Scottish lemmas for "hoose" and "kirk" are both connected to the feature for "yon" and the English lemmas link to "that."

Shared conceptual features offer a mechanism to explain how conceptually-driven lexical selection processes can correctly retrieve words and grammatical features that match the intended dialect. It also explains why translation equivalent distractors compete for selection. Specifically, the addition of an extra conceptual feature renders them semantically similar but not identical. Finally, it also provides a mechanism over which abstract dialect priming can occur. A shared conceptual feature could enable activation from one dialect word to spread to other dialect words. However, it is not yet clear whether a dialect-specifying conceptual feature would bind

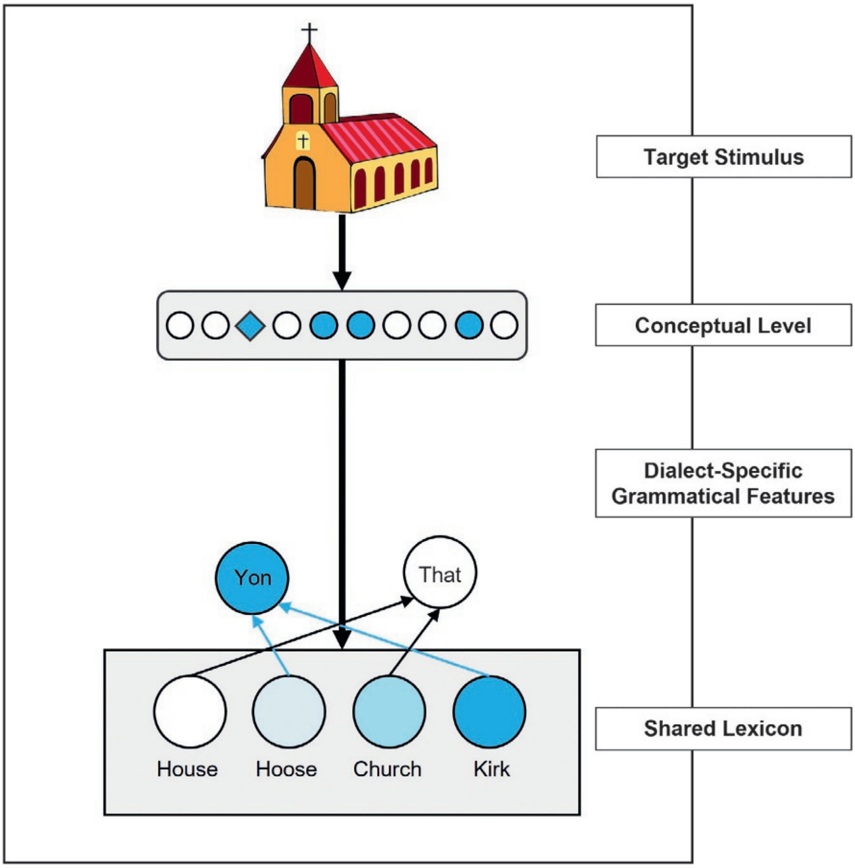

Fig. 5 Proposed model for the organization of the bidialectal language system, showing dialect feature nodes at the conceptual level which connect to all dialect items in the lexicon and dialect-specific grammatical features linked with grammatically-congruent dialectal lemmas.

words together more strongly than other abstract conceptual features such as animacy features or high level category features. In other words, it is speculative to suggest that a single feature at the conceptual level would be sufficient to explain the behavioral patterns reported in the literature, and further empirical work is required to test this proposal. We propose that clusters of dialect-specific grammatical features, which will be shared by all lemmas of a specific dialect and appropriate parts of speech, are another mechanism that supports the separation of representations within a single linguistic system. In some respects, they instantiate the co-occurrence patterns observed in the spoken language into abstract representation which can drive production processes.

5. Conclusion

Bidialectal speakers face many of the same processing challenges as bilinguals. They have distinct rules and representations for their different dialects, and they must be able to retrieve and produce the right forms for the right context. In this paper, we have reviewed the evidence from diverse literatures exploring how speakers select words from their distinct repertoires. We have examined the similarities and differences between bilingual and bidialectal word production and explored the mechanisms that influence dialectal choices made by speakers. Based on the evidence reviewed, we have proposed a model for dialect word selection, based on existing monolingual production models (e.g., Dell, 1986; Levelt et al., 1999), which posits that both dialect membership nodes at the conceptual level and dialect-specific grammatical features at the lemma level serve as representational markers that bind the repertoires together and that both strategic and automatic processes can operate over. As this is a novel proposal, it provides a framework for future research which will hopefully test its explanatory adequacy.

References

Abdel Rahman, R., & Melinger, A. (2008). Enhanced phonological facilitation and traces of concurrent word form activation in speech production: An object-naming study with multiple distractors. *Quarterly Journal of Experimental Psychology*, *61*(9), 1410–1440.

Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language & Cognitive Processes*, *24*(5), 713734. https://doi.org/10.1080/01690960802597250.

Abdel Rahman, R., & Melinger, A. (2019). Semantic processing during language production: An update of the swinging lexical network. *Language, Cognition and Neuroscience*, *34*(9), 1176–1192. https://doi.org/10.1080/23273798.2019.1599970.

Abutalebi, J., & Green, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language & Cognitive Processes*, *23*(4), 557582. https://doi.org/10.1080/01690960801920602.

Antoniou, K., Grohmann, K. K., Kambanaros, M., & Katsos, N. (2016). The effect of childhood bilectalism and multilingualism on executive control. *Cognition*, *149*, 18–30. https://doi.org/10.1016/j.cognition.2015.12.002.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*(4), 437–456. https://doi.org/10.1017/S0047404510000400.

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology Section A*, *58*(4), 667–692. https://doi.org/10.1080/02724980443000142.

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204. http://www.jstor.org/stable/4167516.

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2012). Effects of phonological feedback on the selection of syntax: Evidence from between-language syntactic priming. *Bilingualism: Language and Cognition*, *15*(3), 503–516. https://doi.org/10.1017/S1366728911000162.

Bortfeld, H., & Brennan, S. E. (1997). Use *and acquisition of idiomatic expressions in referring by native and non-native speakers*. *Discourse Processes*, *23*(2), 119–147. https://doi.org/10.1080/01638537709544986.

Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, *12*(2), 189–221. https://doi.org/10.1080/09541446.2000.10590222.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*(2), B13–B25. https://doi.org/10.1016/S0010-0277(99)00081-5.

Breining, B., Nozari, N., & Rapp, B. (2016). Does segmental overlap help or hurt? Evidence from blocked cyclic naming in spoken and written production. *Psychonomic Bulletin & Review*, *23*(2), 500–506. https://doi.org/10.3758/s13423-015-0900-x.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3–4), 155–180. https://doi.org/10.1159/000261913.

Campbell-Kibler, K., Walker, A., Elward, S., & Carmichael, K. (2014). Apparent time and network effects on long-term cross-dialect accommodation among college students. *University of Pennsylvania Working Papers in Linguistics*, *20*(2), 4.

Chen, Y., & Zhou, R. (2022). The mental lexicon features of the Hakka-Mandarin dialect bilingual. *Brain Sciences*, *12*(12). https://doi.org/10.3390/brainsci12121629.

Cholin, J. (2008). The mental syllabary in speech production: An integration of different approaches and domains. *Aphasiology*, *22*(11), 1127–1141. https://doi.org/10.1080/02687030701820352.

Cirillo, G., Runnqvist, E., Strijkers, K., Nguyen, N., & Baus, C. (2022). Conceptual alignment in a joint picture-naming task performed with a social robot. *Cognition*, *227*, 105213. https://doi.org/10.1016/j.cognition.2022.105213.

Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish-English and English-Spanish bilinguals. *Bilingualism: Language and Cognition*, *2*(3), 231–244. https://doi.org/10.1017/S1366728999000334.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, *41*(3), 365–397. https://doi.org/10.1006/jmla.1999.2651.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*(4), 491–511. https://doi.org/10.1016/j.jml.2004.02.002.

Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*(5), 1057–1074. https://doi.org/10.1037/0278-7393.32.5.1057.

Crinson, J., & Williamson, J. (2004). Non-standard dialect in the formal speech of 15-year-olds on tyneside. *Language and Education*, *18*(3), 207–219. https://doi.org/10.1080/09500780408666876.

De Bot, K. (1992). A bilingual processing model: Levelt's 'speaking' model adapted. *Applied Linguistics*, *13*, 1–24. https://doi.org/10.1093/applin/13.1.1.

Declerck, M., Ivanova, I., Grainger, J., & Duñabeitia, J. A. (2020). Are similar control processes implemented during single and dual language production? Evidence from switching between speech registers and languages. *Bilingualism: Language and Cognition*, *23*(3), 694–701. https://doi.org/10.1017/S1366728919000695.

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.
Dylman, A. S., & Barry, C. (2018). When having two names facilitates lexical selection: Similar results in the picture-word task from translation distractors in bilinguals and synonym distractors in monolinguals. *Cognition*, *171*, 151–171. https://doi.org/10.1016/j.cognition.2017.09.014.
Finkbeiner, M., Almeida, J., Janssen, N., & Caramazza, A. (2006). Lexical selection in bilingual speech production does not involve language suppression. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1075–1089. https://doi.org/10.1037/0278-7393.32.5.1075.
Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology. Human Perception and Performance*, *32*(5), 1276–1293. https://doi.org/10.1037/0096-1523.32.5.1276.
Gann, T. M., & Barr, D. J. (2014). Speaking from experience: Audience design as expert performance. *Language, Cognition and Neuroscience*, *29*(6), 744–760. https://doi.org/10.1080/01690965.2011.641388.
Giles, H., & Coupland, N. (1991). *Language: Contexts and consequences*. Thomson Brooks/Cole Publishing Co.
Giles, H., & Powesland, P. F. (1975). *Speech style and social evaluation*. Academic 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.
Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller, & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 159–207). Mouton de Gruyter.
Green, D. W. (1986). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language*, *27*(2), 210–223. https://doi.org/10.1016/0093-934x(86)90016-7.
Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81. https://doi.org/10.1017/S1366728998000133.
Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.), *Syntax and semantics, Vol. 3, Speech acts* (pp. 41–58). Academic Press.
Hall, M. L. (2011). Bilingual picture-word studies constrain theories of lexical selection. *Frontiers in Psychology*, *2*, 381. https://doi.org/10.3389/fpsyg.2011.00381.
Haugen, E. (1966). Linguistics and language planning. In W. Bright (Ed.), *Sociolinguistics* (pp. 50–71). Mouton.
Hazen, K. (2001). An introductory investigation into bidialectalism. *University of Pennsylvania Working Papers in Linguistics*, 7(3), 8.
Heggdal Lønes, E., Kamide, Y., & Melinger, A. (2023). *Dialect-specific grammatical features: Evidence from abstract priming across the lexicon*. Division of Psychology, School of Humanities, Social Sciences and Law, University of Dundee. Unpublished manuscript.
Hermans, D. (2004). Between-language identity effects in picture-word interference tasks: A challenge for language-nonspecific or language-specific models of lexical access? *International Journal of Bilingualism*, *8*(2), 115–125. https://doi.org/10.1177/13670069040080020101.
Hermans, D., Bongaerts, T., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, *1*(3), 213–229. https://doi.org/10.1017/S1366728998000364.
Hernández, J. E. (2002). Accommodation in a dialect contact situation. *Filología y Lingüística*, *28*(2), 93–110. https://doi.org/10.15517/rfl.v28i2.4491.
Hernández, J. E. (2009). Measuring rates of word-final nasal velarization: The effect of dialect contact on in-group and out-group exchanges. *Journal of SocioLinguistics*, *13*(5), 583–612. https://doi.org/10.1111/j.1467-9841.2009.00428.x.

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addresses. *Journal of Memory and Language*, *47*(4), 589–606. https://doi.org/10.1016/S0749-596X(02)00019-0.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*(3), 464–482. https://doi.org/10.1016/j.cognition.2005.02.006.

Hudson, R. A. (1996). *Sociolinguistics*. Cambridge University Press.

Ivanova, I., & Hernandez, D. C. (2021). Within-language lexical interference can be resolved in a similar way to between-language interference. *Cognition*, *214*, 104760. https://doi.org/10.1016/j.cognition.2021.104760.

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

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*, 201–258 (aWJML).

Kirk, N., Declerck, M., Kemp, R., & Kempe, V. (2022). Language control in regional dialect speakers—Monolingual by name, bilingual by nature? *Bilingualism: Language and Cognition*, *25*(3), 511–520. https://doi.org/10.1017/S1366728921000973.

Kirk, N. W., Kempe, V., Scott-Brown, K. C., Philipp, A., & Declerck, M. (2018). Can monolinguals be like bilinguals? Evidence from dialect switching. *Cognition*, *170*, 164–178. https://doi.org/10.1016/j.cognition.2017.10.001.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32. https://doi.org/10.1016/S0749-596X(02)00505-3.

Kuhlen, A. K., & Abdel Rahman, R. (2017). Having a task partner affects lexical retrieval: Spoken word production in shared task settings. *Cognition*, *166*, 94–106. https://doi.org/10.1016/j.cognition.2017.05.024.

Kühne, V., Rosenthal-von der Pütten, A. M., & Krämer, N. C. (2013). Using linguistic alignment to enhance learning experience with pedagogical agents: The special case of dialect. In *International workshop on intelligent virtual agents* (pp. 149–158). Springer.

La Heij, W. (2005). Selection processes in monolingual and bilingual lexical access. In J. F. Kroll, & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 289–307). Oxford University Press.

Labov, W. (1998). Coexistent systems in African-American English. In S. Mufwene, J. Rickford, J. Baugh, & G. Bailey (Eds.), *The structure of African-American English* (pp. 110–153). Routledge.

Labov, W. (2001). *Principles of Linguistic Change: 2*. Social Factors (Wiley-Blackwell).

Labov, W., Karen, M., & Miller, C. (1991). Near-mergers and the suspension of phonemic contrast. *Language Variation and Change*, *3*(1), 33–74. https://doi.org/10.1017/S0954394500000442.

Levelt, W. J. (1993). *Speaking: From intention to articulation*. MIT Press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences*, *22*(1), 1–75. https://doi.org/10.1017/s0140525x99001776.

Levitan, R., Beňuš, Š., Gravano, A., & Hirschberg, J. (2015). Acoustic-prosodic entrainment in Slovak, Spanish, English and Chinese: A cross-linguistic comparison. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (pp. 325–334). https://doi.org/10.18653/v1/w15-4644.

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

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.

Martin, C. D., Garcia, X., Potter, D., Melinger, A., & Costa, A. (2016). Holiday or vacation? The processing of variation in vocabulary across dialects. *Language, Cognition and Neuroscience*, *31*(3), 375–390. https://doi.org/10.1080/23273798.2015.1100750.

McNeill, D. (2000). Analogic/analytic representations and cross-linguistic differences in thinking for speaking. *Cognitive Linguistics*, *11*(1–2), 43–60. https://doi.org/10.1515/cogl.2001.010.

Melinger, A. (2018). Distinguishing languages from dialects: A litmus test using the picture-word interference task. *Cognition*, *172*, 73–88. https://doi.org/10.1016/j.cognition.2017.12.006.

Melinger, A. (2021). Do elevators compete with lifts? Selecting dialect alternatives. *Cognition*, *206*, 104471. https://doi.org/10.1016/j.cognition.2020.104471.

Melinger, A., & Koenig, J.-P. (2007). Part-of-speech persistence: Part-of-speech information as an organizing principle in the mental lexicon. *Journal of Memory and Language*, *56*, 472–489. https://doi.org/10.1016/j.jml.2006.12.001.

Melinger, A. (2023a). Identifying the boundary between dialects and languages: The role of similarity and usage. Manuscript in preparation.

Melinger, A. (2023b). Cross-dialect lexical alignment: Socially motivated or automatically primed? Manuscript in preparation.

Meuter, R. F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, *40*(1), 25–40. https://doi.org/10.1006/jmla.1998.2602.

Meyer, A. S., Roelofs, A., & Levelt, W. J. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language*, *48*(1), 131–147. https://doi.org/10.1016/S0749-596X(02)00509-0.

Monsell, S., Matthews, G. H., & Miller, D. C. (1992). Repetition of lexicalization across languages: A further test of the locus of priming. *The Quarterly Journal of Experimental Psychology Section A*, *44*(4), 763–783. https://doi.org/10.1080/14640749208401308.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*(4), 273–281.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*(3), 9–17. https://doi.org/10.1016/S0010-0277(97)00030-9.

Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology*, *4*(559), 1–5. https://doi.org/10.3389/fpsyg.2013.00559.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/s0140525x04000056.

Poulisse, N. (1997). Language production in bilinguals. In A. M. B. de Groot, & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (1st ed., pp. 201–204). Psychology Press.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics*, *15*(1), 36–57. https://doi.org/10.1093/applin/15.1.36.

Rabovsky, M., Schad, D. J., & Rahman, R. A. (2016). Language production is facilitated by semantic richness but inhibited by semantic density: Evidence from picture naming. *Cognition*, *146*, 240–244. https://doi.org/10.1016/j.cognition.2015.09.016.

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.

Rose, S. B., Aristei, S., Melinger, A., & Abdel Rahman, R. (2019). The closer they are, the more they interfere: Semantic similarity of word distractors increases competition in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 753–763. https://doi.org/10.1037/xlm0000592.

Ross, J., & Melinger, A. (2017). Bilingual advantage, bidialectal advantage or neither? Comparing performance across three tests of executive function in middle childhood. *Developmental Science*, *20*(4). https://doi.org/10.1111/desc.12405.
Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, *54*(2), 199–227. https://doi.org/10.1016/j.jml.2005.10.002.
Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*(1), 86–102.
Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, *72*(2), 111–123. https://doi.org/10.1016/S0010-0277(99)00024-4.
Slobin, D. I. (1987). Thinking for speaking. In *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society* (pp. 435–445). https://doi.org/10.3765/bls.v13i0.1826.
Slobin, D. I. (1996). From “thought and language” to “thinking for speaking”. In J. Gumperz, & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge University Press.
Smith, J., & Durham, M. (2011). A tipping point in dialect obsolescence? Change across the generations in Lerwick, Shetland 1. *Journal of SocioLinguistics*, *15*(2), 197–225. https://doi.org/10.1111/j.1467-9841.2011.00479.x.
Smith, J., & Durham, M. (2012). Bidialectalism or dialect death? Explaining generational change in the Shetland Islands, Scotland. *American Speech*, *87*(1), 57–88. https://doi.org/10.1215/00031283-1599959.
Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, *52*(3), 322–338. https://doi.org/10.1016/j.jml.2004.11.004.
Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, *60*(4), 487–501. https://doi.org/10.1016/j.jml.2009.01.001.
Tannen, D. (1989). *Talking voices: Repetition, dialogue, and imagery in conversational discourse*. Cambridge University Press.
Tobar-Henríquez, A., Rabagliati, H., & Branigan, H. P. (2021). Speakers extrapolate community-level knowledge from individual linguistic encounters. *Cognition*, *210*, 104602. https://doi.org/10.1016/j.cognition.2021.104602.
Trudgill, P. (2000). *Sociolinguistics: An introduction to language and society*. Penguin Books.
Trudgill, P., & Hannah, J. (2008). *International English: A guide to the varieties of standard English* (5th ed.). Routledge. https://doi.org/10.4324/9780203785225.
Vangsnes, Ø. A., Söderlund, G. B., & Blekesaune, M. (2017). The effect of bidialectal literacy on school achievement. *International Journal of Bilingual Education and Bilingualism*, *20*(3), 346–361. https://doi.org/10.1080/13670050.2015.1051507.
Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *48A*(4), 822–848. https://doi.org/10.1080/14640749508401419.
Von Der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S. H. (2010). “It doesn’t matter what you are!” explaining social effects of agents and avatars. *Computers in Human Behavior*, *26*(6), 1641–1650. https://doi.org/10.1016/j.chb.2010.06.012.
Walker, A., & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology*, *6*(546). https://doi.org/10.3389/fpsyg.2015.00546.

Wang, M., Cai, Z. G., Wang, R., Branigan, H. P., & Pickering, M. J. (2020). How do phonology and orthography feed back to influence syntactic encoding in language production? Evidence from structural priming in mandarin. *Quarterly Journal of Experimental Psychology*, *73*(11), 1807–1819. https://doi.org/10.1177/1747021820932125.

Washington, J. A., & Craig, H. K. (1998). Socioeconomic status and gender influences on children's dialectal variations. *Journal of Speech, Language, and Hearing Research*, *41*(3), 618–626. https://doi.org/10.1044/jslhr.4103.618.

Watt, D., Llamas, C., & Johnson, D. E. (2010). Levels of linguistic accommodation across a national border. *Journal of English Linguistics*, *38*(3), 270–289. https://doi.org/10.1177/0075424210373039.

Weatherholtz, K., Campbell-Kibler, K., & Jaeger, T. (2014). Socially-mediated syntactic alignment. *Language Variation and Change*, *26*(3), 387–420. https://doi.org/10.1017/S0954394514000155.

Wei, L. (2000). Dimensions of bilingualism. In L. Wei (Ed.), *The bilingualism reader* (pp. 3–22). London: Routledge.

Woutersen, M., Cox, A., Weltens, B., & De Bot, K. (1994). Lexical aspects of standard dialect bilingualism. *Applied PsychoLinguistics*, *15*(4), 447–473. https://doi.org/10.1017/S0142716400006871.

This page intentionally left blank

CHAPTER FIVE

Linking learning to language production

Amanda C. Kelley*

Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, United States
*Corresponding author: e-mail address: ackelle2@illinois.edu

Contents

Abstract

Statistical regularities are an important part of linguistic knowledge. This chapter examines this issue in-depth by reviewing the literature surrounding verb biases, or cases where a verb that can occur with multiple sentence structures is most likely to occur with only one. Findings in comprehension and production suggest that many kinds of language users rely on verb biases. Verb biases are acquired through distributional learning, which explains the ability to learn complex statistical relationships throughout the lifespan. This learning shares similarities with associative learning, which is highlighted by recent empirical work. Finally, multiple modeling approaches have contributed to our understanding of verb bias learning. These approaches are discussed alongside a series of connectionist models that replicate previously-discussed empirical findings.

All human languages create structure through regularity. Some statistical regularities are simple, like the fact that some words tend to be more frequent than others (e.g., Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). Others are more complex. Examples include the distribution of speech sounds like "ng" in English, which can occur at the beginning but not the end of a syllable. Learning and using these statistical regularities

Psychology of Learning and Motivation, Volume 78
ISSN 0079-7421
https://doi.org/10.1016/bs.plm.2023.03.001

is helpful for both comprehension and production. For instance, more-frequent words are recognized more quickly (e.g., Connine, Mullennix, Shernoff, & Yelen, 1990), and more-predictable sentences are easier to produce (e.g., Ferreira & Schotter, 2013). To effectively use these statistics, language users must learn them. This process is complicated by the fact that language varies considerably between different contexts. For example, whether a person expects to encounter a sentence about "giving them the book" or "giving the book to them" varies based on whether language is formal or informal, and whether it is spoken or written (e.g., Engel, Grafmiller, Rosseel, & Szmrecsanyi, 2022). To compensate for this complexity, language users continue to learn throughout their lifespan.

There are many examples of statistics that both affect language use online and are clearly updated through learning. To demonstrate the importance of synthesizing these viewpoints, this review will focus on verb subcategorization biases, or verb biases. All verbs have subcategorizations, or words and phrases that occur nearby to form a grammatical sentence. Many have more than one. For example, the sentence "I read" is complete—but so is the sentence "I read the book to the audience last night." Some verbs are equally likely to be followed by any subcategorization, but others tend to occur with only one subcategorization most of the time. For example, the verb "give" is used in prepositional dative sentences such as "I gave the apple to her" about as frequently as it is used in double object dative sentences, such as "I gave her the apple." Note that datives are distinguished by word order after the verb. Prepositional datives place the object before the recipient and use the preposition "to," while double object datives place the recipient before the object. Not all dative verbs behave like "give." For instance, the verb "throw" is much more likely to occur in prepositional dative sentences such as "I threw the ball to her." When a verb is preferentially used with only one subcategorization, the verb is said to be biased toward that structure.

