LANGUAGE PRODUCTION UNDER MESSAGE UNCERTAINTY

by

Arella E. Gussow

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Psychology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: 06/28/2023

The dissertation is approved by the following members of the Final Oral Committee:
Maryellen C. MacDonald, Professor, Psychology
Martha W. Alibali, Professor, Psychology
C. Shawn Green, Professor, Psychology
Fernando Tejedo-Herrero, Professor, Spanish & Portuguese

Table of Contents

Acknowledgments

This is, by far, the part of my dissertation that I most looked forward to writing. Yet it now seems impossible, within such limited space, to name and thank all of the people who have contributed to this work and to my PhD journey. The following is only a humble attempt.

I would first like to thank my advisor, Maryellen MacDonald, for trusting me enough to develop these ideas into a research program, for exemplifying what it is to be a good scientist, and for sharing your brilliance and wisdom with me over the past five years (and more to come, I have no doubt!). Thanks also to my committee members, Shawn Green, Martha Alibali, and Fernando Tejedo-Herrero, for the continuous encouragement, support, and motivation to continue developing and improving my work. I am also grateful for the financial support awarded to me 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.

Thanks to my research assistants and mentees at the Language and Cognitive Neuroscience lab over the years: Yuzhe Gu, Lucy Tasker, Augie Bennett, Skye Stanwich, Elena Groves, Evelyn Yee, Sarah McLeod, Carissa Witthuhn, Sarah Wilhite, Carmen Backes, Gabbi Gragg, Jenna McBride, Danielle Miller, Elena Pittman, Tia Sadlon, Alejandra Gennatiempo, Zeqing Li, Marcela Chavez, Neha Prasad, Elijah Fisher, Artem Lyu, Jiaxuan Li, Kim Otto, Taylor Bodart, and Hanna Sirek. I quite literally could not have done this without you. Our group meetings were always the highlight of my week, and I am eternally grateful for the opportunity to meet and learn from each and every one of you.

Thank you to the many graduate students and postdocs at UW-Madison who have helped me along the way, including Steve Schwering, Elise Hopman, Cass Jacobs, Matt Cooper-Borkenhagen, Matt Borman, Mark Koranda, Desia Bacon, Ella Qiawen Liu, Yuyan (Lillian) Xu, Martin Zettersten, Aaron Cochrane, Melina Knabe, and José Luis Garrido Rivera. To the Psychology department staff, and especially Kevin Belt – I still don't understand the system and likely never will, but trusting you meant I could dedicate more time to my work and life and less to trying to understand bureaucracy. To our LCNL lab manager, Misty Kabasa, for making my onboarding so much smoother, for always having my back, for offering me your pink rollerblades during quarantine, and for random life conversations on the office couch.

To my M.Sc mentors at the BCBL in San Sebastián, Effie Kapnoula and Nicola Molinaro, for believing in me, encouraging me to apply to grad school, and guiding me through my first publication when I didn't think that it (or I) was worthy enough. To Sara Guediche, for encouragement and guidance throughout the daunting application process for PhDs in the USA. To my BCBL cohort, *the fam*, and especially Jordi Martorell, my fellow sentence processer. I always saw you not only as a peer but also as a mentor, and I am most grateful to call you my friend.

I am also extremely lucky to be surrounded by a support system of friends and family, who, despite not fully understanding what it is that I do, have accompanied and led me to this point (and even celebrate my accomplishments more than I do). To my friends in Israel; Gaby Wexler-Tanami, Tzipi Huberman, Shani Seri-Damoza, Sofya Plut, Ya'ara Loyfer – thank you for keeping me grounded, and for staying in touch even when I'm so terrible at it. To Ye'ela Asraf-Toledano – you always said I should listen to my heart more and to my brain less, so I majored in Psychology to understand the difference. Ten years and four degrees later, I still think they're the same damn thing. You still think otherwise. To Tal Amichay, for being my person throughout undergrad at Hebrew U, for commiserating when I considered dropping out after a month (and then planning to join me), for our award-winning research proposal two years later, and for that one conversation at the Zigmond bar on Aza street, when we were too scared to imagine the lives we live now. I am so goddamn proud of us.

To Allie Lessard – any attempt to put my gratitude into words will only diminish what it is that I'm thankful for. But I'll say this: Thank you for growing and learning with and alongside me. It's always the next person that gets the better version.

To my brother Ayal, when you asked me why I wanted to do a PhD I answered "because it's kinda my dream", and you said "good, because that's the only good reason to do it". Solid advice as always, and luckily no regrets. To my sister Ella, for entertaining me as I try to express some knowledge of UX/UI, and for cracking me up when describing how Wisconsinites walk in line through the farmers' market ("I'm Israeli, I don't have patience for this! Let's go!"). To my brother Eliran, for being my first ever student, and for calling to ask what it is again that I study when you were a tour guide for a birthright group from UW-Madison (and then discovered that one of my students was on it too!). To my sister Or-Tal, for sending me snippets of new songs to ask my opinion, for keeping me up-to-date with Israeli pop culture, and for shaming me on the family texting groups when I don't answer your private messages (texting is hard!!). To my sister-in-law Melodi, for always welcoming me with open arms, red wine, and hot sauce. To my brother-in-law Avihai, for deep philosophical conversations alongside hilarious military stories, and for trusting me enough to officiate your wedding.

To my Grandparents, Saba and Savta, who keep asking when I'm coming back and why I need to go to school for so long when I could just get a job as a real-estate agent and make a bunch of money fast like the neighbor's kid did (fair point, though I would be a terrible agent!). I wish you had had the opportunities I have. I wish you had had the choices I have. To Grandpa Lenny, who passed away during my PhD, but would entertain me with long phone calls and the cutest greeting cards during my first three years here. To my New York family; Ely, Leora, Maia, & Gabi Pinto, Yossi Pinto, Raphi Uziel, Amy & Bruce Engelberg. To my Miami family; Alegre Uziel, Genie, Max, Alegria, Moshe, & Vidal Bensimon. To my Modi'in family; Joe, Sisi, Tal, Roni, Itai, & Danielle Uziel. Thank you all for keeping me grounded, checking in on me, and sharing your wisdom and love. To my Bubbie – I wish you could see me right now. This accomplishment is yours as much as it is mine. Thank you for teaching me to read and to write, to love language, to love learning, to love teaching.

To my Abba, I'll never forget the Shabbat meal seventeen years ago when I casually mentioned I wanted to learn how to play the saxophone. On Tuesday you called to say you'd found me a teacher and a saxophone rental, which could be bought after three months if I liked it. This is just one example of your never-failing support for me and every goal I set for myself.

To my Ima, I know you think I lead multiple secret lives, and that may or may not be true, but your love and support are felt in each and every one of them. You always say I'm your rock but I don't think you realize that you too are my (and so many others') rock. Thank you for being a role model of determination, strength, resilience, empathy, openness to learn, and unconditional love. This is for you.

Abstract

Language production is assumed to begin with a *message*: a thought or idea that the speaker must turn into a verbal utterance that others can understand. But speakers occasionally begin speaking before they are certain of what to say – that is, before they know what message they want to communicate. Theories of language production do not account for these situations, however, as researchers typically assume that production begins with a pre-determined message.