Verb biases have been studied intensively as both an influence on language use and as an acquired statistic that is dynamically updated over time. This review will begin by summarizing how verb biases interact with other constraints in language comprehension and production. The scope of the scenarios where verb bias effects are found indirectly suggests that the learning mechanisms that govern it are complex. The second half of the chapter addresses the matter of learning mechanisms more directly by reviewing more recent work that shows rapid updating of verb biases, concluding with a series of models that can account for these effects. Both sections will address the benefits of understanding both where language statistics come from and what effects they have once they are learned.

1. Verb bias effects in language comprehension and production

Commonly, language statistics are first observed when they are found to affect either language comprehension or production. Verb bias affects both in many different populations: children and adults, first and second language speakers, neurotypical people and people with aphasia. Studies of these groups suggest that verb biases are learned easily and early, and may share learning mechanisms with other statistics that play an early role in language use.

1.1 Comprehension

Verb biases were originally studied as a factor in adult language comprehension, and their effects have been most fully described in this arena. Although not all studies have found verb bias effects in comprehension (Ferreira & Henderson, 1990; Kennison, 2001), the majority have suggested that they are one of many constraints that guide language comprehension (e.g., Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Lee, Lu, & Garnsey, 2013). This is true not only for typical adult native speakers, but also for children (e.g., Snedeker & Trueswell, 2004), people who have learned English as a second language (e.g., Dussias & Scaltz, 2008), and people with aphasia (e.g., DeDe, 2013). Reading or listening to sentences where a verb's bias matches the sentence's structure is usually facilitated for all these groups, with differences emerging when speakers must weight verb biases against other potential cues to structure. Comprehension studies suggest that verb-structure knowledge is detailed, including different biases for different senses of words. Language comprehension studies demonstrate the scope of verb bias knowledge and the breadth of its effects.

For typical adults who speak English as their first language, verb biases are one cue of many cues to sentence structure. In some early work on language comprehension, it was unclear if lexical items could be cues, or whether more complex structures would always be read more slowly than simple sentences (e.g., Frazier & Rayner, 1982). This perspective predicts that an association between a specific verb and a specific syntactic structure should not influence reading (Ferreira & Henderson, 1990). For example, take the following sentences, which have the same structure but use verbs that are biased toward different sentence types (bias norms from Garnsey et al., 1997):

a. The witness admitted (that) the fraud had started several years earlier. (Sentential Complement bias)

b. The witness confirmed (that) the statements he had made before were false. (Direct Object bias)

Sentences A and B are sentential complement sentences, where the main verb (e.g., "admit") is followed by a clause. Because this clause starts with a noun, sentences like these are initially ambiguous. They may be sentential complement sentences, or they may be a simpler direct object sentence such as, "The witness confirmed the statements." If verb biases are not used as a cue to sentence structure, then it should take about the same amount of time and effort to read both A and B because their sentence structure is equally complex. By contrast, if words can predict sentence structure, then verb biases should ease processing even in cases where the predicted structure is not simple (see Mitchell, 1989, for a contemporary review). In other words, if readers can use the information that "admit" tends to occur in sentential complement sentences, Sentence A should be read more quickly than Sentence B because the verb bias is a clue to the eventual sentence structure.

In the right experimental settings, verb biases immediately facilitate reading sentences where the verb and structure match (Garnsey et al., 1997; Trueswell, Tanenhaus, & Kello, 1993). Whether the second noun in the sentence is a plausible direct object—in other words, whether "The witness confirmed the statements" is also possible sentence—does not affect reading times in cases where a verb is strongly biased toward a sentential complement (Garnsey et al., 1997). However, Garnsey and colleagues note that comprehension is more difficult when the structures suggested by the second noun and the verb's bias conflict, suggesting that these two cues are both active up until the point when the structure of the sentence becomes clear. In conclusion, Garnsey and colleagues found that verb biases played an early and consistent role in sentence comprehension, with a smaller role of plausibility reserved for sentences with particularly ambiguous completion. These findings have generally been replicated in other contexts (e.g., Lee et al., 2013; Wilson & Garnsey, 2009), although not all published work finds verb bias effects (Kennison, 2001). Generally, verb biases can be understood as one cue to sentence structure, but which must be balanced against subtler cues such as the meaning of the next words in a sentence.

Children also use verb biases during language comprehension, but have more difficulty weighing them against other cues. Previous work has identified other types of temporarily ambiguous "garden path" sentences as particularly difficult for children to understand, and suggests that they may have more difficulty using contextual information to constrain what kinds of sentences they are likely to hear (e.g., Trueswell, Sekerina, Hill, & Logrip, 1999). Studies of verb bias use in children have confirmed that these same differences are present when their performance is compared to adults. For instance, Snedeker and Trueswell (2004) investigated how children comprehended ambiguous sentences similar to the sentence presented below:

c. Feel the frog with the feather

This sentence has two potential interpretations; the frog could possess the feather (a modifier interpretation), or the feather could be used to touch the frog (an instrument interpretation). Snedeker and Trueswell presented sentences with verbs that varied in their biases for each interpretation. While they listened to these sentences, participants viewed a display of objects that included a potential instrument (a feather), a target (a frog with a feather), and either the same animal (a frog with another object) or a different animal. Snedeker and Trueswell found that both 5-year-old children and adults used verb biases, but only adults were also able to use information from the scene to guide their interpretation of a sentence. For example, when a verb, like "feel" is instrument biased but the scene contains two frogs, adults were more likely to interpret "feather" as a modifier. Presumably, this is because adults realized that the speaker might be distinguishing between the two frogs using a modifier, even though the verb "feel" tends to describe actions that use an instrument. Children also seem to have difficulty discounting verb biases when plausibility suggests a different interpretation. Kidd, Stewart, and Serratrice (2011) examined sentences with instrument-biased verbs that were paired with instruments of varying plausibility. For example:

d. Cut the cake with the candle (Implausible)

Kidd et al. found that adults were able to somewhat overcome a verb's bias when the potential instrument was implausible, but that 5-year-olds were strongly affected by the verb's bias regardless of plausibility. Jointly, these results suggest that verb biases are available to children relatively early in development, but that the ability to use other cues at the same time develops later.

A similar pattern of competition between verb biases and other cues can be observed in speakers of English as a second language, which may be modulated by similarities between a speaker's first language and English. For example, Dussias and Scaltz (2008) compared the reading strategies of bilingual Spanish-English speakers and monolingual English speakers on sentential complement sentences like those used in Garnsey et al. (1997). Dussias and Scaltz found that bilingual Spanish-English speakers used verb biases similarly to monolingual English speakers if they had learned English-like verb biases. Grammatical similarities between first and second languages may allow language users to transfer strategies from one language to another. When reading sentential complement sentences, bilingual Korean-English speakers (Lee et al., 2013) and bilingual Mandarin-English speakers (Qian, Lee, Lu, & Garnsey, 2019) both had difficulty integrating multiple cues to structure in the way that native speakers do. However, while high-proficiency Korean-English speakers were better able to use cues than lower-proficiency speakers (Lee et al., 2013), the Mandarin-English groups showed sensitivity to verb biases and the presence of the complementizer "that" at all levels of English proficiency (Qian et al., 2019). One possible explanation is that Mandarin word order places verbs in the middle of sentences while Korean places words at the end, which may mean that the Mandarin-English bilingual group is more familiar with using verbs and complementizers as cues (Qian et al., 2019). The effects of transfer are most noticeable when the first language is in a different modality altogether. Anible et al. (2015) found that deaf American Sign Language (ASL)-English bilinguals use verb bias information in some situations, but generally prefer the more-complex sentential complement to the simpler direct object. Anible et al. suggest that this may be the result of parsing strategies that are tailored to visual languages like ASL. Despite varying first languages, second language speakers of English seem to generally acquire some knowledge about verb biases. These findings echo the developmental literature, suggesting that it is more difficult to hold cues like verb biases in mind than to learn them initially.

Verb biases remain an important cue for language understanding even in cases when other language functions are damaged. Several studies have addressed whether people with impaired language abilities due to brain damage (aphasia) still use verb biases during language comprehension. Although results are noisier than for healthy adults, Gahl and Menn (2016) conclude that studies of people with aphasia generally find that they rely heavily on verb biases, perhaps even more extensively than healthy adults. For instance, DeDe (2013) compared the performance of people with aphasia to healthy controls using a subset of sentences from Garnsey et al. (1997). Verb biases still facilitated reading when they matched the actual structure of a sentence, but people with aphasia did not show sensitivity to the presence or absence of "that" or the plausibility of the direct objects (DeDe, 2013). Again, it appears that verb biases are a relatively robust cue to sentence structure, while the ability to consider multiple cues simultaneously is a more taxing skill.

Many studies have focused on pairing a verb with only one structure. However, under the right circumstances, it is also possible to learn multiple biases for the same verb. Verb biases can change depending on the context in which a sentence is presented, particularly when a verb has multiple related meanings or "senses." Hare, McRae, and Elman (2003) collected verbs like "indicate" that change their biases between a direct object (indicating the door) and a sentential complement (indicating that there is a problem) based on which sense of a verb was suggested by context (see Hare, McRae, & Elman, 2004 for further discussion of how verb senses interact with structural biases in corpora). When sentences were presented with context that suggested a particular verb sense, Hare and colleagues found that reading was facilitated for that sense's verb bias. A similar result was found using subjects of sentences that bias participants to expect either a transitive or intransitive sense of the same verb (Hare, Elman, Tabaczynski, & McRae, 2009). Consequently, it appears that even within a single verb, different semantic senses can create a cue to different syntactic structures.

To summarize, verb biases are one of many constraints that are used during language comprehension. They are available relatively early in development (e.g., Snedeker & Trueswell, 2004), are learned by second language speakers of English (e.g., Dussias & Scaltz, 2008; Lee et al., 2013), and remain even when other linguistic abilities are damaged (Gahl & Menn, 2016). While many groups can use verb biases, only the most proficient groups seem to be able to consider other cues at the same time (e.g., Garnsey et al., 1997; Lee et al., 2013). The details of verb bias representations are also clearly tied to linguistic utility; language users maintain multiple representations per verb when it helps them more accurately understand a sentence (Hare et al., 2009, 2003). The range of verb bias effects seen in language comprehension suggests that knowledge of verb bias contributes to successful language use, and hints at a learning process that is capable of acquiring a complex range of statistics.

1.2 Production

In much the same way that verb biases are one cue to structure in language comprehension, they also influence language production. For adults, verb biases facilitate the production of sentences with structures that match the bias of the main verb (Ferreira & Schotter, 2013) and even change the articulatory features of these sentences (e.g., Gahl & Garnsey, 2004). Verb bias effects can be observed in children's production as early as 3-year-olds (Kidd, Lieven, & Tomasello, 2006; Peter, Chang, Pine, Blything, & Rowland, 2015). Jointly, these findings suggest that verb biases facilitate language production for common verb-structure pairings, and that this facilitation is present throughout the lifespan. These findings also extend to speakers of second languages, who may carry some knowledge of both their first and second language verb biases into second language production (e.g., Kootstra & Doedens, 2016). Language production offers further evidence that verb biases are important linguistic knowledge and include a diverse array of probabilistic relationships between verbs and structures.

Verb biases influence language production by encouraging the production of common verb-structure pairings. One line of evidence comes from studies that use sentence completion tasks to derive verb bias norms (e.g., Garnsey et al., 1997). To complete a sentence fragment containing a particular verb (e.g., "The farmer gave …"), one is essentially engaging in production of the completion (e.g., "the hay to the cows"). While sentence completion tasks are not the most sensitive way to measure production, they do suggest that the same statistics affect comprehension and production. Ferreira and Schotter (2013) address why verbs tend to be produced in their preferred structures, observing that including the optional complementizer "that" in sentential complement sentences can index production difficulty. Ferreira and Schotter found that "that" was produced more often with verbs that are biased against a sentential-complement continuation. In other words, when speakers produce a sentence like "She accepted (that) the book was unlikely to be published," they are more likely to include "that." This is because, as mentioned above, "accept" tends to be followed with a simple direct object noun phrase instead of an entire sentence complement. Additionally, producing or not producing "that" was not conditioned on the potential for a sentence to be ambiguous, but instead on how unlikely that verb was to be produced with a sentential complement. In summary, Ferreira and Schotter (2013) suggest that a verb bias toward a sentential complement reduces production difficulty for that sentence, while using a verb with any structure that is not its preferred structure will result in greater production difficulty. Verb biases are reflected in language production not only as the most common completion statistically, but also as the least-effortful completion to produce.

Evidence that verb biases facilitate specific structures is also seen in the articulatory properties of sentences. Using sentential complement and direct object sentences, Gahl and Garnsey (2004) examined phonetic aspects of production, including word duration and the deletion of t's or d's at the end of verbs, such as "maintaine**d**" or "conclude**d**." When the main verb's bias matched the sentence structure, /t-d/ deletion was more common, and the durations of pauses were often shorter. For a sentential complement-biased verb like "conclude," this means that "She concluded the review had been biased" is more likely to be produced without the final "d" in "conclude" and have a shorter pause between "conclude" and "the review." A similar effect was found in Gahl, Garnsey, Fisher, and Matzen (2006), which examined sentences with verbs that could either be transitive or intransitive. Tily et al. (2009) extended this method to spontaneous rather than recited speech. Tily and colleagues did not find a significant effect of verb bias, but suggest that this is because they included information like givenness and animacy in their models, which may have accounted for variance in structure choice that is normally accounted for by verb biases. Collectively, these studies suggest that production is sensitive to the probability that a word will occur in a particular sentence structure. Whether this probability is best explained by statistical co-occurrences or correlated factors that influence production remains an open question, but it is possible to say that verb biases can account for this data when other fine-grained predictors are not included.

Kidd et al. (2006) suggest that young children have fine-grained knowledge of how frequently verbs appear in particular structures. Using a sentence repetition task, Kidd et al. found that a group of children ranging from 2 years, 10 months to 4 years, 2 months were more likely to correctly repeat sentences containing verbs that are likely to appear in that sentence structure. Children were also more likely to substitute verbs biased toward a sentence's structure when they repeated it incorrectly—for instance, incorrectly using the sentential complement-biased "think" instead of "pretend" in a sentence such as "I pretend she is going to the beach." Kidd, Lieven, and Tomasello (2010) followed up on this work using a priming task with groups of 4-year-olds and 6-year-olds, who were primed with sentences that contained verbs that were either likely or unlikely given the sentence structure. Kidd et al. again found that children were more successful when they were asked to recall a sentence with a likely verb and that children tended to substitute likely verbs for unlikely verbs when they recalled incorrectly. Consequently, these studies suggest that children already show knowledge of verb biases in language production almost as early as their production can be reliably measured. In tandem with the comprehension literature, these results suggest that even young children use their knowledge of verb biases during language production.

Beyond verbs and structures that are naturally easier to produce together, experiments have also focused on what happens when structures are intentionally made easier to produce. One way to do this is through syntactic or structural priming, which refers to facilitated processing for recently-seen syntactic structures (Bock, 1986). Although priming can be measured in both comprehension and production, findings that manipulate verb bias tend to focus on how properties of prime sentences change sentence production. First-language speakers are primed more by uncommon verb-structure pairs (Bernolet & Hartsuiker, 2010). For example, if a verb was strongly biased toward the prepositional dative (throwing something to someone), seeing that verb in a double object sentence (throwing someone something) created stronger priming than seeing a verb that was biased toward the double object dative in the same sentence. This is an example of an effect of surprisal, or a greater response to unexpected stimuli. Surprisal effects are predicted by error-based implicit learning models of syntactic priming (e.g., Chang, Dell, & Bock, 2006). Verb biases also influence priming in children's language production. Peter et al. (2015) studied the performance of a group of 3- and 4-year-olds, a group of 5- and 6-year-olds, and a group of adults on a syntactic priming task. They found that both groups of children were more likely to produce the structure that a target verb was biased toward, and that they were more likely to repeat a structure when they were primed with a verb that was biased against that structure. In other words, Peter et al. found that children showed sensitivity to verb biases in language production in ways that are similar to the adults in their study and to the adult participants in Bernolet and Hartsuiker (2010). Other work suggests that executive function helps speakers overcome their tendency to produce dispreferred structures and verb-structure combinations (Thothathiri, Evans, & Poudel, 2017). Using an unexpected verb-structure combination in a prime makes it easier to say uncommon sentences, which under ordinary circumstances might require the use of additional planning resources such as executive function.

Bilingual speakers demonstrate knowledge of verb biases in their second language, although these effects vary in terms of how and when this information is used. For bilingual speakers, the effect of verb biases on syntactic priming is more variable, and studies focus on the likelihood of producing a structure that is congruent with a verb bias rather than the greater effect of uncommon verb-structure combinations in primes found in Bernolet and Hartsuiker (2010). Using a second-language sentence-completion task, Gries and Wulff (2005) found verb bias effects that are better explained by second-language than first-language verb biases. Similarly, Flett, Branigan, and Pickering (2013) found that German and Spanish speakers behaved similarly to English native speakers in an English priming task, suggesting that lexical and syntactic preferences from a first language did not affect second language production. However, studies of cross-language priming suggest a more complex picture. Salamoura and Williams (2006) examined priming from first-language Dutch to second-language English, and found that the biases of Dutch verbs influenced the structures produced in English. Kootstra and Doedens (2016) studied priming from both first-language Dutch to second-language English as well as the reverse. They found Dutch verb bias effects in English-to-Dutch priming, and effects of both English and Dutch verb biases in the Dutch-to-English condition. Kootstra and Doedens suggest that their findings may be influenced by a bilingual experimental context that examines production in both languages. Much like in language comprehension, second-language speakers show verb bias effects that are more varied than those of first-language speakers. Although the current literature offers insight into the sources of that variation, further work is needed to compare first and second-language speakers more fully.

In sum, the effects of verb biases in production parallel their effects in language comprehension. Saying sentences where the structure matches the verb bias is easier than when the structure does not match the bias (Ferreira & Schotter, 2013). Additionally, verb biases change the articulation when sentence structure and verb bias match (e.g., Gahl & Garnsey, 2004). Adults learn which structures and verbs commonly occur together, and this knowledge affects what they choose to produce and how they choose to produce it. Verb bias effects in production begin early, at least by age 3 (Kidd et al., 2006; Peter et al., 2015), and persist into adulthood, forming the basis for production effects documented in adults. Finally, bilingual speakers are influenced by verb biases when producing their second language, although these effects may be modulated by whether this production occurs in a monolingual or bilingual context (Gries & Wulff, 2005; Kootstra & Doedens, 2016). In language production, verb biases continue to play a role in guiding language use for many different kinds of speakers.

2. Learning verb biases

Beyond observing the effects of naturally-occurring verb biases, it is also possible to directly test hypotheses about how verb biases are learned. New verb biases can be learned rapidly in lab environments (e.g., Coyle & Kaschak, 2008) and begin to affect language use even at relatively short time frames (Thothathiri et al., 2017; Wonnacott, Newport, & Tanenhaus, 2008). Along with the findings detailed in previous sections, learning experiments bring the mechanisms of verb bias learning into focus. Any potential mechanism must be able to explain verb bias effects in young children (e.g., Peter et al., 2015; Snedeker & Trueswell, 2004), and must account for not only the acquisition of simple verb-structure associations (e.g., Garnsey et al., 1997) but also the flexibility needed for those associations to vary with context and verb meaning (e.g., Hare et al., 2003). This article will suggest that an associative learning mechanism neatly fits these requirements. Associative learning refers to learning that slowly updates the connection strength between two things—in this case, between a verb and a structure (e.g., Wasserman & Miller, 1997). This section will summarize the learning findings that such a mechanism must explain, along with how these findings can be connected to studies of learning in areas outside of verb biases.

2.1 Experimental investigations of verb bias learning

Verb bias learning mechanisms must explain the presence of verb bias effects in young children and the rapid acquisition of new verb biases by all ages of learners. Establishing an early mechanism is key because children show verb bias effects in comprehension and production by ages three to four (e.g., Kidd et al., 2011; Peter et al., 2015; Snedeker & Trueswell, 2004). Because this knowledge is already adult-like in young children, it is likely that it was acquired even earlier (e.g., Fisher, Jin, & Scott, 2020). This mechanism should also be able to account for learning of new verb biases by both children and adults (e.g., Coyle & Kaschak, 2008; Qi, Yuan, & Fisher, 2011). This section will elaborate on the connection between distributional learning in young children and novel verb bias learning in children and adults.

Babies begin to learn about linguistic statistics from a very early age. This is likely done through distributional learning, a general learning mechanism that allows children to keep track of statistical regularities that they experience (see Gómez & Gerken, 2000 for further review). Distributional learning is particularly helpful for tracking repeating sequential stimuli, which includes information like the words that tend to immediately surround verbs (see Romberg & Saffran, 2010 for further discussion). Children's language use suggests that they apply this learning to naturalistic input relevant to learning verb biases; for instance, they tend to use individual verbs in the same kinds of sentences as their caregivers (Theakston, Lieven, Pine, & Rowland, 2001). Distributional learning may also allow children to combine evidence from multiple types of sentences to narrow down the structural biases of individual verbs (e.g., Twomey, Chang, & Ambridge, 2014).

Experimental work has further narrowed down how children apply distributional learning to verbs. Children learn distributionally about entire groups of verbs, which is related to learning biases for individual verbs. For example, Scott and Fisher (2009) investigated the case of learning distributions for two classes of verbs: causal verbs that highlight changes to an object, and unspecified-object verbs that highlight the action being done. Both types of verbs can occur in transitive sentences, such as e., but alternate to a different intransitive structure as in f. and g:

e. Anne broke/dusted the lamp.

f. The lamp broke. (Causal)

g. Anne dusted. (Unspecified-Object)

By examining verbs in a corpus of child-directed speech, Scott and Fisher found that verbs like "dust" and "break" vary in distributional parameters such as whether their intransitive form takes an animate agent, which could be tracked to differentiate these verb classes. In a separate experiment, 28-month-old children listened to a dialogue that used those distributional parameters to indicate which class a novel verb belonged to. When presented with an ambiguous transitive sentence and either a causal or unspecified-object video, Scott and Fisher found that children tended to look at the video that was congruent with the distribution they had previously heard used with that verb. This suggests that children track relevant parameters that allow them to assign verbs to a correct class, including which syntactic environments they have experienced a verb in. Further work also suggests that 2-year-olds apply this same kind of learning to other verb classes, such as transitive and intransitive verbs (Yuan & Fisher, 2009), and that children may use other cues such as discourse structure to further organize the input they receive to learn about verb behavior (see Fisher et al., 2020 for a comprehensive review of how children learn verb distributions).

Adults and slightly older children can learn new verb biases through distributional learning. Coyle and Kaschak (2008) asked adult native speakers of English to complete a series of dative sentences that always contained only one verb. For example, the verb "send" might occur in stems that were usually completed as double object dative sentences (e.g., I sent her [the package]), and the verb "hand" might only appear in prepositional dative stems (e.g., I sent a package [to her]). After this training, Coyle and Kaschak found that their participants were more likely to produce verbs in the structure they had just been experienced in, suggesting that this training had changed their verb biases. In a comprehension task, Ryskin, Qi, Duff, and Brown-Schmidt (2017) trained participants by showing them trials where verb-structure pairings were accompanied by events, and found that participants used these new verb biases to interpret ambiguous sentences (see also: Ryskin, Qi, Covington, Duff, & Brown-Schmidt, 2018). Children are also able to update verb biases through both comprehending and producing skewed sets of training sentences. Qi et al. (2011) trained 5-year-old children using dialogues that either encouraged an instrument or a modifier interpretation of a verb. Like in Snedeker and Trueswell (2004), children were presented with sentences like "Feel the frog with the feather," where the feather could be either an instrument used to feel a frog, or a modifier used to describe a frog. Qi et al. (2011) found that children began to look toward the interpretation of that verb that they were newly biased toward, suggesting that they had changed their biases. Lin and Fisher (2017) extended these findings with 4- and 5-year-olds, finding that children not only learned new verb biases in production, but also that training children toward an unlikely structure given the verb's pre-existing bias increased the size of the training effect, which was further modulated by how common a particular syntactic structure is. This result suggests that children track the contexts of individual lexical items as well as the relative frequency of the syntactic structures themselves. Lin and Fisher also link this finding to the surprisal effects found in Bernolet and Hartsuiker (2010), which they note are predicted by models that assume an error-based learning mechanism (e.g., Chang et al., 2006). Children and adults update verb biases through distributional learning, and children likely begin to do so before even the youngest groups that have demonstrated verb bias effects.