In this dissertation, I investigate situations of message uncertainty and how they affect utterance formulation. Given the relative lack of research on production under message uncertainty, I begin by reviewing evidence from several neighboring areas to introduce issues of message uncertainty and motivate empirical research. Next, I present results from six experiments designed to investigate the strategies that speakers use for production under message uncertainty, and the resulting speech patterns: In Chapter Two I use a novel picture-naming paradigm to show that message uncertainty affects speakers' word order choices, as they prioritize the more certain components of their utterances and produce them earlier – allowing for production to begin sooner. In Chapter Three I investigate the time course of speakers' utterances in a motion event-description task, manipulating whether the goal of the motion event is apparent from the start or only disambiguated at the end of the movement. Moreover, I use a cross-linguistic comparison between Spanish and English to show how the permitted word order and grammatical properties of each language affect the time course of event description. Finally, in Chapter Four I discuss the implications of a language production system that is highly flexible in adapting its production strategies – depending on goals, contexts, and individual speakers – and how message uncertainty needs to be incorporated into current models.

Chapter One: A review of message uncertainty

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 do not necessarily have to be stressful or high-stakes: imagine your partner asks what you would like to do this weekend, but you're not sure what you're in the mood for. You are unlikely to remain silent, as they might think you didn't hear them, or infer you are intentionally ignoring the question. 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 have 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 dissertation, I investigate situations of message uncertainty and how they affect utterance formulation, whether in the choice of utterance forms or in the time course of production. Given the relative lack of research on production under uncertainty, in Chapter One[1] I review evidence from several neighboring areas to introduce issues of uncertainty and motivate empirical research. Subsequent chapters present six experiments designed to investigate the strategies that speakers use for production under message uncertainty, and the resulting speech patterns: In Chapter Two[2] I use a novel picture-naming paradigm to show that message uncertainty affects speakers' word order choices, as they prioritize the more certain components of their utterances and produce them earlier – allowing for production to begin sooner. In Chapter Three I investigate the time course of speakers' utterances in a motion event description task, manipulating whether a central component of the event is obvious from the start or only disambiguated at the end the movement. Moreover, I use a cross-linguistic comparison of

---

[1] Chapter One is a revised version of my depth preliminary exam. It was previously published and is presented here with minor modifications.
Gussow, A. E. (2023). Language production under message uncertainty: When, how, and why we speak before we think. In K. D. Federmeier & J. L. Montag (Eds.), *Psychology of learning and motivation: Speaking, writing, and communicating* (Vol. 78). Academic Press.

[2] Chapter Two was co-written with Maryellen MacDonald and previously published; a modified version is used here.
Gussow, A. E., & MacDonald, M. C. (2023). Utterance planning under message uncertainty: Evidence from a novel picture-naming paradigm. *Cognitive, Affective, & Behavioral Neuroscience*, 1-16.

Spanish and English to show how the permitted word order of each language affects the time course of event description under message uncertainty. Finally, in Chapter Four I summarize the main takeaways and discuss future directions.

The remainder of Chapter One is organized as follows. I will begin my review 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 et al., 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 et al., 2007), shared knowledge between speakers (Heller et al., 2009), frequency and priming of words (Bock, 1986a; 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 et al., 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 et al., 2013; Gleitman et al., 2007; Jaeger et al., 2012; van de Velde et al., 2014), simple questions to be answered (Chia & Kaschak, 2022), or picture naming (Meyer, 1996; Meyer & Schriefers, 1991; Strijkers et al., 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 & 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; amongst 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 contrastive 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 described earlier – 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 et al., 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 et al., 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 et al., 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 in utterance form decisions.

3. Message uncertainty in context

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 et al., 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 et al., 2016; Bögels, 2020; Corps et al., 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 (Allopenna et al., 1998; Altmann & Kamide, 1999; Kamide, Altmann, et al., 2003; Kamide, Scheepers, et al., 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. Findings from these investigations 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). 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 et al., 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, Schacter et al. (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 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 & 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 (see Chapter Two), 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 of the review, 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 et al., 2020; Lebkuecher et al., 2022; MacDonald, 2016; Rosenbaum et al., 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 et al. (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 et al., 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 et al., 2007; Schwanenflugel & Shoben, 1985; Staub et al., 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 et al., 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 et al., 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 et al., 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 et al., 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 et al., 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 et al., 2021; Gallivan et al., 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 et al., 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 (see Chapter Two). 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) versus 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 et al., 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 et al., 2010, 2015), their proximity to each other (Grent et al., 2015), and even on hand choice for executing the action (van Helvert et al., 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 et al., 2012; de Zubicaray et al., 2006; Piai et al., 2014, 2018, 2020). 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. In Chapters 2 and 3 I present novel paradigms that would be useful tools to begin these investigations, as they can likely be adapted to studies with physiological measures.

5. Review summary

In this chapter I introduced *production under message uncertainty,* suggesting that sometimes speakers begin 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.

In the following Chapters, I begin applying these recommendations to experimental research on language production. In Chapters Two and Three I present six experiments designed to investigate the strategies that speakers use for production under message uncertainty, and how they affect the resulting speech patterns. The goal is to begin explicitly introducing message uncertainty into language production research, and present novel paradigms that bring message uncertainty into the lab. In Chapter Four I summarize the main findings from these experiments and lay out future directions.

Chapter Two: A novel picture-naming paradigm to investigate word order choices under message uncertainty

Production experiments commonly provide participants with the message they need to convey, such as a scene to describe or a picture to name (Bock, 1996). While this method is helpful for experimental control, it is not well-suited for studying planning under message uncertainty. In the current study, we modify standard picture-description tasks to first display several pictures, and only later provide information about which pictures are to be named. Our paradigm is conceptually related to those in 'go-before-you-know' motor reaching studies described earlier, in which it is temporarily unknown which of several locations will be the target of the reaching action (Chapman et al., 2010; Gallivan et al., 2015). The language implementation is very different from reaching tasks, however, and predictions also diverge.

In Experiments 1 and 2, participants view displays of four images, and are later cued to name two of the images. Different types of display allow manipulation of message uncertainty: in one condition (Overlap), participants have early information about one of the two pictures to be named; that is, they already know half of the message. In the other condition (Different,) participants see several options but are uncertain about all of them. Thus our manipulation affects *how much* of the message information participants have at the start of the trial, and consequently, how much of the utterance they can plan. In Experiment 3, we use a semantically constraining question to provide early information about the upcoming targets, manipulating how *likely* a certain target is to be in the message. By carefully controlling the timing of when the full information is revealed, we can examine how speakers incorporate any partial message information into their utterance planning.

Whereas participants in motor reaching studies produce movement trajectories that are in between potential targets, such intermediate strategies don't work for word production – for someone who is uncertain whether to say "hike" or "picnic," saying a blend of the two words is not an effective strategy. Instead, we hypothesize that producers will rely on the known flexibility of sequencing in utterance planning. Specifically, we predict that speakers will begin planning components of the utterance that are certain to be useful, while the rest can be planned later, as more information becomes available. If so, this production strategy will affect speakers' word order, as they prioritize the components that are certain to be useful and choose to place them first in the utterance. Early planning should also allow speakers to begin production sooner, showing a benefit in speech initiation latencies. Such findings would suggest that word order can be shaped by implicit planning strategies that maximize production efficiency, including early planning – even under message uncertainty.