Some researchers have turned to artificial language learning to exert even greater control over the distributional properties of verbs. Naturalistic verb biases already suggest that learners acquire multiple kinds of statistical relationships between verbs and structures, and artificial verb biases add further support to this conclusion. Wonnacott et al. (2008) trained adults using novel sentence structures and verbs, which appeared in artificial languages that varied in whether verb-specific biases or language-general patterns were better predictors of sentence structure. Wonnacott et al. found that adults learned new verb biases, which they demonstrated in production, grammaticality judgments, and eye movements, and generalized to verbs that they had not seen in training. Importantly, adults also learned when verb biases were unreliable predictors of structure, and then instead used language-general patterns to guide language production. This finding also generalizes to children. Wonnacott (2011) presented 5- to 7-year-old children with noun-particle word associations similar to the patterns used in Wonnacott et al. (2008). Again, learners of the lexically-specific language learned noun biases and replicated them in generalization, while learners of the generalist language acquired the general frequency of each particle. Distributional language learning involves not only learning about statistical relationships, but also about how useful those statistics are for language use.

Distributional learning also helps constrain verb use by combining multiple independent sources of evidence. One of these sources involves tracking multiple distributions at once. Twomey, Chang, and Ambridge (2016) examined verbs that can occur in locative structures (e.g., She filled the cup with water), but which also occur in transitive structures that are unique to that type of locative (e.g., She filled the cup vs She poured the water). Twomey and colleagues found that adults are sensitive to these distributions, preferentially using novel verbs with one locative structure when they had heard the verbs previously in the corresponding transitive construction. This suggests that adults use multiple distributions of lexical cues to determine the correct verb biases for a word. Perek and Goldberg (2017) examined cases of verbs that can only appear in one type of construction. They presented participants with one language where verbs alternated between two meaningful constructions, and a similar language with one verb that appeared in only one of the constructions. Perek and Goldberg showed that adults used a majority of the verbs based on the meaning of the construction, but were conservative with the verb that appeared in only one structure. Adults are not only able to learn from lexical distributions of verb-structure co-occurrences, but also from related distributions and negative feedback.

Further work suggests that meaning is also used as a source of information to constrain distributional learning. First, there is evidence that meaning may cause speakers to choose one syntactic structure over another. In an artificial language, Perek and Goldberg (2015) found that when novel structures are associated with a specific meaning, meaning competes with the distribution of verb-structure pairings to determine what structure participants choose. Thothathiri and Rattinger (2016) extended this result by manipulating whether a verb cue or a semantic-context cue was a better indicator of syntactic structure during comprehension of an artificial language, and found that the more valid cue tended to determine which structure a speaker would choose when "producing" this language. This kind of learning can also be induced in a speaker's native language under the right circumstances. When a familiar English dative structure becomes more strongly indicative of meaning than the bias of a particular verb, speakers use meaning rather than verb bias to guide their structural choices when speaking (Thothathiri & Braiuca, 2021). For instance, if the double object dative is repeatedly paired with "completed" transfer events and the majority of verbs occur in both dative alternatives, Thothathiri and Braiuca report that speakers choose to use the double object dative to describe "completed" transfer events. Jointly, these findings suggest that the kinds of meaning that sentences describe can actually be a conditioning environment for particular syntactic structures, and can outcompete lexically-dependent cues (e.g., verb biases) if they provide a stronger cue to structure. Finally, speakers learn about clusters of verbs with similar meanings. For example, participants rated ungrammatical sentences like "She tumbled him" as more acceptable than sentences like "She laughed him," because the semantics of "tumble" are more like verbs that can occur in a transitive sentence (Ambridge, Pine, Rowland, & Young, 2008). Ambridge and colleagues also found that participants assume that novel verbs that describe falling are constrained to the same structures as known verbs of this class. Meaning limits and informs distributional learning, preventing overgeneralization and providing an alternative source of information.

Verb biases are thus acquired through distributional learning, which begins in childhood and continues throughout adult life. While verb biases are often operationalized as a verb-structure co-occurrence, their reality is somewhat more complex. These statistics are clearly created through the long-term storage of many syntactic and semantic contexts, which creates observable facilitation during language comprehension and production. Verb biases are both a result of language experience, and a strategy for successfully predicting the language use of others.

2.2 Characterizing the mechanisms of verb bias learning

For verb bias learning, there are many studies that examine how different distributions affect language use. There are fewer studies that try to directly characterize verb bias learning. However, these studies have suggested that verb bias learning is both incremental and error-based. "Incremental" refers to learning that is updated a small amount during each learning event and accumulates relatively slowly over time. As mentioned previously, error-based learning means that more learning occurs when a system's predictions are more incorrect (e.g., Chang et al., 2006). Diagnosing these types of learning is not only a problem for studying verb biases, and the more general learning literature has developed paradigms that can be used to better understand learning mechanisms.

Since the associations in incremental learning are updated one trial at a time, they are also unlearned in a trial-by-trial manner. This insight leads to a paradigm called reversal learning. In reversal learning, participants are trained on a rule, often to a specific level of performance, and then are trained on the opposite of that rule (Izquierdo, Brigman, Radke, Rudebeck, & Holmes, 2017). Rules vary depending on the species and task but could include discriminating between two different shades of gray (Hoffmann, Perkins, & Calvin, 1956), learning to turn right or left in a maze (McDaniel, 1969), or even learning categories of subtly different Gabor patches (Cantwell, Crossley, & Ashby, 2015). Recovery from the reversal may have one of two outcomes. In fast reversal, learning from the initial rule allows participants to learn what dimensions of a particular problem are important, and when the rule is reversed, participants then quickly learn to use those same dimensions to make an opposite response (Sanders, 1971; see Kruschke, 1996, for review). For instance, human participants in the gray-discrimination task might realize that they are rewarded for choosing a particular shade of gray, and would rapidly switch to choosing the other shade when the reward structure changes. Another possible outcome is a slow reversal. Under these conditions, previous learning does not improve learning the reversed rule; instead, it is common for participants to only gradually begin to give the opposite response (e.g., Hoffmann et al., 1956). The rats in Hoffmann, Perkins, and Calvin consistently took more than twice as many trials to learn a reversed gray discrimination, suggesting that their earlier learning may have actually impeded reversal. The "fast" and "slow" in these situations refers to how quickly participants reach a particular level of performance—in the first case, more quickly than the first time the rule was learned, and in the second, more slowly.

Converging evidence from neuroscience and behavioral studies can help further characterize learning by explaining developmental and evolutionary gradients in reversal behavior. Behavioral work shows two overall trends. When their performance is compared on the same task, children are more likely to reverse slowly than adults (e.g., Kendler & Kendler, 1970). Second, slow reversal is more common among animals of lower order taxa; for instance, rats are more likely to reverse slowly than humans (Sanders, 1971). These findings are complemented by findings from neuroscience. Reversal learning is impaired in patients with ventromedial prefrontal cortex damage (Fellows & Farah, 2003). These areas support the representations that allow learners to compare trials and flexibly shift strategies based on aspects of those sets of trials (Izquierdo et al., 2017; Murray & Gaffan, 2006). If prefrontal areas are impaired, then reversal learning is handled by subcortical structures like the basal ganglia, which requires multiple trials to inhibit previously learned, habitual responses (Frank & Claus, 2006). Consequently, fast reversal occurs when a learner is able to represent a rule and compare across trials (Izquierdo et al., 2017). This is easier for adults than children, and for humans than for animals. Slow reversal occurs when trials cannot be easily represented and compared, and learning is handled by subcortical structures that require multiple trials to re-learn reward structures (Frank & Claus, 2006). These results also allow us to better understand why there are two different kinds of reversal learning. Fast learning is characteristic of the cognitive flexibility derived from the ability to easily compare trials, while slow learning is the result of incrementally unlearning habits on a trial-by-trial basis.

Linking verb bias learning to other reversal learning studies helps determine whether verb bias learning is incremental or not. Using a paradigm like Coyle and Kaschak (2008), Kelley (2019) found incremental verb bias learning in language production. Participants learned new biases for six verbs, three transitive and three dative. Two verbs were biased toward new structures, while one verb appeared an equal number of times in each of its syntactic alternates. For instance, the dative verb "give" would be biased toward the double object dative structure, the verb “hand” would be biased toward the prepositional dative, and the verb “show” would appear in both dative structures an equal number of times. Transitive verbs were theme-experiencer verbs like “surprise,” which are more likely to alternate between active and passive structures (Ferreira, 1994). The dative verbs selected also alternate regularly between the double object dative and prepositional dative. In addition, this study was made up of two blocks, which allowed the associations between verbs and structures to be reversed in the second block. If a participant learned to produce “give” in the double object dative in the first block, they would then learn to produce “give” in the prepositional dative in the second block. As can be seen in Fig. 1, learning was less successful in the second block than the first block for dative verbs, while transitive verbs showed no evidence of learning. Thus, Kelley (2019) provides evidence for slow reversal in dative verb bias learning, where reversal performance is less successful than initial learning. This is likely due to an inability to use attention to detect relevant dimensions

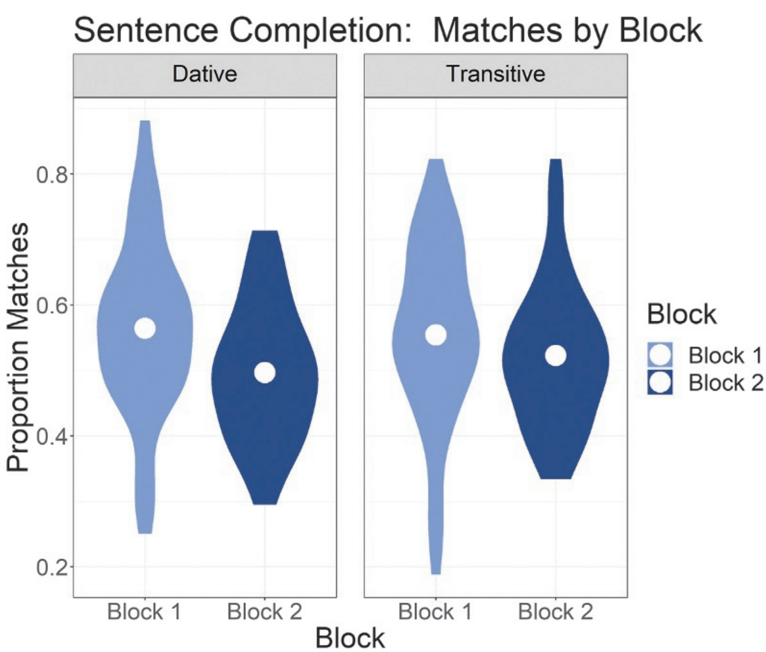

Fig. 1 Behavioral results from Experiment 2 of Kelley (2019). Participants learned more successfully in the first block of dative verb training than the second. However, there was no difference between the two blocks of transitive verb training.

of the stimuli, either due to the complexity of tracking the changing verb biases across multiple trials, or because language production encourages implicit learning.

Distributional learning of verb biases shares similar features with other examples of associative learning. In both cases, statistics are being learned and learning is updated when new information is experienced. As a result, it is possible to apply associative learning paradigms to better understand the distributional learning of verb biases. Kelley (2019) used reversal learning to demonstrate that verb bias distributions are learned incrementally in language production. This finding links distributional learning of verb biases to other types of learning, suggesting a similar mechanism in both cases.

3. Modeling of verb bias use and acquisition

Given the complexity of verb distributions that language users learn and use, another common approach to understanding behavior is cognitive modeling. Multiple architectures, including Bayesian models (e.g., Perfors, Tenenbaum, & Wonnacott, 2010) and neural networks (Ambridge & Blything, 2016), have been used to characterize aspects of verb learning. These models also focus on different aspects of learning about verbs, from the initial discovery of verb classes (Perfors et al., 2010) to understanding what learned factors govern structure selection (Bresnan, Cueni, Nikitina, & Baayen, 2007). Kelley (2022) contributes to this literature by describing a relatively simple set of models that directly address several of the behavioral findings summarized in previous sections. These models unite insights about the incrementality of verb bias learning mechanisms with behavioral findings.

Some models focus on determining what factors in input lead speakers to choose a particular structure. Semantic and discourse-level factors can predict what a speaker will choose to produce; for example, when a noun is “given,” or had been mentioned previously, it is more likely to be said before any non-given nouns (Bresnan et al., 2007). Using features like these in combination with information about the main verb of a sentence, Bresnan and colleagues generated logistic models to predict which dative a sentence would use. After accounting for contextual differences, this model could characterize data from both the spoken Switchboard corpus and the written Wall Street Journal corpus. More recently, this work was generalized to a much larger set of dative judgments using modern language models (Hawkins, Yamakoshi, Griffiths, & Goldberg, 2020). In this study, acceptability judgments from a larger set of verbs were predicted using models that were trained on very large corpora. Hawkins and colleagues found that these models were sensitive to the biases of dative verbs, and that they could predict dative structure about as well as the logistic models in Bresnan et al. (2007). These models do not use hand-selected features like the models used by Bresnan and colleagues, and instead use extensive training and a complex architecture to find features in a corpus. Models are able to use features in written and spoken language to learn how verbs tend to behave.

Another way of characterizing distributional learning is through Bayesian models. These models use Bayes' theorem to update an initial probability distribution, or prior, based on new data. This updated distribution, called the posterior, can then be analyzed to understand what the model has learned. This process can occur at multiple levels—for example, a verb and a particular structure could both have their own distributions—which allows for explicit modeling of the multiple levels of learning (e.g., Barak, Fazly, & Stevenson, 2014; Perfors et al., 2010). Verb biases can be learned as one of these levels; however, the goal of the entire model is to understand what kinds of information need to be collected to successfully model human verb learning and generalization.

Bayesian modeling suggests that verb classes are a necessary component of replicating empirical findings. A hierarchical Bayesian model with three layers can replicate the results of Wonnacott et al. (2008), learning either verb-specific biases or general statistics about verb classes depending on how frequent alternation is in that verb class (Perfors et al., 2010). With the added ability to cluster verbs into classes, this model also learns to differentiate alternating and non-alternating dative verbs; its ability to correctly generalize these verbs increases as it receives more input. Finally, the addition of semantic features to the input further prevents overgeneralization. While the input in Perfors et al. (2010) was limited to dative structures, models that create verb classes can also learn the dative alternation from input that mixes dative structures and other kinds of structures (Parisien & Stevenson, 2010). Bayesian models can also be used to model how verb generalization changes over development. Models that update verb clusters incrementally show that the ability to generalize develops over time as the model gains more knowledge about general verb classes and can move past biases in its input (Barak et al., 2014). Bayesian models offer insight into the levels of abstraction at which humans collect statistics, and can explicitly test whether these levels are necessary to replicate experimental findings. In particular, they underscore the need for a level of abstraction that gathers similar verbs and learns about the behavior of the entire class.

By contrast, connectionist approaches have not produced the same kind of comprehensive models of verb learning as Bayesian approaches. However, connectionist approaches have recreated aspects of verb learning, and have provided insights into the kinds of architectures that can and cannot reproduce previous experimental results. Language models based on connectionist or, neural, nets use arrays of artificial neurons that collect activation and then pass it along to other artificial neurons through adjustable weighted connections. Different arrangements of weights and neurons yield nets with different kinds of behaviors that are suited for different kinds of tasks. For example, feed-forward networks simply pass information "forward" from an input vector to an output layer, and may be used to model processes like selecting one construction from several potential options (e.g., Ambridge & Blything, 2016). By contrast, more complex architectures allow models to complete more complicated tasks, like predicting the next word in a sequence (e.g., Chang, 2002; Chang et al., 2006; Elman, 1990). Models like these offer insights into how verb bias learning might proceed with less prior structure than is typically found in Bayesian models.

There is no reason why a connectionist model cannot learn to associate a verb with a particular structure, but some connectionist architectures, surprisingly, may prevent learning verb biases. For example, the Dual-Path model replicates many findings related to syntactic priming (Chang et al., 2006). However, because it separates syntax from meaning, this model does not learn syntactic information associated with particular words even though it implements a learning mechanism that, as will be described later, can explain verb bias learning. Despite its difficulty learning verb biases, the Dual-Path model can acquire verb classes based on syntactic alternations, such as the English locative (e.g., She filled the cup *with water* vs She poured water *into the glass*) (Twomey et al., 2014). In that study, the model was able to learn five different classes of locative verbs with varying levels of bias toward each of the locative structures. A simpler connectionist model can learn dative preferences for individual verbs in a human-like way, demonstrating early overgeneralization and naturally exhibiting behaviors like preemption (Ambridge & Blything, 2016). Preemption occurs when there are multiple possible constructions that could be used, but one is used preferentially and consequently blocks the usage of the other alternative (Goldberg, 1995; Brooks & Tomasello, 1999). One example of this is the past tense of the verb "go," which could be "goed," but is blocked by the semantically-identical form "went" (Goldberg, 1995). Although this model does not perform as well as a comparable Bayesian model when presented with the same information (Barak, Goldberg, & Stevenson, 2016), it also uses a much simpler implementation than typical Bayesian models (Ambridge & Blything, 2016). Consequently, although connectionist models do not offer the same comprehensive solutions as Bayesian models, their simpler mechanisms offer a more parsimonious account of learning.

Other findings that are good candidates for connectionist modeling have not been directly addressed by other models. These include the dative incremental learning and lack of passive verb bias learning reported in Kelley (2019), as well as the surprisal effects found in Lin and Fisher (2017). The following three-layer feed-forward model is extremely simple, but can account for these findings. The key is that its learning mechanism is incremental and error-based. As seen in Fig. 2, the model receives a vector that represents a single verb as input. This information is passed to a three-node hidden layer, and then to a two-node output layer, where each node

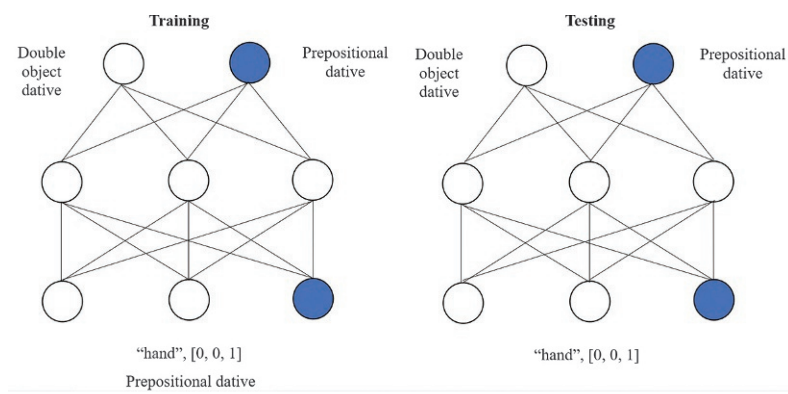

Fig. 2 An example of a model from Kelley (2022). During training, the model accepts vector representations of verbs such as "hand" as input, which it learns to associate with specific outputs representing sentence structures such as the prepositional dative. When the model's learning is tested, it is given only a input vector, and is asked to generate the most likely output. In this case, the model has learned that "hand" is most likely to occur in the prepositional dative after training.

represents a syntactic structure. Using this architecture, the model can learn associations between input verbs and output structures. To model the reversal-learning results of Kelley (2019), the model first learns an association between a particular verb and structure, and then reverses that association. For example, "hand" might first appear only in the prepositional dative, and then only in the double object dative. Under these conditions, the model learns the verb bias less successfully during the reversal block. Despite its simplicity, the model clearly replicates the pattern of behavioral results (Fig. 3).

With a slightly more complex input, this model also accounts for surprisal effects. It demonstrates more learning when it encounters an unexpected verb-structure pairing, much like the behavioral results found in Bernolet and Hartsuiker (2010) or Lin and Fisher (2017). Lin and Fisher (2017) selected verbs with different biases toward dative structures and then trained participants to learn new biases for these verbs. They found that more verb bias learning occurs when participants are presented with verbs in structures that are against their bias. This effect is also modulated by the

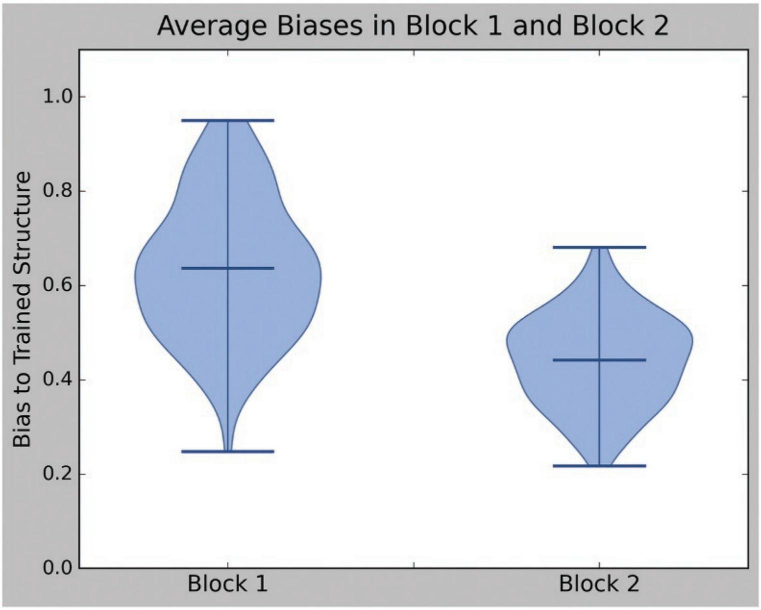

Fig. 3 Average biases obtained by models trained using a paradigm adapted from the behavioral experiment in Kelley (2019).

frequency of a syntactic structure. For example, more learning occurs when a verb is biased toward the overall less-common double object dative than when it is biased toward the overall more-common prepositional dative. These two effects interact, such that learning about a dispreferred (and overall less common) double object structure causes the greatest amount of learning, while biasing a verb toward a preferred (and overall more common) prepositional dative structure creates very little learning. To model this, Kelley (2022) first trained the model so that different verb inputs had different baseline "biases"—for example, "throw" might only appear in the prepositional dative, while "give" would appear in both datives half the time. Then, the verbs were trained toward either the structure they preferred—for example, "throw" would be trained further toward the prepositional dative—or toward the structure they dispreferred. Like Lin and Fisher (2017), Kelley (2022) found that the model's verb biases changed the most when a verb was unexpectedly trained toward an uncommon structure (Fig. 4).

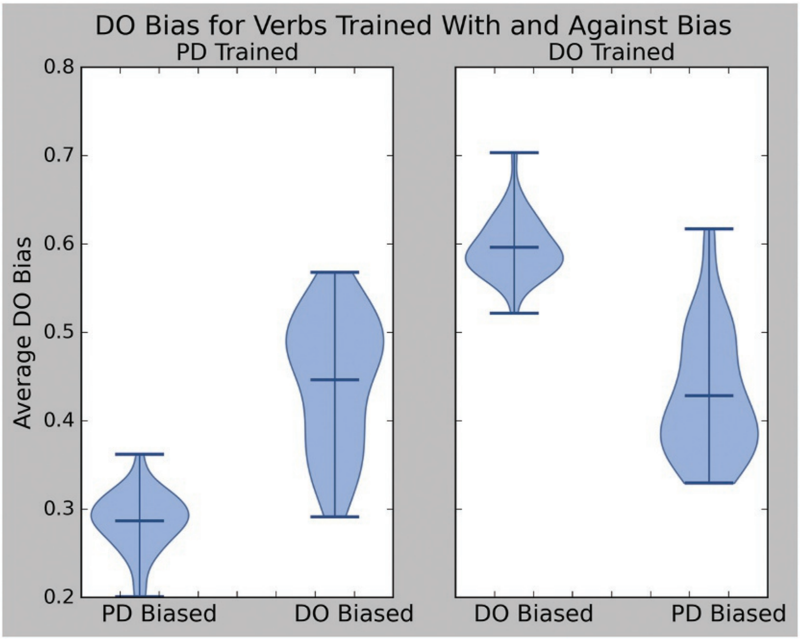

Fig. 4 Average biases obtained from models trained using a paradigm adapted from Lin and Fisher (2017). The pattern of model result closely resembles their behavioral results.

Finally, the model can also account for the lack of verb bias learning for transitive verbs. Kelley (2019) found that transitive verbs do not acquire new verb biases through training, a finding that was replicated twice in Kelley (2022). If it is given the correct input, the model can learn biases for a class of dative verbs like "give," while also not learning biases for transitive verbs like "surprise" or "chop." To simulate this, the model received training about how likely a dative verb was to appear in either the double object or prepositional dative. The model also received information about transitive verbs, which could appear in an active sentence like "She surprised the dog," or a passive sentence like "The dog was surprised by her." Consequently, this version of the model had four possible outputs, one for each possible structure. Along with information about which structure the verb appeared in, the input also indicated whether a sentence was patient-first. Sentences like "She gave a ball to the dog" or "She surprised the dog" both place the thing doing an action—the agent—in the first position in the sentence, and the patient—the thing that is acted upon—later on in the sentence. Only passive sentences like "The dog was surprised by her" place the patient in the first position. Therefore, use of a passive sentence was perfectly predicted by the patient-first cue. After this initial training was done, the model was trained to learn new biases, much as in a behavioral experiment. For example, a verb like "give" might be pre-trained to occur in each dative structure half the time, but would only occur in the prepositional dative during the second phase of experimental training. As can be seen in Fig. 5, the dative verbs acquired a new verb bias after experimental training, while the transitive verbs did not. Further work showed that it was the pre-experimental training in which the model learned from "life experience" how structural cues predict sentence structure, that blocked the model from learning a new verb bias with transitive verbs. In essence, for actives and passives, the model had learned that the verb is not a particularly useful cue, and thus learning a new verb bias in the laboratory was blocked. With the right kinds of training and input, the model presented in Kelley (2022) was able to account for the behavioral data.

Thus, many different models have contributed to explaining how humans learn about verbs. This includes determining what kinds of features may be learned from natural-language input (e.g., Bresnan et al., 2007) and what kind of structure a model must be capable of learning (e.g., Perfors et al., 2010). The models in Kelley (2022) contribute by directly

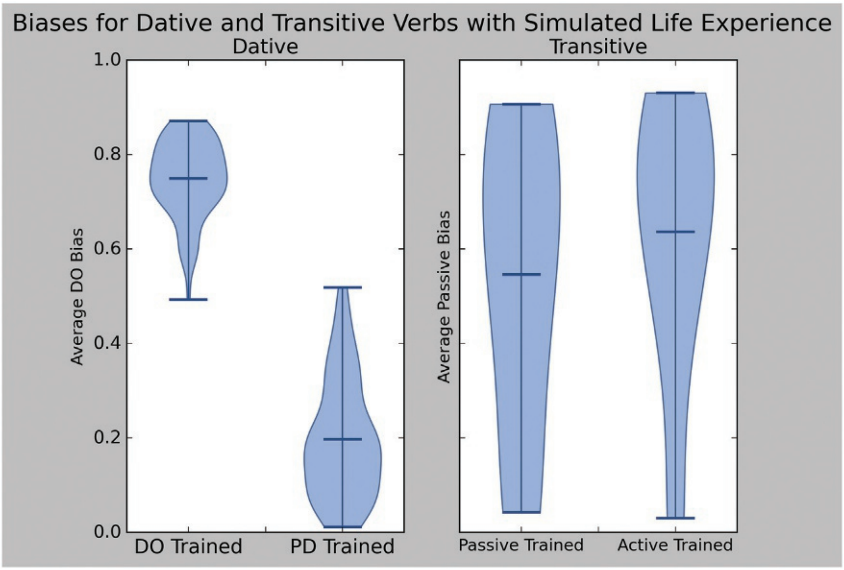

Fig. 5 Average verb biases from models which first learned semi-realistic verb biases, and then were taught novel verb biases as in Block 1 of Kelley (2019). The model shows that dative verbs can acquire these novel experimental biases, while transitive verbs cannot.

simulating behavioral experiments that manipulated verb biases. These findings suggest that the small connectionist models are particularly useful for characterizing the trial-by-trial changes that occur.

4. General conclusions

This review has focused on verb biases as a case study of statistical learning in language. Verb biases impact comprehension and production, and the mechanisms that learn and update them appear to be quite similar to the mechanisms that acquire other kinds of regularities, both linguistic and nonlinguistic.

The first section considered how verb biases are used in language. Both language comprehension and language production are easier when the main verb of a sentence is more likely to occur in that sentence's structure. Interestingly, this is the case for all language users examined, rather than being confined to adults or first-language speakers of a language. Instead, differences between groups usually arose when those with less experience relied on verb bias too much. Although this has been found multiple times, it is not clear what drives this effect or if it is the same factor in all groups. For example, one possibility is that verb biases are easier to learn about than factors like semantic plausibility. Consequently, while adult first-language speakers have received enough information to consider both verb biases and the meaning of the words in a sentence (Garnsey et al., 1997), second-language speakers and children may not have enough language experience to know that both factors are important (e.g., Kidd et al., 2011; Lee et al., 2013). It is also possible that verb biases are easy to access, which might explain why they remain accessible to people with aphasia in comprehension (DeDe, 2013), and why they create memory errors for children who are recalling sentences (Kidd et al., 2006). As it is, the literature only hints at why verb biases have the potential to drown out other sources of linguistic knowledge. Empirical work is needed to distinguish between these possibilities and others.

Next, potential learning mechanisms of verb bias were discussed. Experimental work suggests that verb biases are learned using distributional learning. This mechanism can further be described as error-based and incremental. Finally, this learning mechanism shares attributes, such as incremental updating, with associative learning. Although this review focuses narrowly on verb biases, close examination of learning mechanisms holds insight for other kinds of language statistics. Learning of phonotactic rules, which constrain how sounds can be combined in a language, also changes with experience (Dell, Kelley, Bian, & Holmes, 2019; Dell, Kelley, Hwang, & Bian, 2021) and shows evidence of slow reversal (Anderson & Dell, 2018). However, not all language statistics show ready updating. For example, English speakers cannot learn phonotactic rules that differ based on speech rate (Warker, Dell, Whalen, & Gereg, 2008) or lexical tone (Bian & Dell, 2020). Where and why these kinds of constraints exist is an important issue for the learning of many kinds of language statistics, as well as for learning more generally.

Finally, models of verb bias learning were reviewed and compared. These models take very different approaches, but all make important contributions to understanding the information that humans must gather in order to learn verb biases. However, it is important to continue comparing models to human behavior. Human learners clearly consider many different sources of information when they learn about verb behavior, and it is almost certain that new empirical discoveries about these sources of information will drive changes and improvements to current models.

References

Ambridge, B., & Blything, R. P. (2016). A connectionist model of the retreat from verb argument structure overgeneralization. *Journal of Child Language*, *43*(6), 1245–1276.

Ambridge, B., Pine, J. M., Rowland, & Young, C. R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure overgeneralization errors. *Cognition*, *106*(1), 87–129.

Anderson, N. D., & Dell, G. S. (2018). The role of consolidation in learning context-dependent phonotactic patterns in speech and digital sequence production. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(14), 3617–3622. https://doi.org/10.1073/pnas.1721107115.

Anible, B., Twitchell, P., Waters, G. S., Dussias, P. E., Piñar, P., & Morford, J. P. (2015). Sensitivity to verb bias in American Sign Language–English bilinguals. *Journal of Deaf Studies and Deaf Education*, *20*(3), 215–228. https://doi.org/10.1093/deafed/env007.

Barak, L., Fazly, A., & Stevenson, S. (2014, June). Learning verb classes in an incremental model. In *Proceedings of the fifth workshop on cognitive modeling and computational linguistics* (pp. 37–45). http://www.aclweb.org/anthology/W/W14/W14-2005.

Barak, L., Goldberg, A., & Stevenson, S. (2016, November). Comparing computational cognitive models of generalization in a language acquisition task. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 96–106). https://doi.org/10.18653/v1/D16-1010.

Bernolet, S., & Hartsuiker, R. J. (2010). Does verb bias modulate syntactic priming? *Cognition*, *114*(3), 455–461. https://doi.org/10.1016/j.cognition.2009.11.005.

Bian, Y., & Dell, G. S. (2020). Novel stress phonotactics are learnable by English speakers: Novel tone phonotactics are not. *Memory & Cognition*, *48*(2), 176–187. https://doi.org/10.3758/s13421-019-01000-9.

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

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In *Cognitive foundations of interpretation* (pp. 69–94). KNAW. http://hdl.handle.net/11858/00-001M-0000-0013-1A32-0.

Brooks, P. J., & Tomasello, M. (1999). How children constrain their argument structure constructions. *Language*, 720–738.

Cantwell, G., Crossley, M. J., & Ashby, F. G. (2015). Multiple stages of learning in perceptual categorization: Evidence and neurocomputational theory. *Psychonomic Bulletin and Review*, *22*, 1598–1613. https://doi.org/10.3758/s13423-015-0827-2.

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, *26*(5), 609–651. https://doi.org/10.1207/s15516709cog2605_3.

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

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(6), 1084.

Coyle, J. M., & Kaschak, M. P. (2008). Patterns of experience with verbs affect long-term cumulative structural priming. *Psychonomic Bulletin & Review*, *15*(5), 967–970. https://doi.org/10.3758/PBR.15.5.967.

DeDe, G. (2013). Effects of verb bias and syntactic ambiguity on reading in people with aphasia. *Aphasiology*, *27*(12), 1408–1425. https://doi.org/10.1080/02687038.2013.843151.

Dell, G. S., Kelley, A. C., Bian, Y., & Holmes, E. W. (2019). Tuning the blueprint: How studies of implicit learning during speaking reveal the information processing components of the production system. *Language, Cognition and Neuroscience*, *34*(9), 1246–1256.

Dell, G. S., Kelley, A. C., Hwang, S., & Bian, Y. (2021). The adaptable speaker: A theory of implicit learning in language production. *Psychological Review*, *128*(3), 446–487. https://doi.org/10.1037/rev0000275.

Dussias, P. E., & Scaltz, T. R. C. (2008). Spanish–English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychologica*, *128*(3), 501–513. https://doi.org/10.1016/j.actpsy.2007.09.004.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211. https://doi.org/10.1016/0364-0213(90)90002-E.

Engel, A., Grafmiller, J., Rosseel, L., & Szmrecsanyi, B. (2022). Assessing the complexity of lectal competence: The register-specificity of the dative alternation after give. *Cognitive Linguistics*, *33*(4), 727–766. https://doi.org/10.1515/cog-2021-0107.

Fellows, L. K., & Farah, M. J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: Evidence from a reversal learning paradigm. *Brain*, *126*, 1830–1837. https://doi.org/10.1093/brain/awg180.

Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, *33*(6), 715–736. https://doi.org/10.1006/jmla.1994.1034.

Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 555. https://doi.org/10.1037/0278-7393.16.4.555.

Ferreira, V. S., & Schotter, E. R. (2013). Do verb bias effects on sentence production reflect sensitivity to comprehension or production factors? *Quarterly Journal of Experimental Psychology*, *66*(8), 1548–1571. https://doi.org/10.1080/17470218.2012.753924.

Fisher, C., Jin, K. S., & Scott, R. M. (2020). The developmental origins of syntactic bootstrapping. *Topics in Cognitive Science*, *12*(1), 48–77. https://doi.org/10.1111/tops.12447.

Flett, S., Branigan, H. P., & Pickering, M. J. (2013). Are non-native structural preferences affected by native language preferences? *Bilingualism: Language and Cognition*, *16*(4), 751–760.

Frank, M., & Claus, E. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision-making, and reversal. *Psychological Review*, *113*(2), 300–326. https://doi.org/10.1037/0033-295X.113.2.300.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1.

Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, 748–775. https://www.jstor.org/stable/4489780.

Gahl, S., Garnsey, S. M., Fisher, C., & Matzen, L. (2006). That sounds unlikely: Syntactic probabilities affect pronunciation. In *Proceedings of the 27th meeting of the cognitive science society*. https://escholarship.org/uc/item/1h9879s9.

Gahl, S., & Menn, L. (2016). Usage-based approaches to aphasia. *Aphasiology*, *30*(11), 1361–1377. https://doi.org/10.1080/02687038.2016.1140120.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*(1), 58–93. https://doi.org/10.1006/jmla.1997.2512.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*(5), 178–186. https://doi.org/10.1016/S1364-6613(00)01467-4.

Gries, S. T., & Wulff, S. (2005). Do foreign language learners also have constructions? *Annual Review of Cognitive Linguistics*, *3*(1), 182–200.
Hare, M., Elman, J. L., Tabaczynski, T., & McRae, K. (2009). The wind chilled the spectators, but the wine just chilled: Sense, structure, and sentence comprehension. *Cognitive Science*, *33*(4), 610–628. https://doi.org/10.1111/j.1551-6709.2009.01027.x.
Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language*, *48*(2), 281–303. https://doi.org/10.1016/S0749-596X(02)00516-8.
Hare, M., McRae, K., & Elman, J. (2004). Admitting that admitting verb sense into corpus analyses makes sense. *Language and Cognitive Processes*, *19*(2), 181–224. https://doi.org/10.1080/01690960344000152.
Hawkins, R. D., Yamakoshi, T., Griffiths, T. L., & Goldberg, A. E. (2020). Investigating representations of verb bias in neural language models. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)* (pp. 4653–4663). https://doi.org/10.18653/v1/2020.emnlp-main.376.
Hoffmann, F. K., Perkins, M. J., & Calvin, A. D. (1956). The effect of problem difficulty on discrimination reversal. *Journal of Comparative and Physiological Psychology*, *49*(6), 547–548. https://doi.org/10.1037/h0041304.
Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., & Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience*, *345*, 12–26. https://doi.org/10.1016/j.neuroscience.2016.03.021.
Kelley, A. C. (2019). *Investigating the mechanisms of verb bias learning* (Master's thesis). University of Illinois at Urbana-Champaign. IDEALS. http://hdl.handle.net/2142/105091.
Kelley, A. (2022). *Modeling the mechanisms of verb bias learning*. (Doctoral dissertation).
Kendler, T. S., & Kendler, H. H. (1970). An ontogeny of optional shift behavior. *Child Development*, *41*(1), 1–27. https://www.jstor.org/stable/1127386.
Kennison, S. M. (2001). Limitations on the use of verb information during sentence comprehension. *Psychonomic Bulletin & Review*, *8*(1), 132–138.
Kidd, E., Lieven, E., & Tomasello, M. (2006). Examining the role of lexical frequency in the acquisition and processing of sentential complements. *Cognitive Development*, *21*(2), 93–107. https://doi.org/10.1016/j.cogdev.2006.01.006.
Kidd, E., Lieven, E. V., & Tomasello, M. (2010). Lexical frequency and exemplar-based learning effects in language acquisition: Evidence from sentential complements. *Language Sciences*, *32*(1), 132–142. https://doi.org/10.1016/j.langsci.2009.05.002.
Kidd, E., Stewart, A. J., & Serratrice, L. (2011). Children do not overcome lexical biases where adults do: The role of the referential scene in garden-path recovery. *Journal of Child Language*, *38*(1), 222–234. https://doi.org/10.1017/S0305000909990316.
Kootstra, G. J., & Doedens, W. J. (2016). How multiple sources of experience influence bilingual syntactic choice: Immediate and cumulative cross-language effects of structural priming, verb bias, and language dominance. *Bilingualism: Language and Cognition*, *19*(4), 710–732.
Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science*, *8*(2), 225–248. https://doi.org/10.1080/095400996116893.
Lee, E. K., Lu, D. H. Y., & Garnsey, S. M. (2013). L1 word order and sensitivity to verb bias in L2 processing. *Bilingualism: Language and Cognition*, *16*(4), 761–775. https://doi.org/10.1017/S1366728912000776.
Lin, Y., & Fisher, C. (2017). Error-based learning: A mechanism for linking verbs to syntax. In G. Gunzelman, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th annual conference of the cognitive science society* (pp. 754–759). Austin, TX: Cognitive Science Society. https://cogsci.mindmodeling.org/2017/papers/0151/paper0151.pdf.

McDaniel, J. W. (1969). Reversal learning in the sow bug. *Psychonomic Science*, *16*(5), 261–262. https://doi.org/10.3758/BF03332678.

Mitchell, D. C. (1989). Verb guidance and other lexical effects in parsing. *Language and Cognitive Processes*, *4*(3–4), SI123–SI154. https://doi.org/10.1080/01690968908406366.

Murray, E. A., & Gaffan, D. (2006). Prospective memory in the formation of learning sets by Rhesus monkeys (Macaca mulatta). *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(1), 87–90. https://doi.org/10.1037/0097-7403.32.1.87.

Parisien, C., & Stevenson, S. (2010). Learning verb alternations in a usage-based Bayesian model. In *Vol. 32(32). Proceedings of the Annual Meeting of the Cognitive Science Society* (pp. 2674–2679). https://escholarship.org/content/qt186313ch/qt186313ch.

Perek, F., & Goldberg, A. E. (2015). Generalizing beyond the input: The functions of the constructions matter. *Journal of Memory and Language*, *84*, 108–127. https://doi.org/10.1016/j.jml.2015.04.006.

Perek, F., & Goldberg, A. E. (2017). Linguistic generalization on the basis of function and constraints on the basis of statistical preemption. *Cognition*, *168*, 276–293. https://doi.org/10.1016/j.cognition.2017.06.019.

Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language*, *37*(3), 607–642. https://doi.org/10.1017/S0305000910000012.

Peter, M., Chang, F., Pine, J. M., Blything, R., & Rowland, C. F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language*, *81*, 1–15. https://doi.org/10.1016/j.jml.2014.12.002.

Qi, Z., Yuan, S., & Fisher, C. (2011). Where does verb Bias come from? Experience with particular verbs affects online sentence processing. In N. Danis, K. Mesh, & H. Sung (Eds.), *BUCLD 35: proceedings of the 35th annual Boston university conference on language development*. http://web.mit.edu/zqi/www/uploads/1/4/8/9/14891652/bucld_final.pdf.

Qian, Z., Lee, E. K., Lu, D. H. Y., & Garnsey, S. M. (2019). Native and non-native (L1-mandarin) speakers of English differ in online use of verb-based cues about sentence structure. *Bilingualism: Language and Cognition*, *22*(5), 897–911. https://doi.org/10.1017/S1366728918001001.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906–914. https://doi.org/10.1002/wcs.78.

Ryskin, R., Qi, Z., Covington, N. V., Duff, M., & Brown-Schmidt, S. (2018). Knowledge and learning of verb biases in amnesia. *Brain and Language*, *180*, 62–83. https://doi.org/10.1016/j.bandl.2018.04.003.

Ryskin, R. A., Qi, Z., Duff, M. C., & Brown-Schmidt, S. (2017). Verb biases are shaped through lifelong learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(5), 781.

Salamoura, A., & Williams, J. N. (2006). Lexical activation of cross-language syntactic priming. *Bilingualism: Language and Cognition*, *9*(3), 299–307.

Sanders, B. (1971). Factors affecting reversal and nonreversal shifts in rats and children. *Journal of Comparative and Physiological Psychology*, *74*(2), 192–202. https://doi.org/10.1037/h0030336.

Scott, R. M., & Fisher, C. (2009). Two-year-olds use distributional cues to interpret transitivity-alternating verbs. *Language and Cognitive Processes*, *24*(6), 777–803. https://doi.org/10.1080/01690960802573236.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299. https://doi.org/10.1016/j.cogpsych.2004.03.001.

Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language*, *28*(1), 127–152. https://doi.org/10.1017/S0305000900004608.

Thothathiri, M., & Braiuca, M. C. (2021). Distributional learning in English: The effect of verb-specific biases and verb-general semantic mappings on sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(1), 113. https://doi.org/10.1037/xlm0000814.

Thothathiri, M., Evans, D. G., & Poudel, S. (2017). Verb bias and verb-specific competition effects on sentence production. *PLoS One*, *12*(7), e0180580. https://doi.org/10.1371/journal.pone.0180580.

Thothathiri, M., & Rattinger, M. G. (2016). Acquiring and producing sentences: Whether learners use verb-specific or verb-general information depends on cue validity. *Frontiers in Psychology*, 7, 404. https://doi.org/10.3389/fpsyg.2016.00404.

Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*, *1*(2), 147–165. https://doi.org/10.1515/LANGCOG.2009.008.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89–134. https://doi.org/10.1016/S0010-0277(99)00032-3.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 528–553. https://doi.org/10.1037/0278-7393.19.3.528.

Twomey, K. E., Chang, F., & Ambridge, B. (2014). Do as I say, not as I do: A lexical distributional account of English locative verb class acquisition. *Cognitive Psychology*, *73*, 41–71. https://doi.org/10.1016/j.cogpsych.2014.05.001.

Twomey, K. E., Chang, F., & Ambridge, B. (2016). Lexical distributional cues, but not situational cues, are readily used to learn abstract locative verb-structure associations. *Cognition*, *153*, 124–139. https://doi.org/10.1016/j.cognition.2016.05.001.

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190.

Warker, J. A., Dell, G. S., Whalen, C. A., & Gereg, S. (2008). Limits on learning phonotactic constraints from recent production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1289.

Wasserman, E. A., & Miller, R. R. (1997). What's elementary about associative learning? *Annual Review of Psychology*, *48*(1), 573–607.

Wilson, M. P., & Garnsey, S. M. (2009). Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language*, *60*(3), 368–392.

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language*, *65*(1), 1–14. https://doi.org/10.1016/j.jml.2011.03.001.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, *56*(3), 165–209. https://doi.org/10.1016/j.cogpsych.2007.04.002.

Yuan, S., & Fisher, C. (2009). "Really? She blicked the baby?" two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*, *20*(5), 619–626. https://doi.org/10.1111/j.1467-9280.2009.02341.x.

This page intentionally left blank

CHAPTER SIX

Production, processing, and prediction in bilingual codeswitching

Jorge R. Valdés Kroff[a,*] **and Paola E. Dussias**[b]

[a]Spanish and Portuguese Studies, University of Florida, Gainesville, FL, United States
[b]Department of Spanish, Italian, and Portuguese, Pennsylvania State University, State College, PA, United States
[*]Corresponding author: e-mail address: jvaldeskroff@ufl.edu

Contents

Abstract

Bilinguals in the presence of other bilinguals engage in codeswitching, the fluid and intentional alternation between languages during bilingual speech or text. Codeswitching has most prominently been studied from a theoretical and sociolinguistic perspective, but over the last 20 years, psycholinguists have increasingly turned their attention to understanding the cognitive and neural underpinnings of codeswitching. Despite its common use among bilinguals, understanding the production and comprehension of codeswitching presents unique challenges to current theories of sentence processing. We overview the planning, production, and comprehension of codeswitching, discussing the complex interaction between linguistic, cognitive, and social factors that modulate its use and its comprehension. This overview brings to the forefront an apparent paradox between measurable processing costs and the ease with which bilinguals engage in codeswitching, namely, that codeswitching introduces greater ambiguity into the linguistic signal yet does not lead to disruptive

Psychology of Learning and Motivation, Volume 78
ISSN 0079-7421
https://doi.org/10.1016/bs.plm.2023.02.004

processing delays. To account for this contradiction, we propose the Adaptive Predictability hypothesis with two premises: bilinguals adapt predictive cues in sentence processing as a consequence of exposure to distributional regularities in production, and they recruit greater cognitive control in the service of rapidly integrating codeswitches in real time. We end the chapter by illustrating recent findings that support the Adaptive Predictability hypothesis and areas for future directions.

1. Introduction

Speaker A. …And that's going to include the message, o sea… quién lo vendió, el combo… todo eso.

Speaker A. Después, I'm going to have a spreadsheet para… keep track of nada ehm cuántos combos tenemos ehm who do we give it to porque el takeoff va a ser así. Like, tú pusiste nombre, yo tengo la lista, ah ok, tú pediste un combo, toma. ¿Entiendes? It's gonna be like that so…

Speaker B. [overlapping with prior turn] Básicamente… como que

Speaker A. Uh huh, in order to have todo de esta información, pues la mamá… la esposa del papá de C. me va a pasar todos esos datos, and I'm just gonna organize it in a spreadsheet. If you guys want, I can eh send it to you when it's… done.

Caption: Transcript of a segment of bilingual conversation happening between Puerto Rican middle school students during a meeting for a school fundraising event.