Experiment 1

Method

Participants

Sixty-six undergraduate students, all native English speakers, participated for course credit. Sample size was determined based on pilot testing and a power analysis conducted using PANGEA (Westfall, 2016), aiming for at least 80% power with a moderate effect size, and allowing leeway in case of participant exclusions. Data from seven additional participants were excluded from analyses: four who did not follow instructions, two due to equipment failure, and one who was a nonnative English speaker.

Stimuli

Object images from the MultiPic database (n=320, Duñabeitia et al., 2017) were used as stimuli, assigned to 80 displays of four objects each. Object names within a display were matched on syllable count and word frequency (Balota et al., 2007). For each display, two objects were randomly chosen as the target pair, and the remaining two were the competitor pair. Objects were never repeated across displays.

Each display could appear in one of two conditions: 1) Overlap; where one of the target pictures was repeated in the display, replacing one of the competitors, 2) Different; where there was no overlap between targets and competitors. Each display contained one pair of images appearing on either side of a computer screen, as in Figure 2.1. The two images within each pair rotated around each other to avoid position effects on naming order.

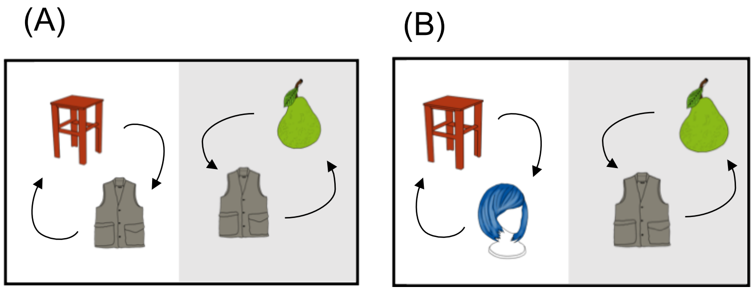

Figure 2.1. Examples of visual displays in the (A) Overlap condition, (B) Different condition. Every two images rotated around each other, as illustrated by the arrows. Arrows were not visible to participants. The gray background appeared after 2.2 seconds of exposure, indicating the targets.

Procedure

Participants came into the lab and were seated in a sound-proof booth with a computer and a freestanding microphone. They first completed a familiarization phase, where they viewed each of the images with its name on screen and said the name aloud. Each image appeared for three seconds with a one second inter-stimulus fixation; the familiarization phase lasted approximately fifteen minutes.

Participants then received the main task instructions, including example displays and four practice trials with feedback. In each trial, a display was presented, and after 2.2 seconds, a gray background appeared on one side of the screen, indicating the target pair. Participants' task was to answer the question "Which are the target images?", and they were told that another participant would later listen to their recordings to identify the targets. This cover story was added to encourage participants to treat the task as information sharing with another individual. Participants were instructed to produce a full phrase, such as "The vest and the pear" for the example in Figure 2.1, rather than simply naming the objects. Participants were told that sometimes pictures would be duplicated in displays, but that they should name the pair cued by the gray background. Participants were not told anything about uncertainty or early language planning and remained blind to the goal of the experiment.

The test phase of 80 trials lasted approximately twenty minutes. Participants were randomly assigned to one of eight lists, counterbalancing which picture was doubled, which competitor was replaced, and condition. Images were offset from the center at the start of the trial. The target side and the initial position of each image within the pair were pseudorandomized, i.e., appearing on each side of the screen (left vs right) and in each position within the pair (top vs bottom) in equal proportions across trials. Order of items (displays) was randomized per participant. After completing the experiment, participants answered a survey to provide feedback, including their guess about the experiment purpose and/or any strategies they developed.

Analyses and Results

Preprocessing

Speech files were manually transcribed. An automated script then coded each trial for accuracy in object naming. Object names that differed from the names learned in the familiarization phase were coded as inaccurate, even if the produced name was generally acceptable in English. Trials in which one or both objects were named inaccurately (25%) were removed from analyses, as were trials with disfluencies (3%).

The FAVE Program Suite (Rosenfelder et al., 2014) was used to extract onset and offset times of each word in each utterance. Before performing latency analyses, we further removed responses that did not follow the 5-word conjoined noun phrase structure (21%), trials where participants began speaking before the cue (6%), and responses with a total duration of more than 2.5 SD above the mean duration (3%). Note that percentages refer to the percentage of remaining observations after the prior step of data cleaning. This left 2797 observations (1419 Different, 1378 Overlap; 53% of all observations) from 55 participants for each word position in the latency analyses.

Word Order

We first tested whether participants were more likely to name the overlapping object first in their responses in the Overlap condition. We ran a mixed-effects logistic regression model regressing Order (non-overlapping vs. overlapping object produced first) on the intercept, including by-participant and by-item random intercepts (the maximal random effects structure; Barr, 2013). The coefficient is reported in log-odds and represents the model intercept. The null model for comparison sets the intercept at zero, which in log-odds is equivalent to 50%. A positive significant intercept therefore indicates above-chance preference for one word order over the other. As expected, participants were significantly above chance in producing the overlapping object first in their utterance (b = .75, *SE* = .1, *z* = 7.28, *p* < .0001), see Figure 2.2.