Examine the bilingual exchange above,[a] which involves several middle school-aged Spanish-English bilingual children at an after-school meeting to discuss logistics for a fundraising event. We have transcribed just under a minute of the recording. To visually aid the reader, we have color-coded the transcription to highlight the use of both languages. Sections that appear

[a] A colleague who works at a private English-medium middle school in Puerto Rico personally shared the recording of the conversation and provided permission to use it for illustrative purposes of bilingual speech.

in red are identifiably Spanish elements, and sections in black are identifiably English elements; elements that appear in blue are filler words and arguably language neutral; and purple indicates lexical items that potentially may be considered established borrowings. Just this short example underscores the complexities that can arise in bilingual speech and shows the nature of bilingual exchanges: common, grammatical, and nondisruptive to conversational exchanges.

One characteristic of bilingual speech that features prominently above is *codeswitching*, which we functionally define as the *fluid* and *intentional* switching between languages in bilingual speech and text[b] (e.g., Lipski, 1978; Poplack, 1980). Other researchers have offered broader definitions of codeswitching, including any instances in which the two languages may appear. The extent to which codeswitches include other bilingual language phenomena such as borrowings, loanwords, and calques remains highly debatable within the field (Bullock & Toribio, 2009). For us, "fluid" and "intentional" are key features of our definition. By fluid, we refer to the grammatical and phonological integration that seemingly occurs at the moment of the codeswitch. While there may be phonetic cues that subtly and reliably change before codeswitches (Fricke et al., 2016; Johns & Steuck, 2021), the phonologies of the two languages are respected in codeswitching. In other words, in the exchange above, the Spanish elements are not an imposition of English phonology on Spanish (aside from individual differences in one's own accent). We further specify intentionality. Codeswitching occurs in pragmatic contexts in which bilingual speakers dialogue with other bilingual speakers. One strong piece of evidence underlying this intentionality is that bilinguals can shift to speaking in one language alone when necessary and do not uncontrollably produce codeswitches in the presence of monolingual speakers. This intentionality also separates codeswitching from other bilingual phenomena such as lexical gap switches, in which a speaker may switch into a single word because they cannot access the lexical item in the target language in the moment.

The most widely-cited classification of codeswitches appeared in Muysken's (2000) classic monograph *Bilingual Speech*, which outlines a

[b] One important note: throughout the remainder of the paper, we will use the term "speech" to refer to both written and spoken modalities in oral languages. If a distinction needs to be made, we will distinguish between spoken modality and written modality. This terminological choice is in part motivated by the observation that there are minor structural differences in codeswitching when comparing spoken language and written language corpora (Callahan 2004; Guzzardo Tamargo et al., 2016; Montes-Alcalá 2000,2001).

typology of codeswitches that is linked to linguistic, psycholinguistic, and sociolinguistic factors. Built from analyses of bilingual corpora, Muysken proposes that codeswitches be classified into three possible categories: insertional, alternational, and congruent lexicalization. Insertional codeswitches involve the embedding of single phrases or words into an otherwise unilingual stretch of speech. Alternational codeswitches occur when speech starts in one language and continues into another. Congruent lexicalization is a special case of codeswitching that most obviously occurs when the two languages are structurally similar, and thus permits freer and more copious switching between languages (see examples 1–3 from Spanish-English bilingual codeswitching cited in Muysken, 1997):

1. *Yo anduve* in a state of shock *pa' dos días* (*Pfaff, 1979*) *[Insertion]*

   "I walked in a state of shock for two days"
2. *Ándale pues* and do come again (*Peñalosa, 1980*) *[Alternation]*

   "That's all right then and do come again"
3. *Bueno,* in other words, *el* flight *que sale de Chicago* around three o'clock (*Pfaff, 1976*) *[Congruent Lexicalization]*

   "Good, in other words, the flight that leaves Chicago around three o'clock"

Muysken's typology promotes that several factors can affect the incidence and type of codeswitching. A sociolinguistic study comparing two bilingual communities, French-English bilinguals in the Ottawa-Hull region of Canada and Spanish-English bilinguals in New York City, illustrates this point (Poplack, 1987). The language pairings represent an interesting comparison between English and a romance language in two regions that are economically comparable. Despite these similarities, Poplack documented differences across the two bilingual groups. The Spanish-English bilingual group engaged in more frequent and intricate codeswitching, demonstrating high flexibility in the syntactic junctures where codeswitching could occur. In contrast, the French-English bilinguals codeswitched less frequently and were more limited in the types of codeswitching that they produced. Poplack explains the difference as deriving from socio-political differences between the two communities. The Ottawa-Hull region exerted a greater social dispreference to codeswitching, likely reflecting the political tension that exists between French and English, especially in a border region between English-majority Ontario (Ottawa) and French-majority Quebec (Hull). This study demonstrates that beyond the linguistic factors that may contribute to codeswitching, a powerful social component must additionally be taken into account.

The complexities underlying bilingual speech that are reflected in Muysken's typology, and that can be surmised in the Ottawa-Hull/New York City example, lead to several questions. Why does the speaker choose to codeswitch, and what factors affect when codeswitches will be produced? Is there a communicative benefit to the listener or the speaker when bilinguals purposefully codeswitch? How does the listener anticipate or prepare for a codeswitch? Our goal is to provide a psycholinguistically plausible account, which we term adaptive predictability, for how bilinguals attune their production and comprehension systems for bilingual exchanges. We do not claim that this mechanism is specific to bilingual speakers who codeswitch; rather, codeswitching is a unique window that affords language scientists the opportunity to investigate connections between production and comprehension due to the heightened ambiguity that occurs during bilingual exchanges. We begin with a brief overview of theoretical models of codeswitching, followed by a closer look at speech planning and production of codeswitches. We then turn to the comprehension system and examine past studies that have focused on the real-time processing of codeswitched speech or text. The next section introduces the Adaptive Predictability hypothesis, which we illustrate with two recent studies. We then offer concluding remarks and future directions.

2. Theoretical accounts of codeswitching

Over the past four decades, linguists have proposed theoretical accounts for grammatical and ungrammatical instances of codeswitching. Researchers have used a variety of methods, including acceptability judgments (e.g., Liceras et al., 2008; González-Vilbazo et al., 2013), corpus-based analyses (e.g., Deuchar, 2020; Poplack, 1980, Torres Cacoullos & Travis, 2018), and production-based elicitation tasks (e.g., Gullberg et al., 2009; Munarriz-Ibarrola et al., 2022; Sarkis & Montag, 2021). Because it is uncontroversial that codeswitches are not random but are instead systematic and grammatical, these theoretical accounts have as a primary goal to propose unified accounts of codeswitching. However, the common ground stops there. Two primary issues have dominated theoretical debates. One is the contribution of each of the participating languages in codeswitching. Another issue is whether grammatical constraints are specific to bilingual grammars or not.

On the first issue, we have termed the distinction as symmetric and asymmetric approaches to codeswitching (Valdés Kroff, 2012). This distinction centers on the theoretical stance of the researcher as to whether both languages neutrally contribute to the grammatical properties of codeswitching or if one language "controls" the grammatical frame whereas the other language plays a more limited role. Symmetric approaches include early descriptive accounts most famously embodied by Poplack's (1980) Equivalence Constraint, which claims that codeswitches are only permissible if the two grammars are congruent or equivalent at the point of the codeswitch. Spanish and English differ in word order in the placement of object pronouns that are complements to the verb (4).

4a. *I bought it at the store yesterday.*
4b. *(Yo) lo compré en la tienda ayer.*
(I)it bought at the store yesterday

The placement of the object pronoun 'it' is post-verbal in English (4a) but pre-verbal in Spanish (4b). The Equivalence Constraint thus posits that a codeswitch before or after the verb is not licensed in Spanish-English codeswitching, i.e., **Yo lo bought at the store yesterday* or **I bought lo en la tienda ayer*. Symmetric approaches remain agnostic to the participatory role of the two languages or take a stance that they can equally contribute.

Asymmetric approaches are most clearly represented by Myers-Scotton's Matrix Language Framework (MLF, Myers-Scotton, 1993; Myers-Scotton & Jake, 2000). This framework takes direction from psycholinguistic models of production, in particular the separation between grammatical and lexical morpheme selection (Levelt, 1989). This approach is partially inspired by well-known speech error phenomena showing that lexical elements are swapped systematically but often leave their grammatical elements (inflectional morphology) in place, e.g., *the moving company lamped the ships* instead of *the moving company shipped the lamps*. The MLF attributes a single matrix language that sets the grammatical frame for codeswitched speech. Consequently, grammatical elements (e.g., conjunctions, complementizers, inflectional morphology, determiners) should come from one language alone. The embedded language plays a more limited role, contributing lexical items into a matrix language frame. In (5), for example, the matrix language, Acholi (underlined below), sets the grammatical frame while the embedded language, English, only contributes lexical items.

5. *chances <u>me</u> accident <u>pol ka i-</u> boarding taxi* (Myers-Scotton & Jake, 2009)
chances ASSOC accident many if 2SG-boarding taxi
"[The] chances of [an] accident [are] many if you board [a] taxi"

The second issue concerns how specific the constraints are to bilingual grammars. Theoretical constraints proposed in the 1980s and 1990s, such as the Functional Head Constraint (Belazi et al., 1994), relied upon language-specific mechanisms such as a language feature checking mechanism or a language node (Sankoff & Poplack, 1981) to account for codeswitched utterances. To illustrate, the Functional Head Constraint extended Abney's (1987) proposal that functional heads such as determiners and complementizers are generally required to select the features of their complement (a process that Abney refers to as f-selection) to include language index in bilingual speech as one of the features being checked. However, proponents of constraint-free accounts (Mahootian & Santorini, 1996; MacSwan, 1999) criticized these approaches as introducing bilingual-specific machinery, thus questioning its universality within the language faculty. Constraint-free approaches instead propose that bilingual codeswitching should be fully accountable simply by the grammatical properties of the participating languages. Constraint-free approaches now dominate theoretical accounts of bilingual codeswitching, and include minimalist approaches to codeswitching (e.g., González-Vilbazo & López, 2011) and approaches working within Distributed Morphology (López, 2020; Cruz, 2021).

3. Planning and production of codeswitched speech

Bilingual speech corpora (e.g., the New Mexico Spanish-English Bilingual corpus or NMSEB, Torres Cacoullos & Travis, 2018; the Bangor Miami corpus of Spanish-English bilingual speech, Deuchar et al., 2014) demonstrate that when bilinguals interact with other bilinguals, they use stretches of unilingual discourse, alongside fluid alternations between their two languages, adapting their linguistic demands on the fly according to their patterns of language use, their communicative intentions, and the immediate spoken context (Green, 2018:1). What is remarkable about this feat is that bilinguals seldom make language selection errors. One important question, then, is what mechanism bilinguals engage to move effortlessly in and out of their two languages.

Green and Abutalebi (2013) argue that eight domain-general control processes—*goal maintenance*, *interference suppression*, *salient cue detention*, *selective response inhibition*, *task engagement*, *task disengagement* and *opportunistic planning*—are recruited to different degrees by bilinguals to regulate their two languages in everyday interactions (Green & Abutalebi, 2013:518–528). In their Adaptive Control hypothesis, three distinct interactional contexts, reflecting everyday conversational use of language, will require the differential recruitment of these control processes. In *dual language contexts*, where codeswitching is expected because both languages are used, the demands placed on control processes will be the greatest due to the need to engage and disengage the two linguistic systems. Control processes such as goal maintenance and interference control will be more pertinent in the *single-language contexts*, where only one language is used exclusively in a particular environment (e.g., speaking one language at home and another one at work), and where codeswitching would be viewed as an instance of a language intrusion error (Gollan et al., 2014). Finally, in *dense-codeswitching contexts*, where bilinguals routinely mix freely between their two languages within a single utterance, the demands are expected to be the smallest. The Adaptive Control Hypothesis also recognizes that bilinguals may find themselves predominantly in only one of these three contexts or may shift from context to context, requiring cognitive control mechanisms to adapt accordingly. More recently, Green (2018:12) has argued that changes in bilinguals' language use specifically—not just the greater social context itself—may also differentially engage cognitive control mechanisms. To illustrate, whereas dense codeswitching may instantiate an open control state, where monitoring and interference suppression at a local (i.e., lexical) level would be required, insertional codeswitching would necessitate the engagement and disengagement of the participating languages, therefore requiring global inhibition. As such, dense codeswitching would promote a broader attentional state relative to insertional switching, which would require a narrow attentional state. Importantly, these control mechanisms are argued to be domain general rather than language specific (Green & Abutalebi, 2013), with the implication that the process of adapting to different task demands during the production of codeswitched language may have ramifications not only for language-related processes but also for non-linguistic tasks.

Although studies testing the Adaptive Control Hypothesis have been difficult to carry out given the methodological challenges that come with recording brain activity while participants speak to other bilinguals, comprehension studies that simulate bilingual discourse and naturalistic interactions are beginning to provide the first pieces of evidence that bilinguals' interactional contexts modulate the engagement of language control networks. Kaan et al. (2020) examined whether bilinguals could dynamically shift between attentional control states depending on whether another bilingual speaker or a monolingual speaker was present during a reading task involving unilingual and codeswitched sentences. The EEG findings showed an early frontal positivity effect that was largest when bilinguals read codeswitched sentences in the presence of a known monolingual interlocutor. Because past research had shown that the prefrontal cortex and anterior cingulate cortex are primarily involved in language control and cognitive control, the finding provided evidence that knowledge of the interlocutor's language modulates the engagement of language control in bilinguals. A second piece of particularly strong evidence comes from a study by Blanco-Elorrieta and Pylkkänen (2017), who recorded MEG responses in a study investigating cued language switching and natural language switching in a setting that simulated (or not) naturalistic interactions. In production, participants' language choice (Arabic or English) was determined by a color cue or was socially-conditioned by a facial cue of an individual introduced as a monolingual or bilingual speaker. In the comprehension study, the processing of isolated word switches was compared to the comprehension of language switches in natural Arabic-English conversations. Blanco-Elorrieta and Pylkkänen found that in production, voluntary switching did not engage the prefrontal cortex or elicit behavioral switch costs. The comprehension study showed that the anterior cingulate and prefrontal cortex regions were more involved when language switching was externally-cued by a color stimulus relative to listening to naturalistic codeswitched conversations. The evidence stemming from these comprehension studies only provides indirect support for the differential recruitment of different control processes during the production of codeswitched language. Notwithstanding, support from related fields of study (e.g., motor control) have demonstrated that imagery engages a network of cortical, subcortical, and cerebellar regions that largely overlaps with the network for actual execution (see, for example, Hardwick et al., 2018; Jeannerod, 2001). In this sense, these results described here support the hypothesis that the diversity of bilingual experiences differentially modulates the engagement of cognitive processes (see also Beatty-Martínez & Dussias, 2017 for related findings).

The premises of the Adaptive Control Hypothesis highlight the fact that being bilingual does not, in and of itself, require the engagement of a unique pattern of cognitive control. An important future line of research is to explicitly test the prediction of the hypothesis against different types of codeswitches. This is a point raised by Treffers-Daller et al. (2020), who discuss that Green and Abutalebi's (2013) reference to "dense codeswitching" remains rather underspecified: whereas in some cases it refers to insertional codeswitching (example (3) above), other times its reference includes alternations such as those shown in example (4). Correlational results reported in Hofweber, Marinis, and Treffers-Daller (2016) suggest that bilinguals' performance on executive control tasks (measured in their study by a Flanker task) is affected by the frequency with which bilinguals engage in codeswitching, along with the specific type of intrasentential codeswitching they use within their social network. A next step would be to find direct evidence by capitalizing on the method with high temporal resolution to track the unfolding of neural activity as it occurs millisecond by millisecond.

3.1 Why do bilinguals codeswitch?

One approach to answer this question has been to ask codeswitching bilinguals why they codeswitch. As with most metalinguistic reflections, the answers are after-the-fact explanations that are unreliable. Bilinguals often surmise that they codeswitch to fill linguistic gaps (Heredia & Altarriba, 2001). Fueling this assumption is the persistent, yet mistaken, belief among parents, educators, and healthcare providers that codeswitching reflects mental laziness and is a sign of alingualism or an inability to maintain the two languages separate. As critical linguistic, psycholinguistic, and neurolinguistic attention has been directed to the study of codeswitching, one unsurprising result has been that only those bilinguals who are linguistically competent in their two languages produce fluid and rule-governed language alternations (Lipski, 1982; Myers-Scotton, 1993; Poplack, 1980), thus directly challenging these ill-formed perceptions of bilingual language use. Instead, bilingual codeswitching should be seen as a complex and skillful speech act that requires a high degree of proficiency and coordination across a bilingual's languages. Below we discuss several bottom-up processes that may partially explain why bilinguals codeswitch.

3.1.1 Lexical accessibility

An oft-cited reason for engaging in codeswitching relies on the notion of *lexical accessibility*—codeswitching takes place when bilinguals experience difficulty retrieving a word in the current language that best conveys the intended meaning. Bilinguals, like monolinguals, also experience *tip-of the-tongue* (TOT) word retrieval failures, which they might resolve by resorting to codeswitching. And indeed, lab research has shown that bilinguals experience more TOTs than monolinguals. Several reasons have been proposed to explain this difference: competition for selection between translation equivalents (Kroll et al., 2006; Sarkis & Montag, 2021), reduced frequency of use of words in each language (Gollan et al., 2011; Pyers et al., 2009), and interference with the accessibility of one of the languages due to immersion in the other one (Linck et al., 2009). While the evidence that bilingual speakers experience word retrieval difficulties is undisputed, the notion that bilinguals predominantly switch as a strategy to make up for word finding difficulties faces substantial challenges. First, inspection of naturally produced bilingual speech (e.g., Torres Cacoullos & Travis, 2018; Chapters 2 and 3) has documented that single other-language noun insertions are the most frequent type of language switch (Poplack, 1980; Torres Cacoullos & Travis, 2018). If switching exclusively fills in lexical gaps, it would mean that bilinguals are "in a constant TOT state"—a proposition that would be at odds with the observation that codeswitching is fluid and intentional. Second, the lexical accessibility account predicts that only words that are subject to frequency effects (e.g., nouns, verbs, etc.) should participate in codeswitching; however, evidence from bilingual corpora suggests that closed-class words are also switched, as shown in (6) below (switched closed class lexical item in italics):

6. *I could understand* que *(that) you don't know how to speak Spanish, ¿verdad? (Poplack, 1980: 596).*

Third, in balanced bilingual communities such as the Spanish-English bilingual community living in northern New Mexico where codeswitching is remarkably balanced with respect to switch direction and the nearly even distribution of unilingual English and unilingual Spanish clauses, we would expect similar rates of noun insertions in mixed noun phrases (e.g., *el shoe*, *the zapato*). However, analyses of these bilinguals' spontaneous speech reveal that they disproportionately opt for inserting English nouns when speaking Spanish over inserting Spanish nouns when speaking English (Torres Cacoullos et al., 2022). Furthermore, English noun insertions into Spanish include high frequency words such as kinship terms, suggesting that language-use norms, coupled with the control demands of bilinguals' interactional context also play an important role (Green & Abutalebi, 2013; Green &Wei, 2014; Beatty-Martínez et al., 2020b).

3.1.2 The Triggering Hypothesis

The *Triggering Hypothesis*, introduced in Clyne (1967), proposes that when bilinguals using one language access a word whose language-specific membership is ambiguous, such as homophones (e.g., *fin* meaning *end* in Spanish; or *pan* meaning *bread* in Spanish), proper nouns (e.g., *Walmart*; *Target*), borrowed words (*taco, tortilla, queso*), or lexical transfers, the raised level of activation of the language not in use sometimes precipitates (i.e., triggers) a switch into that language. Cognate words are thought to be predilect triggers due to their overlapping phonology, orthography, and meaning in the bilinguals' languages. In the trilingual Spanish-Italian-English example below, *disprezzare*, an Italian cognate with Spanish (*despreciar*) and English (*despise*—although the valence in English is more negative than in Italian or Spanish) is interpreted as triggering the codeswitch into English in (7):

7. *no porque quiero* disprezzare *a mi language, Italian*
not because (I) want to despise (to) my language Italian
"Not because I want to belittle my language, Italian"
Clyne (1997:109)

Testing the predictions of the triggering hypothesis has been the object of many recent studies (Broersma, 2009, 2011; Broersma et al., 2009, 2020; Broersma & De Bot, 2006; Bultena et al., 2015; Fricke & Kootstra, 2016; Gullifer & Titone, 2019; Li & Gollan, 2018; Neveu et al., 2022), with corpus data generally confirming the privileged status of cognate words as triggers to codeswitches. For example, Broersma and De Bot (2006) report that codeswitches occur more frequently directly after a cognate or a proper name, and Broersma et al. (2020) found that codeswitches occur more frequently in clauses containing cognates than in clauses without cognates. These authors also found that the number of cognate words within a clause increases the likelihood of codeswitches downstream (see Van Hell, 2023, for a review). These findings suggest that language switches are more likely to occur when the target word is relatively more accessible to the speaker. However, recent evidence has challenged the conclusion that certain words serve as triggers for codeswitches. At stake is the role of single-word insertions in mixed speech. Past studies that have classified single word insertions as codeswitches have found support for the triggering hypothesis. However, some researchers argue for the need to distinguish between single word insertions and multi-word codeswitches (e.g., Poplack et al.,1988). When this distinction is made, rates of codeswitching do not vary according to the presence or absence of trigger words; instead, codeswitches co-occur with codeswitches more than with trigger words (Trawick, 2021), leading to the hypothesis that perhaps priming promotes codeswitching.

3.1.3 Priming effects

Within a dominant language, some sentence structures are more frequent than others. For example, dominant English-speaking adults are more likely to use the active sentence 'The chairman is suggesting a compromise' over the passive sentence 'A compromise is being suggested by the chairman' (examples taken from Bock, 1986: 383). In a classic study, Bock (1986) showed that it was possible to override the bias for active sentences if a speaker first heard and then repeated a sentence spoken in passive voice. This syntactic repetition effect, referred to as *syntactic priming*, is thought to serve several cognitive and social functions, among which are promoting fluent language production (MacDonald, 2013) and coherence during social interactions (Pickering & Garrod, 2004), supporting the learning of syntax (Dell & Chang, 2014), and promoting and guiding predictive language processing (e.g., Dell & Chang, 2014, Pickering & Gambi, 2018).

In the bilingual literature, scholars have asked whether languages that share syntactic structures lead to cross-linguistic priming. To illustrate, Hartsuiker, Pickering and Veltkamp (2004) examined whether English passive sentences were produced more frequently following a Spanish passive than a Spanish intransitive or active sentence. The results showed that cross-language priming effects were similar to the within-language priming effects, leading Hartsuiker et al. to the proposal that the syntactic computations for each of the bilingual's two languages make use of the same abstract syntactic-level information. This finding begs the question of whether priming might be a mechanism that helps explain why codeswitching occurs.

In their analysis of naturally produced speech from the Bangor Miami corpus (Deuchar et al., 2014), Fricke and Kootstra (2016) found that signature priming effects first described in studies with monolingual speakers were present in bilingual codeswitched discourse. Of note in their study, they found that bilingual speakers were more likely to codeswitch when the immediately preceding utterance also contained a codeswitch. They also observed that lexical overlap boosted the tendency to codeswitch, consistent with past experimental work showing that cross-language priming effects are generally stronger when there is lexical overlap and shared word order across languages (Kootstra et al., 2010). In addition, a speaker's tendency to codeswitch is primed more by their own speech (within-speaker priming) than by the speech of others (between-speaker priming or comprehension-to-production priming). In a recent lab-based study, Kootstra, Dijkstra, and Van Hell (2020) used the confederate-scripted priming task, where a naïve participant engages in a referential communication task with a lab member who presents as another participant, to investigate codeswitching behavior in a group of Dutch-English bilingual speakers. Analyses of a picture description task showed codeswitching was more common when the confederate had codeswitched in the preceding trial; in addition, the likelihood of a codeswitch was even greater when a cognate was also present, demonstrating the impact of cross-language priming on codeswitching (see also Torres Cacoullos & Travis, 2018).

While the studies reviewed in this section underscore the impact that bottom-up processes have on codeswitching behavior, it is also important to recognize bilinguals most often produce unilingual speech. The estimated rate of codeswitches in the New Mexico Spanish-English bilingual corpus is 11% (Trawick, 2021: 82), and Fricke and Kootstra (2016) note that only about 5.8% of the utterances in the Bangor Miami corpus contains intra-sentential codeswitches. Leaving aside the specifics of how the percentages were calculated (a matter that is not trivial), these figures strongly indicate that despite high levels of cross-language activation, priming effects—no matter how robust they might be—are insufficient to account for how bilinguals use their two languages. Instead, a speaker's communicative intentions exert top-down control over codeswitching practices to achieve communicative objectives, "opportunistically threading together words and phrases from each language in order to convey the intended meaning" (Beatty-Martínez et al., 2020a: 2; cf. Fricke & Kootstra, 2016: 195).

Studies examining speech articulation in bilingual speakers provide some evidence of a certain degree of cross-language effects adjacent to a codeswitch. When English and Spanish have been the languages under investigation, voice onset time has been the preferred target of study, given obvious differences between the two languages. Voice onset time (VOT) refers to the temporal lag of vocal fold vibration (voicing) following the occlusion produced in a stop consonant (e.g., /p/, /t/, /k/). Spanish VOT of voiceless stops is characterized by a short lag of approximately 0–30 milliseconds (Lisker & Abramson, 1964; Abramson & Lisker, 1973), whereas English VOT has a lag duration between 30 and 90 milliseconds. In this respect, Spanish is considered a 0-lag language because voicing begins shortly after the release of the voiceless stop, whereas English is considered a positive-lag language because in certain phonological contexts some time passes between the release of the stop and voicing, often realized as a puff of air (e.g., Spanish *pata* [pata], English *pot* [p^{ha}t]).

Examining spontaneously produced speech, Balukas and Koops (2015) found interesting asymmetries in VOT values between English and Spanish for words in close proximity of a codeswitch. Whereas Spanish words near a codeswitch did not result in a significant increase in Spanish VOT values in the direction of English, the converse was true such that the VOT of English words in the vicinity of a codeswitch into Spanish were significantly reduced (i.e., more Spanish-like). Fricke, Kroll, and Dussias (2016) similarly found that words in the proximity of codeswitches showed slower speech rate and exhibited cross-language phonological influence on consonant VOT (Bullock, 2009; Olsen 2013).

These changes in the properties of acoustic features in bilingual speech could be viewed either as a consequence of the cognitive demands associated with codeswitching, which result in the unintended activation of the non-target languages, or as processes related to opportunistic planning in anticipation of producing an intended codeswitch (Beatty-Martínez et al., 2020a; Beatty-Martínez et al., 2020b). To shed light on this question, Johns and Steuck (2021) conducted analyses of the codeswitched speech in the New Mexico Spanish-English bilingual corpus, couching their hypothesis on the premises of MacDonald's (2013) Production-Distribution-Comprehension (PDC) model. The PDC model proposes that production is shaped by a combination of an individual's linguistic experience and by cognitive limitations on production: production is difficult, but our own experiences as a language user, along with the cognitive strategies speakers employ, can alleviate this difficulty. One such strategy is the *easy first bias:* easier-to-produce features of language will tend to occur before harder-to-produce ones in a production episode. The goal is to 'buy time' for the production system to plan the more difficult elements without disrupting the fluidity characteristic of spontaneous speech. Johns and Steuck reasoned that if switching languages is cognitively more demanding than staying in one language, then it should occur later in a production episode, which they confirmed in their analysis. Importantly, while this may suggest that codeswitching is difficult to plan and produce, an alternative interpretation is that codeswitching could serve as one of several strategies that bilinguals use to alleviate difficulties when they arise. To test this, they further examined speech rates in the codeswitched utterances and compared them to carefully matched Spanish- and English-only production from the same speakers. They found that speech rates surrounding a codeswitch were significantly *faster* compared to unilingual production. Thus, codeswitching may be more likely to occur when production gets difficult, however, not as the source of the difficulty but as a mechanism to alleviate it. In other words, codeswitching is one of various tools that bilingual speakers employ to facilitate effortful speech planning and production (cf. Sarkis & Montag, 2021).

This section has focused on discussing findings that shed light on a question that has been of great interest to scholars, parents, and teachers alike: why do bilinguals codeswitch? We have provided a cursory overview of the research that contributes to our understanding of the bottom-up factors that contribute to the production of a codeswitch and have also presented arguments to show that these factors are insufficient to explain the overall pattern of data available to date. Ultimately, bilinguals codeswitch "because they can."

3.2 What do naturally produced codeswitches look like?

One important finding in the codeswitching literature is the presence of asymmetries in the production of codeswitched speech. Earlier we discussed asymmetries in the phonetic realization of VOT when voiceless stop consonants in Spanish and English are in the vicinity of a codeswitch. Here, the focus will be on different a type of asymmetry: one that arises from bilinguals' syntactic choices at particular codeswitching sites. We will limit our discussion to two examples often cited in the literature of Spanish-English codeswitching, a pair of languages that has been examined extensively in codeswitching studies.

Several studies examining the production of Spanish-English intrasentential codeswitches (e.g., Lipski, 1978; Poplack, 1980) have alluded to an asymmetry involving alternations within the auxiliary phrase. Specifically, when the direction of the codeswitch is from Spanish into English, switches in which the Spanish auxiliary verb *estar* (English *be*) is followed by an English present progressive verb (examples (8) through (11) below) are produced with similar frequency as switches at the verb phrase boundary (e.g., example (12) below):

8. *Mi marido está working on his Master's (Lipski, 1978: 252)*
"my husband is working on his Master's"
9. *Los están bussing pa' otra escuela (Pfaff, 1979: 296)*
"(They) are bussing them to another school"
10. *Estaba training para pelear (Pfaff, 1979: 296)*
"(He) was training to fight"

11. *Siempre está promising cosas.* (*Poplack, 1980*: *596*)
"(He) is always promising things."

12. *El niño is trying to catch mariquitas. (Rodrigo & Dussias, in prep)*
"The boy is trying to catch ladybugs."

In contrast to *estar/be* + English verb switches, switches in which the Spanish auxiliary *haber* (*to have*) is followed by an English participle verb form are rare in Spanish-English codeswitching corpora. One of the few cases found in production data is shown in (13):

13. Yo creo que apenas se había *washed* out. (*Pfaff, 1979*: *300*)
"I think that it had just washed out."

Guzzardo Tamargo, Valdés Kroff, and Dussias (2016) examined available oral and written Spanish-English codeswitching corpora to confirm the distributional patterns involving switches at these two syntactic sites. The oral corpus was the Bangor Miami Corpus of Spanish-English bilingual speech mentioned earlier (Deuchar et al., 2014). Twenty-six transcriptions (approximately 390,000 words) were analyzed. The written corpus was extracted from a weekly editorial column of a bilingual newspaper published in Gibraltar. Eighty-eight editorial codeswitched column entries (approximately 25,300 words) were examined. Despite differences in modality between written and spoken language, the data extracted from the two corpora confirmed the existence of asymmetric patterns.

A second well-documented asymmetry observed in virtually every corpus that has been examined to date concerns mixed noun phrases (or mixed NPs). Corpus analyses have reported the systematic and widespread use of gender-marked masculine determiners (*el*/the$_{\mathrm{MASC}}$; *este*/this$_{\mathrm{MASC}}$; *muchos*/many$_{\mathrm{MASC}}$) with English nouns whose Spanish translation equivalents are marked with masculine or feminine grammatical gender (See (14) and (15) below). However, codeswitches involving gender-marked feminine determiners (*la*/the$_{\mathrm{FEM}}$; *esta*/this$_{\mathrm{FEM}}$; *muchas*/many$_{\mathrm{FEM}}$) have been observed to occur less frequently and are restricted to English nouns whose translation equivalents are marked with feminine grammatical gender (see (16) and (17) below). This asymmetry differs from that in unilingual contexts, where masculine and feminine nouns are evenly distributed (Eddington, 2002), suggesting a codeswitching strategy that results from the interaction between the two languages.

14. *En el* winter
"In the winter"

15. *El* flag
"the flag"

16. *con la* flashlight *en una mano*
"with the flashflight in one hand"

17. *We went to that floating garden*
que hay muchas *flowers*
"that there are many flowers"

Why are asymmetries important? First, as discussed in Beatty-Martínez et al. (2020a), bilingual speakers use their languages in different ways; not all syntactic sites participate equally in codeswitching, just like not all contexts of language use incur the same cognitive demands during the production of codeswitched language. The task at hand is to uncover "… of the places where bilinguals can switch, where they prefer to do so" (Torres Cacoullos & Travis, 2018: 175). Second, mounting evidence has shown an intimate connection between production patterns and comprehension difficulty. One influential model of monolingual language processing that has sought to capture this link is MacDonald's (2013) PDC model, mentioned earlier. Briefly, the PDC proposes that cognitive limits on language planning and production, such as those related to memory and retrieval, shape the distributional properties of language. If indeed comprehension skill is tied to individuals' production preferences, documenting codeswitching asymmetries in production choices is important because distributional patterns in the production of code-switches lead to the clear prediction that codeswitching bilinguals should demonstrate asymmetric processing patterns that do not match unilingual processing. We take up this issue again in Section 5.1 below.

4. The real-time processing of codeswitched speech

A cornerstone finding in sentence processing is that the human sentence parser is incremental and predictive, continuously and dynamically updating interpretations as new linguistic input unfolds. The parser builds interpretations guided by phonological, lexical, syntactic, and semantic information, as well as extra-linguistic information such as speaker, setting, or pragmatic context (e.g., Altmann & Kamide, 1999; Garnsey et al., 1997; Trueswell et al., 1999). The field continues to hold important debates in terms of when linguistic and extra-linguistic information come online to aid parsing, as well as whether prediction is lexically specific or more probabilistic (Kuperberg & Jaeger, 2016). Nevertheless, there is strong evidence for the observation that speakers of a first or dominant language can anticipate speech under certain conditions. Meanwhile whether the same characteristics are observable in speakers of a second, late acquired language remains more debatable, but recent work reveals the capacity to efficiently anticipate upcoming linguistic input under certain constraints and with sufficient proficiency in the second language (Kaan & Grüter, 2022).

To demonstrate incremental parsing, we discuss a classic study on verb bias (i.e., a verb's preferential subcategorization frame) by Garnsey et al. (1997). Verbs such as "discover" and "believe," as in examples (18) and (19) below, appear in different structural configurations. Both verbs can optionally select a direct object complement (18a, 19a) or sentence complement (18b, 19b). Yet examination of different production corpora coupled with sentence completion studies indicate that some verbs are more likely to co-occur with specific structures. Thus, in English, "discover" is often followed by a direct object (18a) but "believe" is more often followed by an embedded sentential complement (19b; examples taken from Garnsey et al., 1997):

18a. *The scuba diver discovered the wreck behind the hidden coral reef.*
18b. *The scuba diver discovered the wreck was caused by a collision.*
19a. *The job applicant believed the interviewer despite her pointed questions.*
19b. *The job applicant believed the interviewer was dishonest with her.*

Of note, this frequency difference is, in part, language-specific (i.e., translation equivalents across languages may show different frequency co-occurrences, Cuetos & Mitchell, 1988; Dussias et al., 2010).

According to constraint-based or experiential theories of sentence processing, such distributional differences affect online sentence processing at the earliest moment. Such theories predict that L1 speakers of English will demonstrate differential processing speeds based on whether the structural configuration matches the speaker's expected verb bias. In line with such predictions, Garnsey et al. (1997) showed that in 18b, participants initially *garden path* (Bever, 1970) interpreting "the wreck" as a direct object, leading in turn to slower reading times when encountering the embedded verb "was caused." In contrast, participants did not encounter this slowdown in (19b), presumably because the post-verbal noun phrase was assigned the role of sentential subject. This finding has subsequently been confirmed using other methodologies such as EEG (Román et al., 2022) and for L2 readers who have familiarity with a verb's distributional bias (Dussias & Cramer Scaltz, 2008).

This line of research demonstrates two important points. First, during sentence processing, humans exploit probabilistic information from their linguistic input and use it to generate expectations about upcoming input (e.g., Gahl & Garnsey, 2006; MacDonald, 2013). Second, the observation that humans experience delayed processing or garden-path effects strongly suggests that comprehenders do not wait to reach the end of a clause to begin to build interpretations, instead making commitments incrementally while new information becomes available. When comprehenders encounter incorrect initial interpretations, they revise them based on later-arriving information.

The fundamental characteristic of incremental parsing in human sentence processing raises further issues when considering codeswitching. As codeswitching involves alternations from one language into another at variable syntactic positions, its use entails the constant coordination of two grammars at multiple linguistic levels (e.g., phonology, morphology, syntax, semantics). This coordination across grammars potentially introduces greater ambiguity into the linguistic signal. It may also momentarily lead to representational conflict if the produced codeswitching structure differs from unilingual structures (we discuss this point in more detail below). Moreover, the psycholinguistics of bilingualism robustly finds that bilingual lexical access and processing is non-selective; even when bilinguals intend to stay in one language alone, their non-target language remains co-active to varying degrees, and only under very limited circumstances is it possible to completely "shut off" the other language (Kroll & Gollan, 2014; Kroll et al., 2015). Consider the logical alternative: for comprehension, it should be simpler to remain in the same language within a conversation than to introduce additional ambiguity into the speech signal by codeswitching. Nevertheless, we know that bilinguals choose to ubiquitously codeswitch under the right pragmatic contexts and that such practice does not visibly result in confusion or a breakdown to comprehension. The question, then, is how?

Codeswitching is presumably under the speech planning control of the speaker, as discussed in Section 2; however, there are potentially no clear, explicit signals to the comprehender that a codeswitch may occur, nor that the switch may potentially induce conflict across the two languages. We demonstrate this potential conflict in example (20), taken from the Bangor Miami Corpus (Deuchar et al., 2014):

20. pero no tenían el *flag out there? [sastre 09]*[c]

but NEG have-3PL.IMP the-MASC flag out there

"but didn't they have the flag out there?"

[c] The name in brackets is the corresponding file in the Bangor Miami corpus from which the example is extracted, see https://biling.talkbank.org/access/Bangor/Miami.html.

The utterance was made without any apparent pauses or hesitations by the speaker; the switch occurred at the Spanish definite masculine article *el* and then continued into English. This example highlights the potential challenges for the comprehender. One well-known cross-linguistic difference between Spanish and English is the presence of grammatical gender in Spanish and its absence in English. All Spanish nouns are classified in a binary grammatical gender category (feminine, masculine), and modifying elements such as determiners and adjectives must agree in grammatical gender with the corresponding noun. Psycholinguistic studies have demonstrated consequential effects of this characteristic of gendered languages for sentence processing. When processing a gender-marked pre-nominal word such as an article, listeners anticipate the gender of the upcoming noun (Dussias et al., 2013; Hopp, 2016; Lew-Williams & Fernald, 2007, Morales et al., 2016; Paolieri et al., 2020). This effect, known as the grammatical gender congruency effect (see Beatty-Martínez & Dussias, 2019 for review), leads to bilinguals experiencing significant delays in processing when encountering a noun that does not agree in gender with the preceding element (Barber & Carreiras, 2005). Consequently, pre-nominal elements that overtly mark grammatical gender serve as predictive or facilitatory cues in sentence processing.

Returning to Example (20), there are two reasons to assume that switching at this site ought to give rise to processing difficulties. First, the robustness of the gender congruency effect presupposes a strong dependency between Spanish articles and Spanish nouns. Therefore, an English noun after a Spanish article should be an unexpected event. A second, more complex observation is that the Spanish translation equivalent of *flag* is a feminine noun, *bandera*. Consequently, it conflicts with the preceding gender-marked definite article, *el*, which is masculine. This and other similar examples should be problematic during online processing for Spanish-English bilinguals. Essentially, if Spanish-speaking individuals commit to Spanish-like processing, then they should expect Spanish masculine nouns to follow masculine articles. Nevertheless, this specific type of switch is highly frequent among Spanish-English bilingual speakers and within certain bilingual communities (e.g., Poplack, 1980; Torres Cacoullos & Travis, 2018) and does not saliently appear to disrupt communication.

As will be discussed in greater detail below, experience with codeswitching can lead to changes in the time course of how these structures are processed (e.g., Valdés Kroff et al., 2017). This specific example illustrates the need for bilinguals to avoid costly garden paths in codeswitching contexts, especially when these contexts differ in numerous ways from unilingual contexts. Our claim is that bilinguals adapt and shift the cues that they rely upon during sentence processing, a phenomenon that we term *adaptive predictability*. Adaptive predictability stems from the same mechanisms operant under constraint satisfaction accounts, thus extending these notions into bilingual contexts.

To further underscore the seemingly challenging demands that face the bilingual comprehender, lab-based studies on sentential codeswitching find apparent switch costs to comprehension. Using time-based measures such as reaction times (e.g., self-paced reading), fixations and regressions (eye-tracking), pupil dilation (pupillometry), and neurocognitive measures (e.g., event-related potentials, ERPs), encountering a switch into another language can lead to a slowdown or an unexpectancy in processing (e.g., Altarriba et al., 1996; Moreno et al., 2002; Bultena et al., 2015; Johns & Dussias, 2022; Litcofsky & Van Hell, 2017). However, some studies show that these switch costs can be attenuated under certain circumstances, such as when the codeswitch is more frequent (e.g., Beatty-Martínez & Dussias, 2017; Guzzardo Tamargo et al., 2016; Kheder & Kaan, 2019; Salig et al., under review; Valdés Kroff et al., 2017, 2020), when taking into account language dominance and direction of the switch (e.g., Litcofsky & Van Hell, 2017; Fernández et al., 2019), or when considering the ecological validity or the pragmatic context underlying the experimentally presented codeswitch (Blanco-Elorrieta & Pylkkänen, 2017; Johns et al., 2019; Kaan et al., 2020; Tomić & Valdés Kroff, 2022; Tomić & Kaan, 2022. See Beatty-Martínez et al., 2018; Blanco-Elorrieta & Pylkkänen, 2017; Valdés Kroff et al., 2018 for extended discussions on these issues and why they may arise). These findings more broadly align with how humans engage with general switching phenomena. In cued language switching tasks where participants are asked to name a picture or digit in which a language is externally signaled via a visual or auditory cue (e.g., color of background screen), bilinguals are slower to name on switch trials as compared to same-language trials (e.g., Meuter & Allport, 1999). Similarly, in studies where participants are asked to indicate a response that switches between two dimensions, such as indicating the color or shape of a figure, participants are slower to respond on trials that switch between tasks relative to responding to the same task (e.g., Monsell, 2003). Yet there are also appreciable differences between these highly controlled externally cued languages switching paradigms and the naturalistic switching that occurs within bilingual speech. For one, switch costs in cued language switching are robust in production but are weak or non-existent in comprehension (Declerk et al., 2019; Declerck, 2020), while the opposite appears true for sentential codeswitching (Altarriba et al., 1996; Moreno et al., 2002; Beatty-Martínez et al., 2020a). Second, cued language switching studies are mostly focused on lexical switching and overwhelmingly represent one grammatical category, nouns, while sentential code-switching occurs at a variety of syntactic switch junctures and can involve multiple words. Thus, even though we can draw certain inferences between cued language switching and sentential codeswitching paradigms, the extent to which the same cognitive and neural processes are involved in both phenomena is not yet well-understood.

In this section, we have discussed how incremental parsing and codeswitching highlight an apparent paradox in that it should be costly for comprehension, yet its ubiquity in bilingual and multilingual speakers suggests that it is not. We argue that these contrasting findings require a different approach to how we consider these issues, one that focuses on how bilinguals better prepare the comprehension system for bilingual language use. In the next section we outline the Adaptive Predictability hypothesis.

5. Adaptive prediction in codeswitching

We propose the Adaptive Predictability hypothesis to account for the real-time integration of codeswitched speech in online comprehension. The hypothesis is composed of two core premises. First, bilinguals adapt the way in which they process bilingual speech, especially how they predict upcoming linguistic information, because of accumulated experience with codeswitched speech (i.e., adaptability). Second, cognitive control is the primary domain-general cognitive mechanism that supports rapid integration of other-language information in online comprehension. Underlying the hypothesis is the assumption that codeswitching is a highly skilled speech act that one acquires through experience that is not immediately accessible to all bilinguals in online processing. Our hypothesis is an extension of experience-based accounts that considers ways in which ongoing language experience shifts the cues that comprehenders rely upon to not only guide prediction but also actively not to predict in circumstances that would otherwise lead to non-optimal processing of bilingual input.

5.1 Adaptive prediction

Bilinguals demonstrate adaptability in codeswitching by shifting the weighting of different sentence processing cues. For prediction in sentence processing, this shifting entails adapting when to predict and when not to predict, even in contexts that are predictive in monolingual processing. This adaptive behavior, which results from cumulative experience, best prepares the comprehension system for a *possible but not guaranteed* codeswitch in upcoming speech. In other words, the bilingual listener must consider a trade-off between actively predicting or holding off on predicting upcoming information. The intention to delay active prediction becomes an optimal and efficient strategy for the bilingual comprehender if it aids in avoiding disruptions arising from predictive commitments that ultimately are wrong (i.e., predicting Spanish-like processing when instead, a codeswitch into English occurs). This adaptation will require dynamic changes from those cues that facilitate monolingual sentence processing, thus fundamentally making bilingual sentence processing qualitatively different on the surface, even when comparing the dominant language of the bilingual. Yet, the mechanisms that bilinguals rely upon are the same as those that monolinguals recruit; simply put, the language input bilinguals encounter blurs the line between languages, speakers, and contexts thus leading to varied experiences from monolinguals and even from other bilinguals.

At the phonological level, we see evidence for facilitated processing arising thanks to the existence of cross-linguistic phonetic differences that can signal to the listener an impending codeswitch on the part of the speaker. Corpus and experimental studies have demonstrated that subtle shifts in VOT can occur before a codeswitch (Balukas & Koops, 2015; Bullock, 2009). Recent research has shown that these shifts are used by bilingual listeners, along with a host of other cues (e.g., slower speech rate, disfluencies, cross-linguistic differences in the permissibility of complex consonant clusters) to anticipate an upcoming codeswitch (Fricke, et al., 2016; Li, 1996; Olson 2013; Shen et al. 2020). Whether these sub-phonemic and speech rate differences are the result of articulatory pressures or are a pragmatic means of signaling to the listener an impending codeswitch remains an open question; however, because these physical properties can uniquely differentiate between unilingual and bilingual speech, listeners are able to exploit them to aid comprehension. We find similar patterns in monolinguals and the inferences that they make when speakers are disfluent. Listeners interpret disfluencies such as "uh" and "uhm" as signals that new or less frequent information is upcoming (Arnold et al., 2003, 2004, 2007) but only do so when the speaker is a dominant speaker and not when an L2-accented speaker produces similar disfluencies (Bosker et al., 2014).

At the morphosyntactic level, bilingual speakers capitalize on codeswitching asymmetries that surface in production (see Section 3.2). Grammatical notions of constituency, clause boundary, and congruence are likely to play a role in determining common versus less common codeswitches. For example, both children and adults are sensitive to the processing cost of inter- v. intra-sentential codeswitching, as reflected by pupillometry (Byers-Heinlein et al., 2017), an implicit measure of the diameter of the pupil that serves as an index of cognitive or processing effort. On the other hand, in structurally similar constructions (such as within the noun phrase in Spanish and English), community practices may instead lead to distributional asymmetries, which in turn, should lead to differential processing of these structures (Beatty-Martínez & Dussias, 2017; Valdés Kroff et al., 2017). To further uncover these grammatical asymmetries, additional work triangulating between corpus analyses, speaker intuition and production, and experimental-based tasks will be necessary (Beatty-Martínez et al. 2018; Valdés Kroff et al. 2018). Here, we claim that morphosyntactic cues tune the bilingual parser to delaying active prediction because likelier or more frequent codeswitch junctures lead to increased ambiguity as to whether an utterance will continue in the same language or codeswitch. In essence, experience with codeswitching can guide listeners to optimally reweight processing cues that lead to actively predicting or not under appropriate contexts.