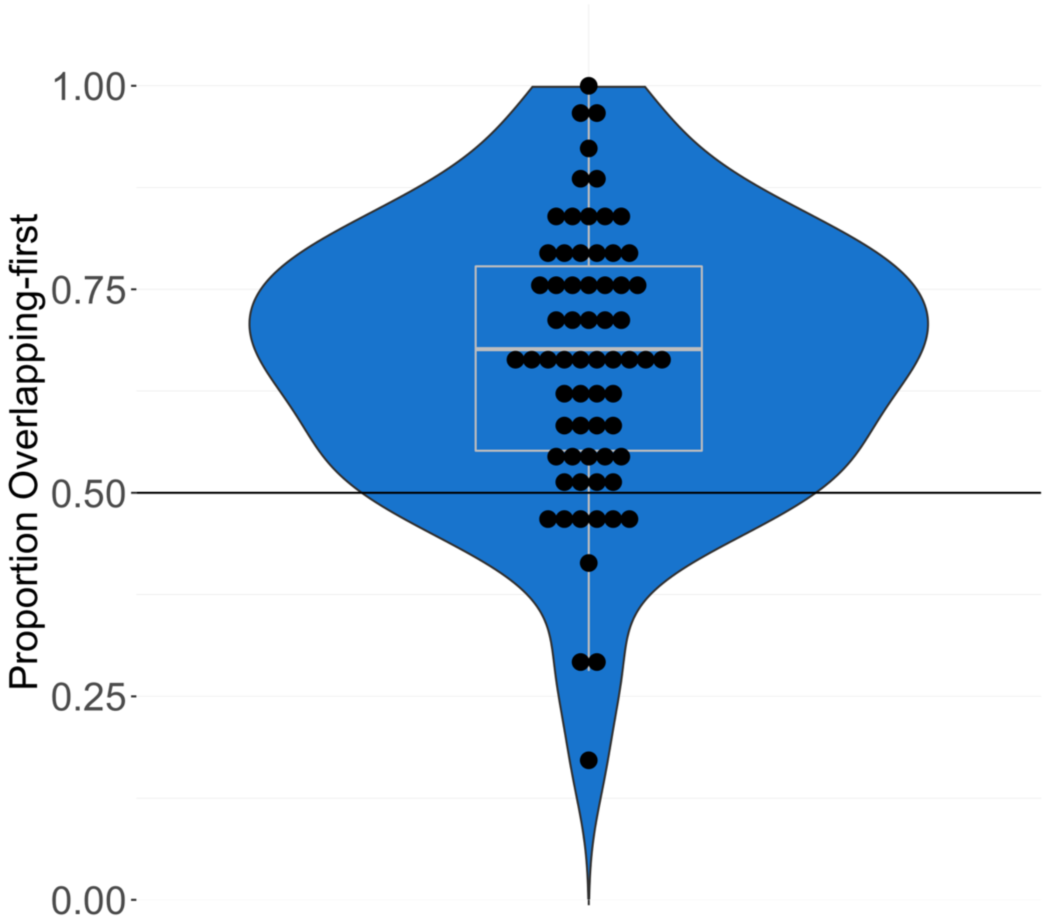

Figure 2.2. Violin plot of the proportion of trials in the Overlap condition in which participants named the overlapping target first (Experiment 1). A boxplot is overlayed in gray to identify quartiles and the median proportion. The horizontal black line identifies chance level. Points reflect individual participant means.

Latencies

We next examined speech initiation latencies (onsets) for each word in participants' utterances. Onsets were calculated as the time between the cue appearing and the first phoneme of each word and were log-transformed for analyses. The predictor variables were word Position (1-5), Condition (Different, Overlap; coded [0,1]), and Order of naming (non-overlapping first, overlapping first; coded [0,1]). Order was nested within condition (Condition/Order) because it was only relevant in the Overlap condition. Position was coded using successive difference contrasts, where each word position is compared to the preceding word position (2-1, 3-2, 4-3, 5-4). Each contrast therefore measures the difference between the onset of a given word and the onset of the prior word in that utterance.

We ran a mixed-effects linear regression model regressing Onset on Position, Condition/Order, and the interaction between Position and Condition/Order. Including the nested term (Condition/Order) is equivalent to including fixed effects of Condition and the interaction between Condition and Order. The maximal random effects structure to converge included by-subject and by-item random intercepts, by-subject random slopes for Condition/Order and for the interaction between Condition:Order:Position, and a by-item random slope for Condition/Order. Significance values are reported using Satterthwaite's method t-tests.

Results show a significant interaction between Condition and Order, *b* = -.01, SE = .002, *t* = -4.36, *p* < .0001; speakers had shorter word onsets when the overlapping object was named first, see Figure 2.3. There was no significant effect of Condition (*b* = -.0003, SE = .001, *t* = -.2, *p* =.8), suggesting the advantage in onset times is specific to trials where the overlapping object was named first, and not an overall difference between Overlap and Different conditions. There was also no significant interaction between Condition and Position, or between Condition, Position, and Order (*p*s > .1), indicating the effect of Order was driven by differences in the onset of the first word, and not due to differing word durations or pauses throughout the utterance.

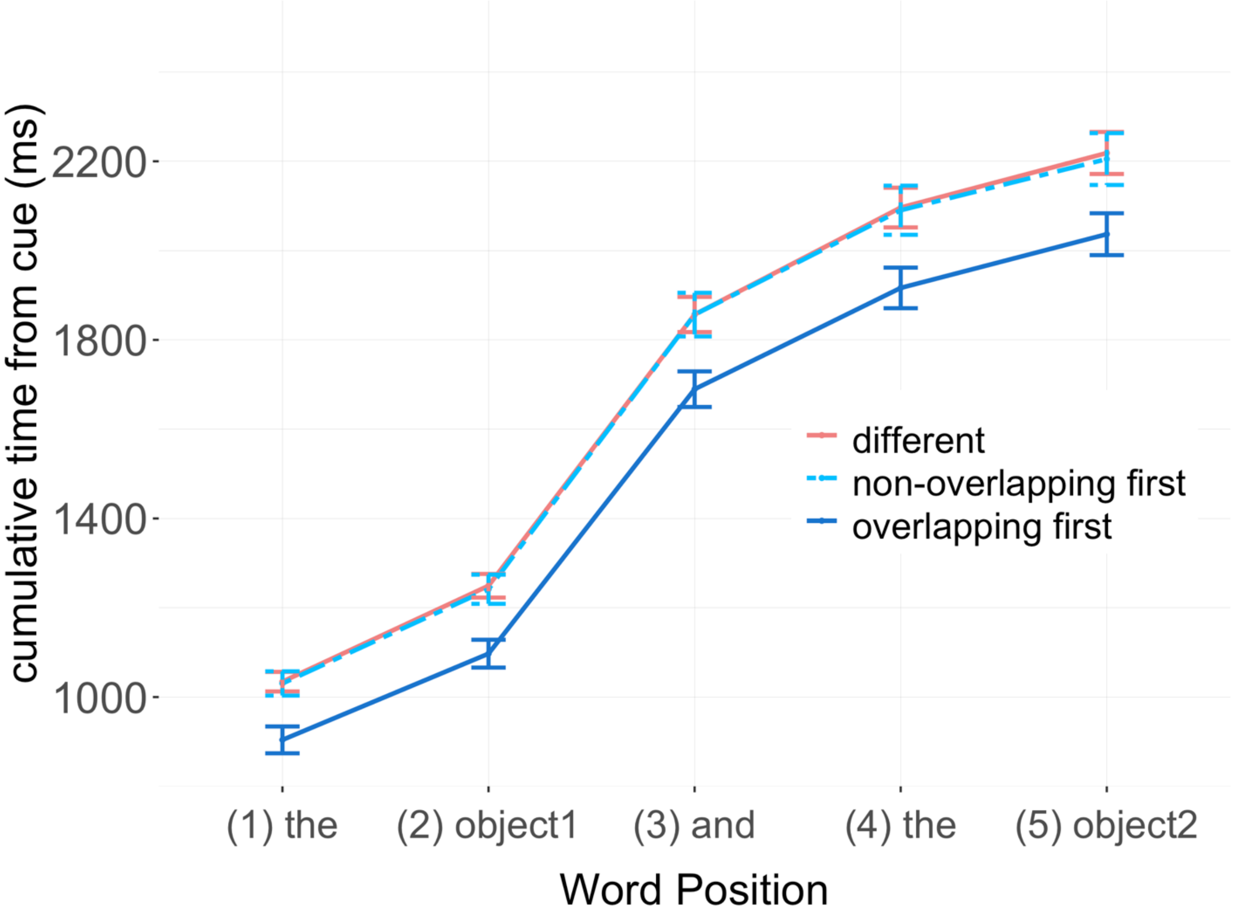

Figure 2.3. Model predictions for onsets in Experiment 1. Data from the Overlap condition are divided into trials where participants named the overlapping target first (dark blue) or the non-overlapping target first (dashed, light blue). Error bars represent ±1 SE.

Responses from the post-experiment survey showed that none of the participants were able to guess the experiment purpose. When participants were specifically asked about strategies or attempts to plan early, thirteen participants mentioned or alluded to the overlapping image as playing a part in their strategy. We re-ran the analyses excluding these participants; results held for both the order analysis ($b = .72$, $SE = .05$, $z = -13.46$, $p < .0001$) and the latencies (Condition by Order interaction: $b = -.008$, SE = .002, $t = -2.98$, $p < .01$).