Other extra-linguistic factors such as speaker and context are also probable cues for the listener. Individual- and community-level factors such as proficiency, codeswitching habits, and community practices can affect the frequency and type of codeswitching, which in turn contribute to adaptive prediction. Bilinguals who are less proficient in a second language are more likely to produce single word or inter-sentential codeswitches (Poplack, 1980; Zentella 1996). Similarly, the extent to which codeswitching is accepted by a community of speakers will further affect codeswitching production (Poplack, 1987). Listeners are guided by this same information, adapting to the environmental and pragmatic context of the setting and/or interlocutor (Blanco-Elorrieta & Pylkännen, 2017; Kaan et al., 2020). Using the visual world eye-tracking paradigm and ERPs respectively, Valdés Kroff et al. (2018) and Beatty-Martínez and Dussias

(2017) demonstrated group differences in the online comprehension of Spanish-English codeswitches involving mixed determiner phrases, showing that only bilinguals who had been exposed to community-specific codeswitching patterns (in this case, overwhelming preference for Spanish masculine articles in mixed noun phrases) also exhibited asymmetric processing patterns that reflected these production biases.

Altogether, we take these varied and intricate linguistic cues as indication that codeswitching is a skillful, interactive speech act. Although bilinguals have strong intuitions for major codeswitch violations occurring at clear, grammatical boundaries (e.g., codeswitches between subject pronouns and predicates are dispreferred; González-Vilbázo & Koronkiewicz, 2016; Koronkiewicz, 2020), only bilinguals who are exposed to more intricate forms of codeswitching (and their resulting asymmetries) will be able to rapidly process these same asymmetries in comprehension. Consequently, one further extension of the Adaptive Predictability hypothesis is that the acquisition of codeswitching constraints does not go hand in hand with bilingual acquisition. The development of sensitivities to preferred codeswitching structures will occur on a different trajectory relative to the development of either language and will be cumulative with experience. This means that the type of language mixing that bilingual children may demonstrate is not necessarily the same type of codeswitching that bilingual adults engage in (e.g., Ribot & Hoff, 2014). Consequently, codeswitching structures produced in bilingual adult speech may not be easily processed in child sentence processing. Initial indications support these claims, at least in the case of bilingualism in North America (Byers-Heinlein, 2013; Gross & Kaushanskaya, 2015; Quick et al., 2021; Smolak, de Anda, Enriquez, Poulin-Dubois, & Friend, 2020). Nevertheless, whether bilingual children acquiring both languages are also sensitive to adult bilingual codeswitching most likely depends on the extent to which codeswitching has broad community support (Myers-Scotton 1993) and where direct transmission from older to younger generations occurs (e.g., Aboh, 2020; Yow et al., 2016, 2018). Within such multilingual communities, children will develop sensitivity to asymmetric codeswitching patterns at a younger age, as their language mixing patterns are also more likely to mirror adult codeswitching within their community. More work directly comparing the processing of codeswitched speech between children and adults is a necessary path to continue to uncover similarities and differences and understand acquisition of codeswitching constraints.

5.2 Cognitive control

The prior section illuminates the multiple moving pieces and possible cues that could potentially signal the listener to impending codeswitches. However, the occurrence of a codeswitch is, of course, not guaranteed. In other words, these linguistic and extra-linguistic cues probabilistically signal an increased likelihood of a codeswitch, but the utterance may continue in the same language. Additionally, bilingual speakers may vary in their scope of planning an upcoming codeswitch, thus leading to moments when codeswitches are planned in advance or occur on the fly and at later stages of speech planning (Valdés Kroff, 2016; Johns & Steuck, 2021). Therefore, a second component to the Adaptive Predictability hypothesis is the upregulation of cognitive control to aid rapid integration of codeswitches in online processing.

Cognitive control is a domain-general executive function that is deployed to resolve representational conflict or to override prepotent biases or responses (Botvinick et al., 2001; Braver 2012). One model of sentence processing proposes that cognitive control deploys in contexts in which conflict resolution is necessary to override dominant interpretations that ultimately are not correct (Novick et al., 2005). Empirically, this framework has been supported by behavioral and neuroimaging work (e.g., Humphreys & Gennari, 2014; January et al., 2009; Zirnstein et al., 2018; Hsu & Novick, 2016) as well as by demonstrating that patients with left inferior frontal damage (a neural area involved in cognitive control) have greater difficulties in resolving garden-path syntactic ambiguities (Novick et al., 2009).

In parallel, we interpret the comprehension of bilingual codeswitching as linguistic contexts that bring bilingual grammars into competition to varying degrees. This competition can occur at multiple levels, with the most basic level being language membership. This competition means that barring any explicit signal, bilingual listeners need to override biases that utterances will continue in the same language. Many of the cues outlined in Section 5.1 make this task easier but may not eliminate it completely. At more complex levels, codeswitches can occur at syntactic sites that result in cross-linguistic conflict that needs to be resolved between the two grammars, as in the case of mixed NPs in Spanish-English codeswitching (Example 20). Consequently, cognitive control is the primary mechanism that helps to overcome same-language lexical and grammatical constraint biases, which aids listeners to shift to processing in the other language after encountering a codeswitch. The upregulation of cognitive control is likely to be gradient due to the varying degrees of conflict induced by codeswitching (cf. Green, 2018). Less cognitive control should be required for inter-sentential codeswitches as compared to intra-sentential codeswitches and especially in environmental contexts that highly support bilingual speech. More complex intra-sentential codeswitches or codeswitches that occur under pragmatically odd or unexpected contexts should require greater cognitive control for comprehension.

Initial evidence in support of this prediction comes from behavioral and neuroimaging work. For example, neural regions implicated in the cognitive control network such as the left inferior frontal gyrus and the dorsal anterior cingulate cortex are reliably recruited when bilinguals are engaged in language switching tasks or when processing auditory codeswitches (Abutalebi et al., 2008; Abutalebi & Green, 2008; Hernandez, et al., 2001; cf. Blanco-Elorrieta & Pylkkänen, 2017; Luk, Green, Abutalebi, & Grady, 2011). Additionally, in between-group designs directly comparing monolinguals and bilinguals or bilinguals with different linguistic profiles—those who are more likely to switch between their languages throughout the day or who engage in more intricate "dense" codeswitching—show reduced conflict effects on non-linguistic tasks requiring cognitive control such as the Stroop and Flanker tasks (Hofweber et al., 2016; Prior & Gollan, 2011).

5.3 Empirical illustrations of the adaptive predictability hypothesis

In this section, we describe two research studies in support of our hypothesis. Our descriptions will summarize overall findings and provide a brief discussion on how these studies illustrate the two central premises of the Adaptive Predictability hypothesis. For further technical details on the studies, readers should consult the original studies (Valdés Kroff, 2016; Valdés Kroff et al., 2017, 2018 for adaptation in predictive processing; Adler et al., 2020 for upregulation of cognitive control while reading codeswitched sentences).

Valdés Kroff et al. (2017) investigated how the purported asymmetry in mixed noun phrase production in east coast US and Puerto Rican Spanish-English bilingual communities affects online processing. Aside from the greater preference for mixed noun phrases to surface with a Spanish determiner and an English noun phrase (*el house* v. *the casa*), these bilinguals overwhelmingly produce Spanish masculine gender-marked determiners in mixed noun phrases (i.e., preferring *el house* v. *la house*, Sp. *la casa, feminine;* Beatty-Martínez & Dussias, 2017; Otheguy & Lapidus, 2003; Valdés Kroff, 2016). Infrequently, feminine-marked mixed noun phrases also are produced but are prohibitively restricted to English nouns whose Spanish translation equivalents are feminine (i.e., *la house* but **la juice*, Sp. jugo/zumo, masculine). These distributional patterns mean that the gender concord system in Spanish-English codeswitching does not simply follow the constraints in place for Spanish or fully neutralize grammatical gender agreement due to English contact. Yet despite its arbitrary assignment, the presence of obligatory grammatical gender in Spanish leads to consequences for the human parser; namely, it serves as a predictive cue for upcoming nouns. Subsequently, any shifts in how grammatical gender is used in codeswitching should have consequences for online processing.

The study employed the visual world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) in eye-tracking to test for asymmetric effects in online processing. Twenty-five bilingual participants were shown a visual scene consisting of two concrete objects while listening to auditory instructions indicating which object to select via computer mouse. Instructions were presented in Spanish only, single word codeswitches, or multi-word codeswitches. Crucially, critical trials consisted of pairs of items that are different gender in Spanish (e.g., *el carro/coche* "the car, masculine" v. *la galleta* "the cookie, feminine") because these trials provide linguistic contexts in which prenominal Spanish articles can be informative. In other words, participants can look towards target items more quickly due to the facilitatory information carried by the Spanish article on different gender trials (Lew-Williams & Fernald, 2007; Dussias et al., 2013). This anticipatory effect was symmetrically found in a control group of 24 Spanish monolinguals, thus confirming the experimental manipulation. The critical test came in the case of Spanish-English bilinguals, since in codeswitched speech masculine articles no longer uniquely signal the upcoming presence of masculine items either in Spanish or in the translation equivalent of the corresponding English target item. Consequently, we predicted an asymmetric processing pattern for the codeswitched trials, whereby bilinguals would continue to exhibit predictive processing towards feminine translation equivalent English target items when hearing the feminine article but would wait until after the onset of target nouns after hearing the masculine article. The results confirmed this prediction, thus indicating that online processing had changed adaptively due to how grammatical gender is used in production.

However, we also had an at-the-time unexpected finding: bilinguals exhibited the same asymmetric processing pattern in the Spanish-only session. Because the monolingual group had shown a symmetric anticipatory effect for both masculine- and feminine-marked articles, we isolated this unexpected finding to the bilingual group. We conducted a post-hoc naming test on the experimental materials by a subset of the participants about 6 weeks after the experiment which confirmed that the asymmetric processing pattern observed in the Spanish-only session was not due to overuse of the masculine gender in production (i.e., due to loss of grammatical gender representation; Montrul et al., 2008). Therefore, we reasoned the adaptation that occurs in the bilingual comprehension system is one that reflects a trade-off between predictive processing and intentional delay in the service of avoiding a costly garden path. In the case of the Spanish grammatical gender system, this means that Spanish masculine gender on prenominal agreement elements will no longer be predictive globally including in Spanish-only contexts because this trade-off best serves the bilingual comprehension system. While Spanish masculine articles continue to uniquely signal upcoming masculine Spanish nouns in Spanish, they may additionally be linked to switches into English and may include English nouns whose translation equivalents are feminine. Rather than having to calculate at any given moment whether an upcoming noun will continue in Spanish, adapting to an intentional delay leads the bilingual comprehension system to be better prepared for a possible codeswitch that may ultimately conflict with Spanish grammatical gender agreement constraints. In contrast, feminine articles continue to uniquely signal upcoming feminine nouns, whether in Spanish or as English translation equivalents; consequently, participants continue to utilize feminine gender as a facilitatory cue. The critical point is that what on the surface may seem like a failure to predict is instead a global shift to actively not predict as an optimal processing strategy.

Adler et al. (2020) directly tested the hypothesis that reading codeswitched language leads to upregulation of cognitive control. The study takes direction from recent work by Novick and colleagues demonstrating cross-modal conflict adaptation in monolingual speakers (Hsu & Novick, 2016; Kan et al., 2013). The basic idea behind conflict adaptation is that participants are more efficient on "harder" trials that require cognitive control after having just triggered cognitive control mechanisms, a phenomenon known as conflict adaptation (Botvinick et al., 2001). For example, Kan et al. (2013) demonstrated that after monolingual readers encountered a garden path from processing temporarily ambiguous sentences (e.g., *Put the frog on the napkin in the box* where *on the napkin* is a modifying prepositional phrase instead of an intended location), they were more efficient (i.e., faster) on an immediately following incongruent Stroop trial. This logic was used to test whether reading codeswitched sentences similarly led to conflict adaptation on an immediately following Flanker task; such a finding would lend support to the hypothesis that encountering codeswitches in comprehension engages cognitive control.

Forty-eight Spanish-English bilingual participants either read Spanish or English unilingual sentences or Spanish-English codeswitched sentences via a non-cumulative moving window self-paced reading paradigm (Just et al., 1982). The sentences were embedded in pseudorandom sequences with a classic Flanker task where participants were instructed to indicate the direction of a central arrow flanked by two arrows on each side. On congruent Flanker trials, the central arrow points in the same direction as the other arrows, whereas on incongruent trials the central arrow points in the opposite direction. Participants respond more slowly and less accurately on incongruent trials in classic Flanker tasks.

Pseudorandomization led to critical trial sequences in which sentence reading was immediately followed by Flanker trials. If participants were indeed upregulating cognitive control while reading codeswitched sentences, then immediately following incongruent but not congruent Flanker trials should show a reduced interference cost as compared to reading unilingual sentences. The results indeed showed an interaction on reaction times between language (unilingual, codeswitched) and Flanker trial type (congruent, incongruent). Importantly, the interaction was not driven by language dominance. Additionally, the effects of conflict adaptation were stronger in the second half of the experimental session, indicating that recruitment of cognitive control while processing codeswitched text does not diminish over time.

While these studies are only illustrative, and more work needs to be carried out to test the two key predictions of the Adaptive Predictability hypothesis, the hypothesis is a first attempt to make sense of the apparent contradiction that results from increasing ambiguity in bilingual codeswitching not leading to massive disruption to comprehension. This hypothesis extends experience-based accounts in sentence processing to bilingual contexts by globally considering the diverse linguistic experiences of bilinguals. Taken together, the adaptive behavior that bilinguals demonstrate during sentence processing and the upregulation of cognitive control suggest that switch costs that are found during comprehension may simply reflect an optimal strategy deployed by bilinguals to avoid disruptive garden paths. This framing moves away from the deficit framing that is prevalent in bilingualism and L2 studies and invites researchers to turn their focus from interpreting switch costs in comprehension as indices of integration difficulty to being an indication of a complex and skillful trade-off between prediction and garden path avoidance.

6. Conclusions

We have provided a broad examination of bilingual codeswitching, reviewing its planning and production and its comprehension in sentence processing; further, we have proposed the Adaptive Predictability hypothesis—an extension of experience-based or constraint-satisfaction approaches in bilingual contexts. Our goal has been to argue that codeswitching is a skillful speech act that bilinguals frequently engage in and to demonstrate how its careful investigation in production and comprehension can reveal the highly dynamic and adaptive mind of humans in language processing and use. There are a myriad of reasons and constraints underlying a speaker's choice to codeswitch. These constraints often conspire to lead to likelier codeswitch junctures in speech, which in turn comprehenders can capitalize on. Nevertheless, codeswitching does not occur in a vacuum, and both speakers and listeners must be prepared for codeswitches to be embedded within larger stretches of speech that continue in a single language.

Our goal in outlining the Adaptive Predictability hypothesis is to provide a psycholinguistically plausible framework for how comprehenders rapidly integrate codeswitched speech without suffering from delayed processing or disruptions to comprehension. Bilinguals learn to adapt how they predict in sentence processing guided by distributional regularities that arise in bilingual language use. Exposure to bilingual language use thus leads bilinguals to shift and reorder the weighting of the cues that they rely upon to facilitate comprehension. This results in a more optimized system, and in some cases leads codeswitching bilinguals to no longer rely upon predictive cues that are typically used during monolingual processing. While this adaptive behavior can help comprehenders avoid costly garden paths, codeswitching heightens conflict between the two languages to varying degrees. Subsequently, the upregulation of cognitive control further supports the rapid integration of codeswitched speech in comprehension.

Why propose a seemingly new hypothesis that appears to be subsumed by constraint-satisfaction approaches that already account for monolingual processing? Our observation is that the monolingual and second language literatures on sentence processing have been primarily focused on what prediction is (e.g., Kuperburg & Jaeger, 2016), debates about whether prediction is necessary (e.g., Huettig & Mani, 2016), and on the cues and contexts that could trigger prediction (Kaan & Grüter, 2022). Indications of failure to predict are often attributed to traits at the participant level, whether it be due to literacy, proficiency, or availability of domain-general cognitive resources. Codeswitching and other instances of bilingual language use open new avenues of interpretation—with implications for monolingual sentence processing—by revealing instances in which not predicting is, in fact, an optimal and more efficient choice. Rather than being tied to individual traits, adaptive prediction is a reflection of humans' susceptibility to shifting linguistic contexts and how these contexts trigger changes in the comprehension system. Codeswitching brings these changes to the surface and demonstrates that our experiences do not just trigger predictive mechanisms but highly affect how sentence processing mechanisms come into play.

Several issues remain to be addressed that will help (dis)confirm the validity of the Adaptive Predictability hypothesis. For one, given the highly experiential basis of the hypothesis, we need a clearer understanding of the transmission of codeswitching between adults and children. Are children sensitive to the same distributional asymmetries as adults are? Will such sensitivities depend on community-level support for bilingual codeswitching? Second, we need to disentangle the apparent tension between phonetic and phonological cues that facilitate a comprehender's expectation for codeswitching with morphosyntactic cues that instead push comprehenders to delay active prediction. Studies that can pit these two cue classes against each other will help elucidate on this question. Finally, taking as a given that bilinguals engage in language practices that vary from monolingual language use, we believe that an important future direction is to investigate whether the codeswitch itself serves as a predictive cue for signaling upcoming content. This line of research would help tie the processing of codeswitching to the functions of codeswitching. Indeed, initial work suggests that this future research area could be fruitful, as recent work demonstrates that bilinguals anticipate harder or less frequent words and experience reduced negative reactivity after processing a codeswitch (Tomić & Valdés Kroff, 2021, 2022). Ultimately, by shining a spotlight on bilingual codeswitching, we hope to highlight how codeswitching is a natural testbed to investigate language processing under heightened uncertainty.

Acknowledgments

The authors would like to thank Edith Kaan, Judy Kroll, Jessica Montag, Jared Novick, M. Carmen Parafita Couto, Aleksandra Tomić and the various audiences where we have presented and discussed the ideas that helped shape the work presented here, including the Bilingualism, Mind, and Brain Speaker Series, the Center for Language Science Distinguished Speaker Series, Bilingualism in the Hispanic and Lusophone World, and the SMART Cognitive Science workshop on the bilingual mind. The writing of this article was supported in part by NSF OISE 1545900 and NSF 2125865 to Paola E. Dussias.

References

Abney, S. P. (1987). *The English noun phrase in its sentential aspect*. (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Aboh, E. (2020). Lessons from Neuro-(a)-typical brains: Universal multilingualism, code-mixing, recombination, and executive functions. *Frontiers in Psychology*, *11*, 488. https://doi.org/10.3389/fpsyg.2020.00488.

Abramson, A. S., & Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics*, *1*, 1–8. https://doi.org/10.1016/S0095-4470(19)31372-5.

Abutalebi, J., Annoni, J., Zimine, I., Pegna, A. J., Seghier, M. L., Lee-Jahnke, H., et al. (2008). Language control and lexical competition in bilinguals: An event-related fMRI Study. *Cerebral Cortex*, *18*, 1496–1505. https://doi.org/10.1093/cercor/bhm182.

Abutalebi, J., & Green, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language & Cognitive Processes*, *23*, 557–582. https://doi.org/10.1080/01690960801920602.

Adler, R. M., Valdés Kroff, J. R., & Novick, J. M. (2020). Does integrating a code-switch during comprehension engage cognitive control? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *46*, 741–759. https://doi.org/10.1037/xlm0000755.

Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory and Cognition*, *24*, 477–492. https://doi.org/10.3758/BF03200936.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1.

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research*, *32*(1), 25–36.

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 914–930.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science*, *15*(9), 578–582.

Balukas, C., & Koops, C. (2015). Spanish-English bilingual voice onset time in spontaneous code-switching. International Journal of Bilingualism, 19, 423-443. https://doi.org/https://doi.org/10.1177/1367006913516035.

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: an ERP comparison. *Journal of Cognitive Neuroscience*, *17*, 137–153. https://doi.org/10.1162/0898929052880101.

Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language*, *95*, 173–189. https://doi.org/10.1016/j.jml.2017.04.002.

Beatty-Martínez, A. L., & Dussias, P. E. (2019). Revisiting masculine and feminine grammatical gender in Spanish: linguistic, psycholinguistic, and neurolinguistic evidence. *Frontiers in Psychology*, *10*, 751. https://doi.org/10.3389/fpsyg.2019.00751.

Beatty-Martínez, A. L., Navarro-Torres, C. A., & Dussias, P. E. (2020). Codeswitching: A bilingual toolkit for opportunistic speech planning. *Frontiers in Psychology*, *11*, 1699. https://doi.org/10.3389/fpsyg.2020.01699.

Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., & Kroll, J. F. (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*, 1022–1047. https://doi.org/10.1037/xlm0000770.

Beatty-Martínez, A. L., Valdés Kroff, J. R., & Dussias, P. E. (2018). From the field to the lab: A converging methods approach to the study of codeswitching. *Language*, *3*, 19. https://doi.org/10.3390/languages3020019.

Belazi, H. M., Rubin, E. J., & Toribio, A. J. (1994). Code switching and X-bar theory: The functional head constraint. *Linguistic Inquiry*, *25*, 221–237.

Bever, T. (1970). The cognitive basis of linguistic structure. In J. R. Hayes (Ed.), *Cognition and the Development of Language* (pp. 279–352). https://doi.org/10.1093/acprof:oso/9780199677139.003.0001.

Blanco-Elorrieta, E., & Pylkkänen, L. (2017). Bilingual language switching in the laboratory versus in the wild: The spatiotemporal dynamics of adaptive language control. *Journal of Neuroscience*, *37*, 9022–9036. https://doi.org/10.1523/JNEUROSCI.0553-17.2017.

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

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language*, *75*, 104–116. https://doi.org/10.1016/j.jml.2014.05.004.

Botvinick, M., Braver, T., Barch, D., Carter, C., & Cohen, J. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652. https://doi.org/10.1037/0033-295x.108.3.624.

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010.

Broersma, M. (2009). Triggered codeswitching between cognate languages. *Bilingualism: Language and Cognition*, *12*, 447–462. https://doi.org/10.1017/S1366728909990204.

Broersma, M. (2011). Triggered code-switching: Evidence from picture naming experiments. In M. S. Schmid, & W. Lowie (Eds.), *Modeling bilingualism from structure to chaos: In honor of Kees de Bot* (pp. 37–57). John Benjamins. https://doi.org/10.1075/sibil.43.

Broersma, M., Carter, D., Donnelly, K., & Konopka, A. (2020). Triggered codeswitching: Lexical processing and conversational dynamics. *Bilingualism: Language and Cognition*, *23*, 295–308. https://doi.org/10.1017/S1366728919000014.

Broersma, M., & De Bot, K. (2006). Triggered codeswitching: A corpus-based evaluation of the original triggering hypothesis and a new alternative. *Bilingualism: Language and Cognition*, *9*, 1–13. https://doi.org/10.1017/S1366728905002348.

Broersma, M., Isurin, L., Bultena, S., & De Bot, K. (2009). Triggered code switching: Evidence from Dutch-English and Russian-English bilinguals. In L. Isurin, D. Winford, & K. de Bot (Eds.), *Multidisciplinary approaches to code switching* (pp. 103–128). John Benjamins. https://doi.org/10.1075/sibil.41.

Bullock, B. E. (2009). Phonetic reflexes of code-switching. In B. E. Bullock, & A. J. Toribio (Eds.), *The Cambridge handbook of linguistic code-switching* (pp. 163–181). Cambridge University Press.

Bullock, B. E., & Toribio, A. J. (2009). Themes in the study of code-switching. In B. E. Bullock, & A. J. Toribio (Eds.), *The Cambridge handbook of linguistic code-switching* (pp. 1–17). Cambridge University Press. https://doi.org/10.1017/CBO9780511576331.002.

Bultena, S., Dijkstra, T., & Van Hell, J. G. (2015). Switch cost modulations in bilingual sentence processing: Evidence from shadowing. *Language, Cognition, and Neuroscience*, *30*, 586–605. https://doi.org/10.1080/23273798.2014.964268.

Byers-Heinlein, K. (2013). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism: Language and Cognition*, *16*(1), 32–48. https://doi.org/10.1017/S1366728912000120.

Byers-Heinlein, K., Morin-Lessarda, E., & Lew-Williams, C. (2017). Bilingual infants control their languages as they listen. *Proceedings of the National Academy of Science*, *114*, 9032–9037.

Callahan, L. (2004). *Spanish/English codeswitching in a written corpus*. John Benjamins. https://doi.org/10.1075/sibil.27.

Clyne, M. (1967). *Transference and Triggering*. Nijhoff.

Clyne, M. (1997). Some of the things trilinguals do. *International Journal of Bilingualism*, *1*, 95–116. https://doi.org/10.1177/136700699700100201.

Cruz, A. (2021). A syntactic approach to gender assignment in Spanish–English bilingual speech. *Glossa: A journal of general linguistics*, *6*(1), 96. https://doi.org/10.16995/glossa.5878.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition*, *30*, 73–105.

Declerck, M. (2020). What about proactive language control? *Psychonomic Bulletin & Review*, *27*, 24–35. https://doi.org/10.3758/s13423-019-01654-1.

Declerck, M., Koch, I., Duñabeitia, J. A., Grainger, J., & Stephan, D. N. (2019). What absent switch costs and mixing costs during bilingual language comprehension can tell us about language control. *Journal of Experimental Psychology: Human Perception and Performance*, *45*, 771–789. https://doi.org/10.1037/xhp0000627.

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

Deuchar, M. (2020). Code-switching in linguistics: A position paper. *Language*, *5*(2), 22. https://doi.org/10.3390/languages5020022.

Deuchar, M., Davies, P., Herring, J. R., Parafita Couto, M. C., & Carter, D. (2014). Building Bilingual Corpora. In M. E. Thomas, & I. Mennen (Eds.), *Advances in the study of bilingualism* (pp. 93–110). Bilingual Matters. https://doi.org/10.21832/9781783091713-008.

Dussias, P. E., & Cramer Scaltz, T. R. (2008). Spanish-English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychologica*, *128*, 501–513. https://doi.org/10.1016/j.actpsy.2007.09.004.

Dussias, P. E., Marful, A., Gerfen, C., & Bajo, M. T. (2010). Usage frequencies of complement-taking verbs in Spanish and English: Data from Spanish monolinguals and Spanish English bilinguals. *Behavior Research Methods*, *42*, 1004–1011. https://doi.org/10.3758/BRM.42.4.1004.

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*, 353–387. https://doi.org/10.1017/S0272263112000915.

Eddington, D. (2002). Spanish gender assignment in an analogical framework. *Journal of Quantitative Linguistics*, *9*, 49–75. https://doi.org/10.1076/jqul.9.1.49.8482.

Fernández, C. B., Litcofsky, K. A., & Van Hell, J. G. (2019). Neural correlates of intra-sentential code-switching in the auditory modality. *Journal of Neurolinguistics*, *51*, 17–41. https://doi.org/10.1016/j.jneuroling.2018.10.004.

Fricke, M., & Kootstra, G. J. (2016). Primed codeswitching in spontaneous bilingual dialogue. *Journal of Memory and Language*, *91*, 181–201. https://doi.org/10.1016/j.jml.2016.04.003.

Fricke, M., Kroll, J. F., & Dussias, P. E. (2016). Phonetic variation in bilingual speech: A lens for studying the production–comprehension link. *Journal of Memory and Language*, *89*, 110–137. https://doi.org/10.1016/j.jml.2015.10.001.

Gahl, S., & Garnsey, S. (2006). Knowledge of grammar includes knowledge of syntactic probabilities. *Language*, *82*, 405–410. https://www.jstor.org/stable/4490163.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*, 58–93. https://doi.org/10.1006/jmla.1997.2512.

Gollan, T. H., Schotter, E. R., Gomez, J., Murillo, M., & Rayner, K. (2014). Multiple levels of bilingual language control: Evidence from language intrusions in reading aloud. *Psychological Science*, *25*, 585–595. https://doi.org/10.1177/0956797613512661.

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

González-Vilbazo, K., Bartlett, L., Downey, S., Ebert, S., Heil, J., Hoot, B., et al. (2013). Methodological considerations in code-switching research. *Studies in Hispanic and Lusophone Linguistics*, *6*(1), 119–138. https://doi.org/10.1515/shll-2013-1143.

González-Vilbazo, K., & Koronkiewicz, B. (2016). Tú y to can codeswitch, nosotros cannot. In R. E. Guzzardfo Tamargo, C. M. Mazak, & M. C. Parafita Couto (Eds.), *Spanish-English codeswitching in the Caribbean and the US* (pp. 237–260). John Benjamins Publishing Company. https://doi.org/10.1075/ihll.11.10gon.

González-Vilbazo, K., & López, L. (2011). Some properties of light verbs in code switching. *Lingua*, *121*, 832–850. https://doi.org/10.1016/j.lingua.2010.11.011.

Green, D. W. (2018). Language control and code-switching. *Language*, *3*, 8. https://doi.org/10.3390/languages3020008s.

Green, D., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*, 515–530. https://doi.org/10.1080/20445911.2013.796377.

Green, D. W., & Wei, L. (2014). A control process model of code-switching. *Language, Cognition and Neuroscience*, *29*, 499–511. https://doi.org/10.1080/23273798.2014.882515.

Gross, M., & Kaushanskaya, M. (2015). Voluntary language switching in English-Spanish bilingual children. *Journal of Cognitive Psychology*, *27*, 992–1013. https://doi.org/10.1080/20445911.2015.1074242.

Gullberg, M., Indefrey, P., & Muysken, P. (2009). Research techniques for the study of code-switching. In B. E. Bullock, & A. J. Toribio (Eds.), *The Cambridge handbook of linguistic code-switching* (pp. 21–39). Cambridge University Press.

Gullifer, J. W., & Titone, D. (2019). The impact of a momentary language switch on bilingual reading: intense at the switch but merciful downstream for L2 but not L1 readers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 236–250. https://doi.org/10.1037/xlm0000695.

Guzzardo Tamargo, R. E., Valdés Kroff, J. R., & Dussias, P. E. (2016). Examining the relationship between comprehension and production processes in code-switched language. *Journal of Memory and Language*, *89*, 138–161. https://doi.org/10.1016/j.jml.2015.12.002.

Hardwick, R. M., Caspers, S., Eickhoff, S. B., & Swinnen, S. P. (2018). Neural correlates of action: Comparing meta-analyses of imagery, observation, and execution. *Neuroscience and Biobehavioral Reviews*, *94*, 31–44.

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages?: Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science*, *15*, 409–414. https://doi.org/10.1111/j.0956-7976.2004.00693.x.

Heredia, R. R., & Altarriba, J. (2001). Bilingual language mixing: Why do bilinguals code-switch? *Current Directions in Psychological Science*, *10*, 164–168. https://doi.org/10.1111/1467-8721.00140.

Hernandez, A. E., Dapretto, M., Mazziotta, J., & Bookheimer, S. (2001). Language switching and language representation in Spanish-English bilinguals: an fMRI study. *NeuroImage*, *2*, 510–520. https://doi.org/10.1006/nimg.2001.0810.

Hofweber, J., Marinis, T., & Treffers-Daller, J. (2016). Effects of dense code-switching on executive control. *Linguistic Approaches to Bilingualism*, *6*, 648–668. https://doi.org/10.1075/lab.15052.hof.

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*, 277–307. https://doi.org/10.1177/0267658315624960.

Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychological Science*, *27*, 572–582. https://doi.org/10.1177/095679761562522.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, *31*(1), 19–31. https://doi.org/10.1080/23273798.2015.1072223.

Humphreys, G. F., & Gennari, S. P. (2014). Competitive mechanisms in sentence processing: common and distinct production and reading comprehension networks linked to the refrontal cortex. *NeuroImage*, *84*, 354–366. https://doi.org/10.1016/j.neuroimage.2013.08.059.

January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience*, *12*, 2434–2444. https://doi.org/10.1162/jocn.2008.21179.

Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *NeuroImage*, *14*, S103–S109.

Johns, M. A., & Dussias, P. E. (2022). Comparing single-word insertions and multi-word alternations in bilingual speech: Insights from pupillometry. *Language*, 7(4), 267. https://doi.org/10.3390/languages7040267.

Johns, M. A., & Steuck, J. (2021). Is codeswitching easy or difficult? Testing processing cost through the prosodic structure of bilingual speech. *Cognition*, *211*, 104634. https://doi.org/10.1016/j.cognition.2021.104634.

Johns, M. A., Valdés Kroff, J. R., & Dussias, P. E. (2019). Mixing things up: How blocking and mixing affect the processing of codemixed sentences. *International Journal of Bilingualism*, *23*, 584–611. https://doi.org/10.1177/1367006917752570.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228–238. https://doi.org/10.1037/0096-3445.111.2.228.
Kaan, E., & Grüter, T. (2022). Prediction in second language processing and learning: Advances and Directions. In E. Kaan, & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 1–24). John Benjamins. https://doi.org/10.1075/bpa.12.01kaa.
Kaan, E., Kheder, S., Kreidler, A., Tomić, A., & Valdés Kroff, J. R. (2020). Processing code-switches in the presence of others: An ERP Study. *Frontiers in Psychology*, *11*, 1288. https://doi.org/10.3389/fpsyg.2020.01288.
Kan, I. P., Teubner-Rhodes, S., Drummey, A. B., Nutile, L., Krupa, L., & Novick, J. M. (2013). To adapt or not to adapt: The question of domain-general cognitive control. *Cognition*, *129*(3), 637–651. https://doi.org/10.1016/j.cognition.2013.09.001.
Kheder, S., & Kaan, E. (2019). Lexical selection, cross-language interaction, and switch costs in habitually codeswitching bilinguals. *Bilingualism: Language and Cognition*, *22*, 569–589. https://doi.org/10.1017/S1366728918000500.
Kootstra, G. J., Dijkstra, T., & Van Hell, J. G. (2020). Interactive alignment and lexical triggering of code-switching in bilingual dialogue. *Frontiers in Psychology*, *11*, 1747. https://doi.org/10.3389/fpsyg.2020.01747.
Kootstra, G. J., van Hell, J. G., & Dijkstra, T. (2010). Syntactic alignment and shared word order in code-switched sentence production: Evidence from bilingual monologue and dialogue. *Journal of Memory and Language*, *63*, 210–231. https://doi.org/10.1016/j.jml.2010.03.006.
Koronkiewicz, B. (2020). Subject-predicate code-switching: Testing the need of a matrix language through embedding. In A. Morales-Front, M. J. Ferreira, R. P. Leow, & C. Sanz (Eds.), *Hispanic linguistics: Current issues and new directions* (pp. 249–264). John Benjamins. https://doi.org/10.1075/ihll.26.12kor.
Kroll, J., Bobb, S., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, *9*, 119–135. https://doi.org/10.1017/S1366728906002483.
Kroll, J. F., Dussias, P. E., Bice, K., & Perrotti, L. (2015). Bilingualism, mind, and brain. *Annual review of linguistics*, *1*, 377–394. https://doi.org/10.1146/annurev-linguist-030514-124937.
Kroll, J. F., & Gollan, T. H. (2014). Speech planning in two languages: What bilinguals tell us about language production. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 165–181). Oxford University Press.
Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.
Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*, 193–198. https://doi.org/10.1111/j.1467-9280.2007.01871.x.
Li, P. (1996). Spoken work recognition of code-switched words by Chinese-English bilinguals. *Journal of Memory and Language*, *35*, 757–774. https://doi.org/10.1006/jmla.1996.0039.
Li, C., & Gollan, T. (2018). Cognates facilitate switches and then confusion: Contrasting effects of cascade versus feedback on language selection. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *44*, 974–991. https://doi.org/10.1037/xlm0000497.

Liceras, J. M., Fernández Fuertes, R., Perales, S., Pérez-Tattam, R., & Spradlin, K. T. (2008). Gender and gender agreement in bilingual native and non-native grammars: A view from child and adult functional-lexical mixings. *Lingua*, *118*, 827–851. https://doi.org/10.1016/j.lingua.2007.05.006.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: evidence for the role of inhibition in second-language learning. *Psychological Science*, *12*, 1507–1515. https://doi.org/10.1111/j.1467-9280.2009.02480.x.

Lipski, J. M. (1978). Code-switching and the problem of bilingual competence. In M. Paradis (Ed.), *Aspects of Bilingualism* (pp. 250–264). Hornbeam Press.

Lipski, J. M. (1982). Spanish-English language switching in speech and literature: Theories and models. *Bilingual Review*, *9*, 191–212.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384–422. https://doi.org/10.1080/00437956.1964.11659830.

Litcofsky, K. A., & Van Hell, J. G. (2017). Switching direction affects switching costs. Behavioral, ERP and time-frequency analyses of intra-sentential codeswitching. *Neuropsychologia*, *97*, 112–139. https://doi.org/10.1016/j.neuropsychologia.2017.02.002.

López, L. (2020). *Bilingual Grammar*. Cambridge University Press. https://doi.org/10.1017/9781108756181.

Luk, G., Green, D. W., Abutalebi, J., & Grady, C. (2011). Cognitive control for language switching in bilinguals: A quantitative meta-analysis of functional neuroimaging studies. *Language & Cognitive Processes*, *27*, 1479–1488. https://doi.org/10.1080/01690965.2011.613209.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226. https://doi.org/10.3389/fpsyg.2013.00226.

MacSwan, J. (1999). *A minimalist approach to intrasentential code switching*. Garland Press.

Mahootian, S., & Santorini, B. (1996). Code Switching and the Complement/Adjunct Distinction. *Linguistic Inquiry*, *24*, 464–479. http://www.jstor.org/stable/4178946.

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, *40*, 25–40. https://doi.org/10.1006/jmla.1998.2602.

Monsell, S. (2003). Task switching. *Trends in Cognitive. Science*, *7*, 134–140. https://doi.org/10.1016/s1364-6613(03)00028-7.

Montes-Alcalá, C. (2000). Written code-switching: Powerful bilingual images. In R. Jacobson (Ed.), *Codeswitching worldwide* (pp. 193–222). Mouton de Gruyter. https://doi.org/10.1515/9783110808742.

Montes-Alcalá, C. (2001). Oral vs. written code-switching contexts in English-Spanish bilingual narratives. In I. Cruz Cabanillas (Ed.), *La lingüística aplicada a finales de siglo XX* (pp. 715–720). Servicios de Publicaciones de la Universidad de Alcalá de Henares.

Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, *58*(3), 503–553. https://doi.org/10.1111/j.1467-9922.2008.00449.x.

Morales, L., Paolieri, D., Dussias, P. E., Valdés Kroff, J. R., Gerfen, C., & Bajo, M. T. (2016). The gender congruency effect during bilingual spoken-word recognition. *Bilingualism: Language and Cognition*, *19*, 294–310. https://doi.org/10.1017/S1366728915000176.

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching palabras (words): An electrophysiological study of code switching. *Brain and Language*, *80*, 188–207.

Munarriz-Ibarrola, A., Ezeizabarrena, M.-J., Arrazola, V. D. C., & Parafita Couto, M. C. (2022). Gender assignment strategies and L1 effects in the elicited production of mixed Spanish- Basque DPs. *Linguistic Approaches to Bilingualism*, *12*(6), 778–815. https://doi.org/10.1075/lab.20016.mun.

Muysken, P. (1997). Code-switching processes: Alternation, insertion, congruent lexicalization. In M. Pütz (Ed.), *Language choices: Conditions, constraints, and consequences* (pp. 361–380). John Benjamins.

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge University Press.

Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in codeswitching*. Clarendon Press for Oxford University Press.

Myers-Scotton, C., & Jake, J. L. (2000). Testing the 4-M model: An introduction. *International Journal of Bilingualism*, *4*, 1–8. https://doi.org/10.1177/13670069000040010101.

Myers-Scotton, C., & Jake, J. L. (2009). A universal model of code-switching and bilingual language processing and production. In B. E. Bullock, & A. J. Toribio (Eds.), *The Cambridge handbook of linguistic code-switching* (pp. 336–357). Cambridge University Press.

Neveu, A., McDonald, M., & Kaushanskaya, M. (2022). Testing the Triggering Hypothesis: Effect of Cognate Status on Code-Switching and Disfluencies. *Language*, 7, 264. https://doi.org/10.3390/languages7040264.

Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology*, *26*, 527–567. https://doi.org/10.1080/026432909035193673.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *5*, 263–281. https://doi.org/10.3758/CABN.5.3.263.

Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, *41*, 407–420. https://doi.org/10.1016/j.wocn.2013.07.005.

Otheguy, R., & Lapidus, N. (2003). An adaptive approach to noun gender in New York Contact Spanish. In R. Cameron, I. López, & R. Núñez Cedeño (Eds.), *A Romance perspective on language knowledge and use* (pp. 209–229). John Benjamins. https://doi.org/10.1075//cilt.238.17oth.

Paolieri, D., Demestre, J., Guasch, M., Bajo, T., & Ferré, P. (2020). The gender congruency effect in Catalan–Spanish bilinguals: Behavioral and electrophysiological evidence. *Bilingualism: Language and Cognition*, *23*, 1045–1055. https://doi.org/10.1017/S1366728920000073.

Peñalosa, F. (1980). *Chicano Sociolinguistics*. Newbury House.

Pfaff, C. W. (1976). Functional and syntactic constraints on syntactic variation in code-mixing. In B. Steever, et al. (Eds.), *Papers from the Parasession on Diachronic Syntax* (pp. 248–259). Chico Linguistic Society.

Pfaff, C. W. (1979). Constraints on language mixing: Intrasentential codeswitching and borrowing in Spanish/English. *Language*, *55*, 291–318.

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

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

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics*, *18*, 581–618.

Poplack, S. (1987). Contrasting patterns of code-switching in two communities. In E. Wande, A. Jan, N. Bengt, L. Steensland, & M. Thelander (Eds.), *Aspects of Multilingualism: Proceedings from the Fourth Nordic Symposium of Bilingualism* (pp. 51–77). Borgströms.

Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, *26*, 47–104. https://doi.org/10.1515/ling.1988.26.1.47.

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: evidence from Spanish-English and Mandarin-English bilinguals. *Journal of the International Neuropsychological Society*, *17*, 682–691. https://doi.org/10.1017/S1355617711000580.

Pyers, J. E., Gollan, T. H., & Emmorey, K. (2009). Bimodal bilinguals reveal the source of tip-of- the-tongue states. *Cognition*, *112*, 323–329. https://doi.org/10.1016/j.cognition.2009.04.007.

Quick, A. E., Backus, A., & Lieven, E. (2021). Entrenchment effects in code-mixing: Individual differences in German-English bilingual children. *Cognitive Linguistics*, *32*(2), 319–348. https://doi.org/10.1515/cog-2020-0036.

Ribot, K. M., & Hoff, E. (2014). "¿Cómo estas?" "I'm good." Conversational code-switching is related to profiles of expressive and receptive proficiency in Spanish-English bilingual toddlers. *International Journal of Behavioral Development*, *38*, 333–341. https://doi.org/10.1177/0165025414533225.

Román, P., Kaan, E., & Dussias, P. E. (2022). Access to verb bias and plausibility information during syntactic processing in adult Spanish-English bilinguals. *Bilingualism: Language and Cognition*, *25*, 417–429. https://doi.org/10.1017/s1366728921000924.

Salig, L. K., Valdés Kroff, J. R., Slevc, L. R., & Novick, J. M. (under review). Linking frequency to bilingual switch costs during real-time sentence comprehension. Manuscript submitted for publication.

Sankoff, D., & Poplack, S. (1981). A formal grammar of code-switching. *Papers in Linguistics: International Journal of Human communication*, *14*(1), 3–46.

Sarkis, J. T., & Montag, J. L. (2021). The effect of lexical accessibility on Spanish-English intra- sentential codeswitching. *Memory & Cognition*, *49*, 163–180. https://doi.org/10.3758/s13421-020-01069-7.

Shen, A., Gahl, S., & Johnson, K. (2020). Didn't hear that coming: Effects of withholding phonetic cues to code-switching. *Bilingualism: Language and Cognition*, *23*, 1020–1031. https://doi.org/10.1017/S1366728919000877.

Smolak, E., de Anda, S., Enriquez, B., Poulin-Dubois, D., & Friend, M. (2020). Code-switching in young bilingual toddlers: A longitudinal, cross-language investigation. *Bilingualism: Language and Cognition*, *23*, 500–518. https://doi.org/10.1017/S1366728919000257.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration Of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634. https://doi.org/10.1126/science.7777863.

Tomić, A., & Kaan, E. (2022). Oscillatory brain responses to processing code-switches in the presence of others. *Brain and Language*, *231*, 105139. https://doi.org/10.1016/j.bandl.2022.105139.

Tomić, A., & Valdés Kroff, J. R. (2021). Code-switching: A processing burden or a valuable resource for prediction? In E. Kaan, & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 139–166).

Tomić, A., & Valdés Kroff, J. R. (2022). Code-switching aids the prediction of the unexpected. *Bilingualism: Language and Cognition*, *25*, 81–92. https://doi.org/10.1017/S1366728921000237.

Torres Cacoullos, R., Dion, N., LaCasse, D., & Poplack, S. (2022). How to mix: Confronting "mixed" NP models and bilinguals' choices. *Linguistic Approaches to Bilingualism*, *12*, 628–656. https://doi.org/10.1075/lab.20017.tor.

Torres Cacoullos, R., & Travis, C. E. (2018). Bilingualism in the community: Codeswitching and grammars in contact. *Cambridge University Press*. https://doi.org/10.1017/9781108235259.

Trawick, S. (2021). *Revisiting the Concept of Triggering of Code-Switching*. [Unpublished doctoral dissertation]. Penn State University. https://etda.libraries.psu.edu/catalog/23596smt51.

Treffers-Daller, J., Ongun, Z., Hofweber, J., & Korenar, M. (2020). Explaining individual differences in executive functions performance in multilinguals: The impact of code-switching and alternating between multicultural identity styles. *Frontiers in Psychology*, *11*, 561088. https://doi.org/10.3389/fpsyg.2020.561088.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition*, *73*, 89–134. https://doi.org/10.1016/s0010-0277(99)00032-3.

Valdés Kroff, J. R. (2012). *Using eye-tracking to study auditory comprehension in codeswitching: Evidence for the link between comprehension and production*. [Unpublished doctoral dissertation], Penn State University.

Valdés Kroff, J. R. (2016). Mixed NPs in Spanish-English bilingual speech: Using a corpus-based approach to inform models of sentence processing. In R. E. Guzzardo Tamargo, C. M. Mazak, & M. C. Parafita Couto (Eds.), *Spanish-English codeswitching in the Caribbean and the U.S* (pp. 281–300). John Benjamins Publishing Company. https://doi.org/10.1075/ihll.11.12val.

Valdés Kroff, J. R., Dussias, P. E., Gerfen, C., Perrotti, L., & Bajo, M. T. (2017). Experience with code-switching modulates the use of grammatical gender during sentence processing. *Linguist Approaches Bilingualism*, 7, 163–198. https://doi.org/10.1075/lab.15010.val.

Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Dussias, P. E. (2018). Experimental contributions of eye-tracking to the understanding of processes while hearing and reading code-switches. *Linguist Approaches Bilingualism*, *8*, 98–133. https://doi.org/10.1075/lab.16011.val.

Valdés Kroff, J. R., Román, P., & Dussias, P. E. (2020). Are all code-switches processed alike? Examining semantic v. language unexpectancy. *Frontiers in Psychology*, *11*, 2138. https://doi.org/10.3389/fpsyg.2020.02138.

Van Hell, J. G. (2023). Code-switching. In A. Godfroid, & H. Hopp (Eds.), *Routledge handbook of second language acquisition and psycholinguistics* Routledge.

Yow, W. Q., Patrycia, F., & Flynn, S. (2016). Code-switching in childhood. In E. Nicoladis, & S. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 81–100). American Psychological Association. https://doi.org/10.1037/14939-006.

Yow, W., Tan, J., & Flynn, S. (2018). Code-switching as a marker of linguistic competence in bilingual children. *Bilingualism: Language and Cognition*, *21*(5), 1075–1090. https://doi.org/10.1017/S1366728917000335.

Zenella, A. C. (1996). *Growing up bilingual: Puerto Rican children in New York*. Wiley-Blackwell.

Zirnstein, M., van Hell, J. G., & Kroll, J. F. (2018). Cognitive control ability mediates prediction costs in monolinguals and bilinguals. *Cognition*, *176*, 87–106. https://doi.org/10.1016/j.cognition.2018.03.001.

This page intentionally left blank

Serial Editor
KARA D. FEDERMEIER
University of Illinois, USA

ISBN 978-0-443-13409-8

9 780443 134098

ELSEVIER

ACADEMIC PRESS
An imprint of Elsevier
elsevier.com/books-and-journals