Experiment 1 Discussion

Results of Experiment 1 suggest that speakers prioritized planning words that they were certain to need (the overlapping object), placing them first in their utterance plan. This early planning based on partial message information provided a benefit in onset latencies, even though speakers only began articulation after the visual cue appeared and the full message was available. As Figure 2.3 shows, latency differences emerged on the very first word, “The”. This effect is interesting because “the” is the first word to be said in every trial, independent of condition or order of the pictures described. The shorter latencies when the overlapping picture was said first suggests that speakers do not begin to produce the first phrase (e.g., “the vest”) until they have planned the noun component; that is, they do not immediately produce “the” and then pause until the noun (“vest”) is ready. Although speakers can sometimes pause or extend the pronunciation of “the” to provide more planning time for a noun, as in “Have you seen theee, uh, remote?”

(Arnold et al., 2003; Clark & Fox Tree, 2002), most production studies using pictured stimuli have found that speakers typically initiate production of phrases (e.g., “the vest”) after the entire minimal phrase has been planned (Allum & Wheeldon, 2007). The data from Experiment 1 conform to that pattern.

With the goal of replicating and extending the results of Experiment 1, we next conducted Experiment 2 as an online study that uses typed responses instead of speech. Typing presents an interesting alternative to speech because it is slower and more protracted, which could result in different planning strategies (Snyder & Logan, 2014). In our study, the slower time course of typing might affect participants’ production decisions because they have more time to plan upcoming components even after production has begun. This could provide more time to perceive the information needed for their utterance, and/or to make production decisions while typing is ongoing. Moreover, differences in the effort and time needed for typing versus speaking might affect participants’ production decisions because of the costs and benefits involved in beginning production early. For example, the increased effort and time needed for typing might encourage participants to begin production early so they can keep up with time pressures, leading to more extreme production biases compared to speech. Alternatively, it might deter participants from beginning production early because of the risk of error, as it is harder and more time-consuming to correct typed errors compared to spoken errors. Thus although our predictions for typing are more exploratory, extending our results from spoken to typed productions not only allows replicating the effect found in Experiment 1, but could also provide new insights into how decision making in language planning compares across modalities.

Experiment 2

To convert the task in Experiment 1 to an online format with typed responses, changes to the design and materials were needed. To limit the experiment length, we eliminated the familiarization phase and reduced the number of trials to 62, compared to 80 in Experiment 1. In creating the new set of items, we prioritized images that had shown high naming accuracy in Experiment 1.

We expected that participants would again prioritize more certain information and place overlap pictures first in their responses. Our predictions for latencies were less clear, given that typing is less practiced than speaking and likely to yield more variable data.

Method

Participants

Eighty-two undergraduate students, all native English speakers, participated for course credit. Given the online format of the study and concerns about participants' attention, we aimed for a larger sample size than Experiment 1. Data from 25 additional participants were excluded from analyses: four who reported technical difficulties, four due to a script error, eight non-native English speakers, three who guessed the experiment goal, and six who did not complete all trials.

Procedure

The experiment was coded in jsPsych (de Leeuw, 2015) and run online, via a lab server. Trial procedure was identical to Experiment 1, except that once the gray background appeared, a text box appeared at the bottom center of the screen. Participants could then type in their responses using any of the letter keys, the spacebar, and the delete key. The experiment lasted approximately 20 minutes.

Analyses and Results

Preprocessing

The experiment script logged each keypress and its time stamp relative to the cue appearance. A pre-processing script then parsed the key strings into words, checked naming accuracy, and extracted time stamps for the first and last keys of each word. Trials were manually coded if participants used the delete key, misspelled a word, or did not use the instructed 5-word structure (43%).

Trials in which one or both objects were named inaccurately were removed from analyses (22%); misspellings were retained. For latency analyses, responses that used a different sentence structure than instructed (18%), and responses with a total duration of more than 2.5 SD above the mean duration (2%) were removed. In cases of misspellings or use of the delete key (38%), we analyzed latencies only for the first word in the sentence, which was always the word “the”. Note that percentages refer to the percentage of remaining observations after the prior step of data cleaning. This left 3198 observations (1579 Different, 1619 Overlap; 63% of all observations) from 68 participants for latency analyses of the first word, and 1966 observations for each word in positions two through five (976 Different, 990 Overlap). Analyses were identical to Experiment 1, except that in the latency model we removed the by-item random intercept to allow convergence.

Results

As in Experiment 1, participants were significantly above chance in naming the overlapping object first (b = .78, *SE* = .1, $z = 8.06$, $p < .0001$), see Figure 2.4.

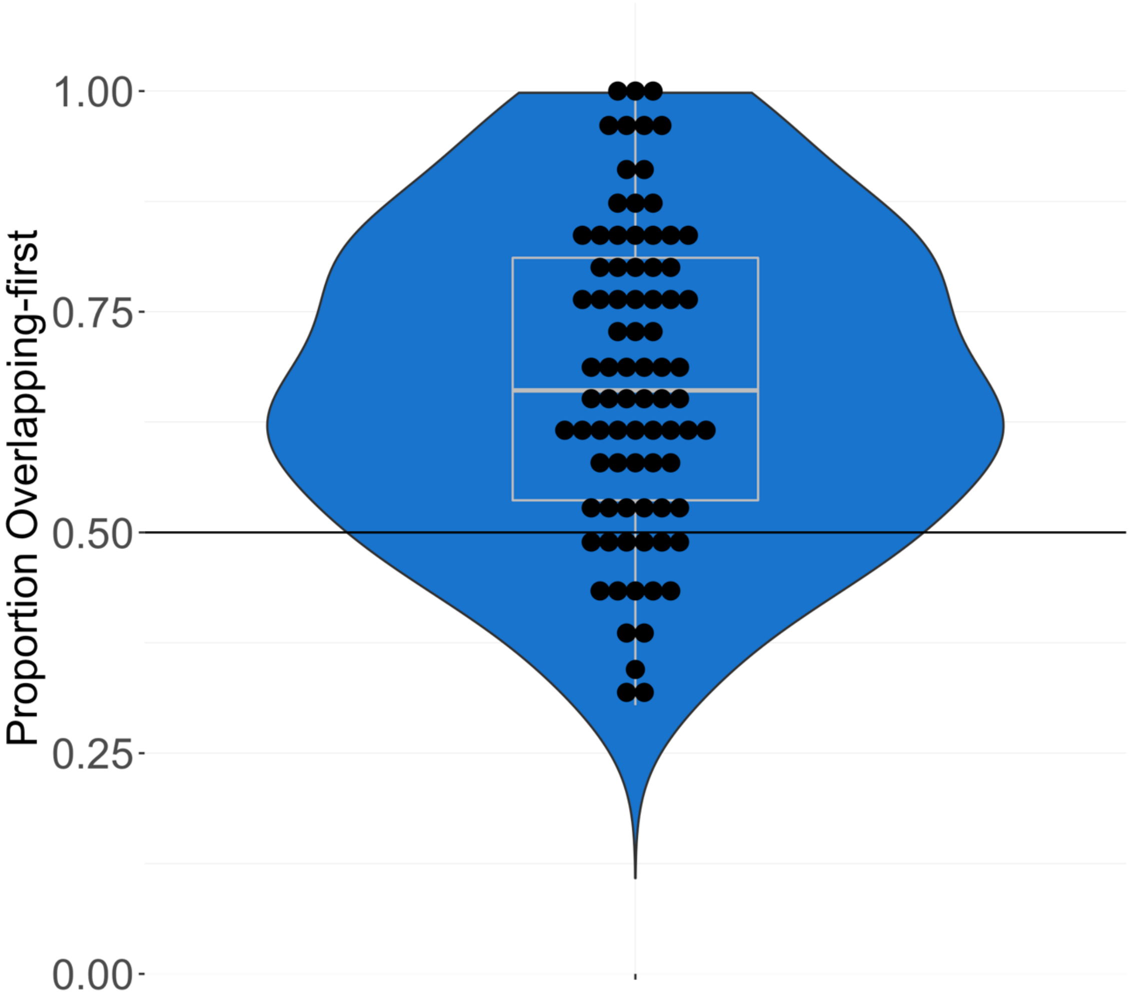

Figure 2.4. Violin plot of the proportion of trials in the Overlap condition in which participants typed the overlapping target first (Experiment 2). A boxplot is overlayed in gray to identify quartiles and the median proportion. The horizontal black line identifies chance level. Points reflect individual participant means.

Latency analyses showed a significant interaction between Condition and Order, indicating shorter onset latencies when the overlapping object was produced first in the typed response, *b* = -.006, SE = .002, *t* = -2.94, *p* < .01, see Figure 2.5. There was no significant effect of Condition (*p* > .1), but there was a marginally significant three-way interaction between Condition, Order, and the contrast of Positions [3-2], *b* = .004, SE = .002, *t* = 1.68, *p* = .09. Given the small number of observations available for word positions Two to Five, especially for the rare non-overlapping first responses, we are wary of interpreting interactions involving positions Two to Five. Future studies might explore late-position latencies further, as position interactions could suggest differences in planning during production itself. In either case, initiation latencies were shorter for overlapping-first sentences, resembling the Experiment 1 results. Moreover, when excluding twenty-two participants who alluded in the post-experiment survey to any strategy involving the overlapping image, results held for the order analysis (b = .61, *SE* = .1, *z* = 5.86, *p* < .0001), and the latencies (Condition by Order interaction: b = -.006, SE = .002, *t* = -2.63, *p* < .05, n = 52).

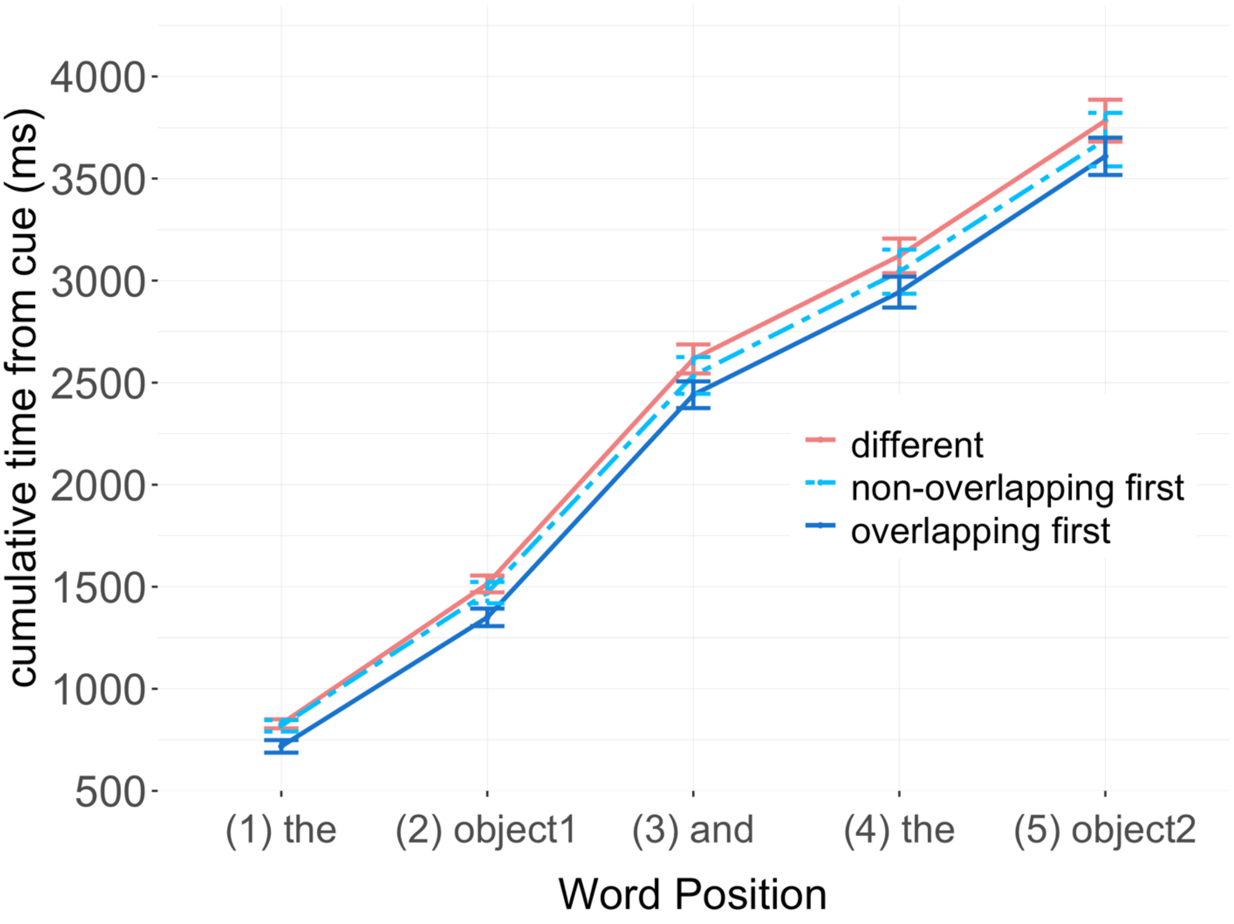

Figure 2.5. Model predictions for onsets in Experiment 2. Data from the Overlap condition are divided into trials where participants named the overlapping target first (dark blue) or the non-overlapping target first (dashed, light blue). Error bars represent ±1 SE.

Experiment 2 Discussion

Results from Experiment 2 largely replicated those of Experiment 1: participants were more likely to name the overlapping object first, and this provided a benefit in typing initiation latencies. Despite numerous differences between speech and typing that could result in different strategies for production planning (Snyder & Logan, 2014), we find similar word order biases across modalities, indicating a robust planning bias. The converging findings support our hypothesis that producers can use partial message information to begin planning their responses early, with implications for word order and response times.

One question about our findings in Experiments 1 and 2 is whether production of the overlapping object first reflects conscious strategizing and/or implicit production decisions. While conscious strategizing certainly exists in everyday language use, we do not believe that it has a major role in our findings. First, when participants were probed about strategies used to begin planning their responses early, most did not report consciously planning the overlapping object first, and our results held even when excluding those that did. This is not surprising, given that evidence of early planning is found in natural conversation (Bögels, 2020), and is therefore unlikely to implicate deliberate strategies for task performance. Second, our results suggest that participants made order choices rapidly, at an early stage during the trials. Rapid utterance planning is generally thought to proceed via implicit planning decisions (Bock, 1996; Levelt, 1989) and we propose that at least part of the bias we find emerged out of these implicit decisions rather than conscious strategizing.

However, one concern with our method is that the effects could stem from visual salience of the duplicated picture instead of, or in addition to, message uncertainty. We do not think visual salience can entirely explain our results, given that earlier speech onset was found only for overlapping-first utterances in the Overlap condition. If visual salience were playing a major role in our experiment, we would expect to find that speakers are faster to begin speaking in the Overlap condition even when they name the non-overlapping object first, because there is a duplicated picture in the display that can facilitate processing. Instead, we find that speech initiation latencies in the non-overlapping first productions are equivalent to the Control trials, where there was no visual salience of a duplicated picture.

To further rule out the possibility of a visual salience confound, we designed Experiment 3. Instead of duplicated pictures, in Experiment 3 we use a semantically-constraining question to provide early message information about the upcoming targets that participants are required to name. Using a semantically-constraining question also creates a context that better resembles natural production contexts, improving ecological validity.

Experiment 3

In Experiment 3, participants read a question under one of two conditions: 1) Semantic, where the question is semantically constraining and informative about the participant's upcoming response (e.g., *In the kitchen, what might be used to bake?*), and 2) Control, where the question does not provide any information about the upcoming targets (e.g., *In this display, what are the targets?*). Next, participants view four images on the screen. Out of the four images in a given display, one image is strongly associated with the context in the Semantic question (e.g., oven), another two are plausible responses (e.g., apple, spoon), and one is unrelated (e.g., pill). After a brief preview, a cue indicates which two images are the targets to be named in a spoken utterance. In all trials, the two targets are the image that is strongly associated with the semantically constraining question (e.g., oven) and one of the other plausible responses (e.g., apple).

We hypothesize that in the Semantic condition participants will be more likely to name the strongly-associated target first in their response, before the other target, producing utterances like *the oven and the apple*. This pattern would suggest that participants use the semantic information in the question to begin planning their responses in advance of the appearance of the visual cue indicating the targets. This result would suggest that the results of Experiment 1 and Experiment 2 were not solely due to the visual salience of the overlapping images, but rather reflect a strategy for language production under uncertainty. We also expect participants to begin speaking sooner in the Semantic condition compared to the Control condition, consistent with the results of the prior studies. Within the Semantic condition, we expect participants will begin speaking sooner when naming the strongly-associated target first, again reflecting patterns in Experiments 1 and 2. This result would suggest that early utterance planning based on a semantic cue could provide a benefit in latencies, making it an efficient strategy for planning under uncertainty.

Method

Participants

Undergraduate students (n = 101), all native English speakers, participated in the experiment online for course credit. Sample size was determined using a power analysis conducted using PANGEA (Westfall, 2016), aiming for at least 80% power with a medium effect size (*d* = 0.5) in the analysis of word order. Data from 46 additional participants were excluded from analyses: 21 who had empty audio files, 14 who did not comply with task instructions, 7 who did not complete all trials, and 6 who had corrupt audio files. Experiment 3 was pre-registered prior to beginning data collection (https://osf.io/th9ck).

Stimuli

Object images from the MultiPic database (n=160, Duñabeitia et al., 2017) were used as stimuli, assigned to 40 item displays of four objects each. Each display was assigned a Semantic question (e.g., *In the kitchen, what might be used to bake?*), providing a semantic cue for the upcoming target images; and a Control question, with no information about the upcoming images (e.g., *In this display, what are the targets?*). Out of the four images in a given display, one image was strongly associated with the context in the Semantic question (e.g., oven), another two were plausible responses (e.g., apple, spoon), and one was unrelated (e.g., pill).

Before using the stimuli in our main experiment, we ran a norming study on a separate sample of 60 participants. We first created a list of the semantically-constraining questions and four object images that could answer each question. Each participant was presented with each question and asked to rank the four images in order from 1 (the most appropriate for answering the question) to 4 (least appropriate). The image with the highest average ranking was then designated as the strongly-associated object (*M* = 1.42, *SD* = 0.36). Images ranked third (*M* = 2.82, *SD* = 0.28) and fourth (*M* =3.42 , *SD* = 0.34) were used as plausible responses. The image ranked second (*M* = 2.21, *SD* = 0.46) was discarded from that display in order to reduce competition from the first-ranked image and increase the strength of the manipulation. Instead, each second-ranked image was used as the distractor of another display, for a total of four images in each trial display: one strongly-associated target, one plausible target, one plausible competitor (non-target), and one distractor.

The mean number of words per question in the experiment was 7.42, *SD* = 1.54 (Semantic *M* = 8, *SD* = 1.34; Control *M* = 6.8, *SD* = 1.54). Questions in the Semantic condition were designed to include the constraining information early, in the first half of the question.

Questions in the Control condition asked about images in the display without mentioning semantic content, such that they could be plausibly answered with any of the image names.

Procedure

The experiment was coded in jsPsych (de Leeuw, 2015) and run online, via a lab server. Participants first read the task instructions on screen, including example displays and one practice trial with feedback, and then proceeded to the experiment. Figure 2.6 illustrates the trial sequence. Each trial began with a question presented visually in the center of the screen, either in the Semantic or Control condition. Next, participants clicked on a green button in the center of the screen, and the question was replaced by a display of four images rotating clockwise in a circle. After a brief exposure of approximately 3.8 seconds, the two target images were framed with black squares. Participants' task was to name these two target images in a conjoined noun phrase ("the oven and the apple"), as a response to the prompt question. Participants were told that another participant would later listen to their recordings to identify the targets, and that they should use the simple single-word labels for the objects (e.g., *oven* and not *gray oven*). The trial ended approximately 5.8 seconds after the targets were cued with the black frames.

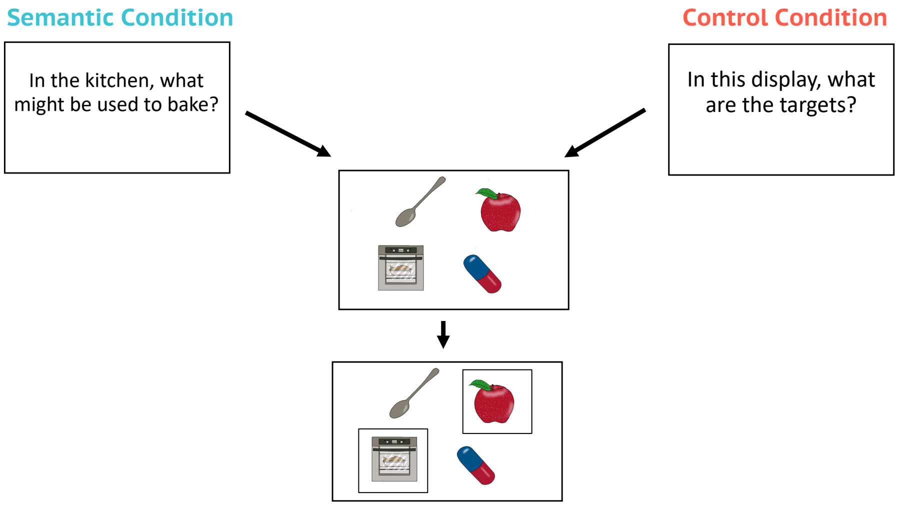

Figure 2.6. Example of a trial sequence under the Semantic and Control conditions. Participants first read the question on screen, and then were presented with the four images. The four images rotated clockwise to avoid screen position effects. After 3.8 seconds of a preview, square frames appeared around each of the target images, and participants named the targets in a conjoined noun phrase.

Participants were randomly assigned to one of two lists, each consisting of twenty trials in the experimental condition and twenty trials in the control condition. Condition was counterbalanced across stimulus lists, such that each participant only viewed a given display once, but each display appeared in both conditions approximately equally across participants. The starting position of each image in each display was randomized by the experimental script. Trial order was randomized per participant. The entire experiment lasted under 15 minutes.

Analyses and Results

Preprocessing

Speech files were transcribed as in Experiment 1. Trials in which one or both objects were named inaccurately (38%) were removed from analyses. Montreal-Forced-Aligner (MFA; McAuliffe et al., 2017) was used to extract onset and offset times of each word in each utterance. At the time of the analysis for Experiment 3, MFA was more easily compatible with newer versions of Python compared to other aligners and therefore preferred. Before performing latency analyses, we further removed trials with disfluencies (6%), responses that did not follow the 5-word conjoined noun phrase structure (65%), and trials where participants began speaking before the cue (8%). Note that percentages refer to the percentage of remaining observations after the prior step of data cleaning. This left 759 observations (372 Semantic, 387 Control; 19% of all observations) from 63 participants for each word position in the latency analyses.

Word Order

We first tested whether participants were more likely to name the strongly-associated object first in the Semantic condition compared to the Control condition. We ran a mixed-effects logistic regression model regressing Order (strongly-associated object first vs. plausible object first) on Condition (Semantic vs. Control). The model included by-participant and by-item random intercepts, and by-participant and by-item random slopes for Condition (the maximal random effects structure; Barr, 2013). As expected, participants were significantly more likely to produce the strongly-associated object first in the Semantic condition compared to the Control (b = .27, *SE* = .1, *z* = 2.75, *p* < .01), see Figure 2.7.

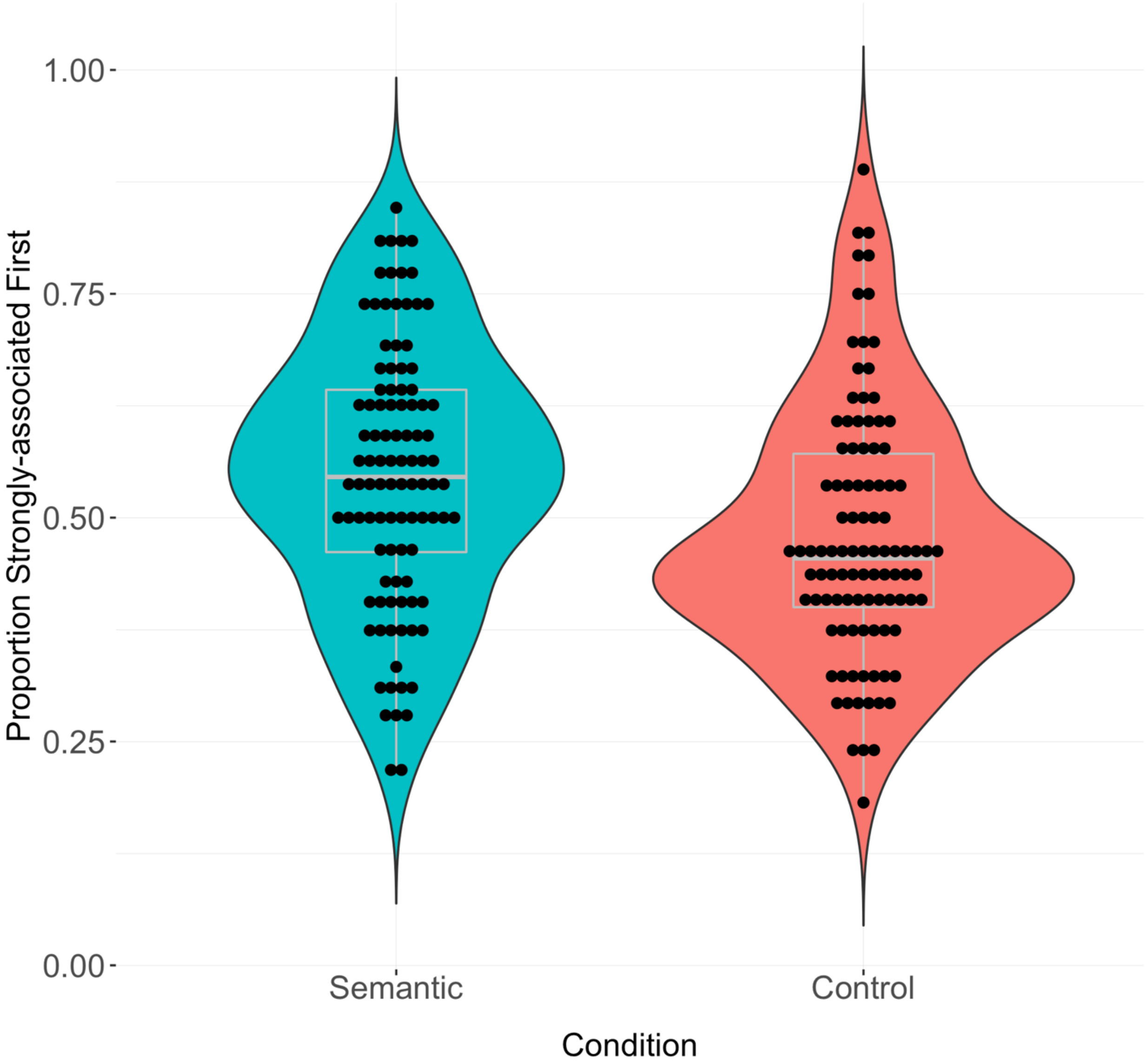

Figure 2.7. Violin plots of the proportion of trials in which participants named the strongly-associated object first, in the Semantic condition and the Control condition in Experiment 3. Boxplots are overlayed in gray to identify quartiles and the median proportion. Points reflect individual participant means.

Latencies

We next examined speech initiation latencies (onsets) for each word in participants' utterances. As in the previous experiments, onsets were calculated as the time between the cue appearing and the first phoneme of each word and were log-transformed for analyses. The predictor variables were word Position (1-5), Condition (Control, Semantic; coded [0,1]), and Order of naming (plausible first, strongly-associated first; coded [0,1]). Position was coded using successive difference contrasts, where each word position is compared to the preceding word position (2-1, 3-2, 4-3, 5-4).

We ran a mixed-effects linear regression model regressing Onset on Position, Condition, Order, and their interaction. The random effects structure included by-subject and by-item random intercepts, and by-subject and by-item random slopes for Condition, Order, and their interaction. Significance values are reported using Satterthwaite's method t-tests.

There was no significant effect of Condition, Order, or their interaction (*p*s > .05), see Figure 2.8. Numerically, the estimated marginal means for onsets at the first word (latencies to begin speaking) were shorter when the strongly-associated object (e.g., oven) was named first in the Semantic condition compared to all other cases. This pattern was maintained throughout all five word positions and even increased with sentence progression. The numerical pattern therefore aligns with our hypothesis, but the interaction between Order and Condition was not reliable.

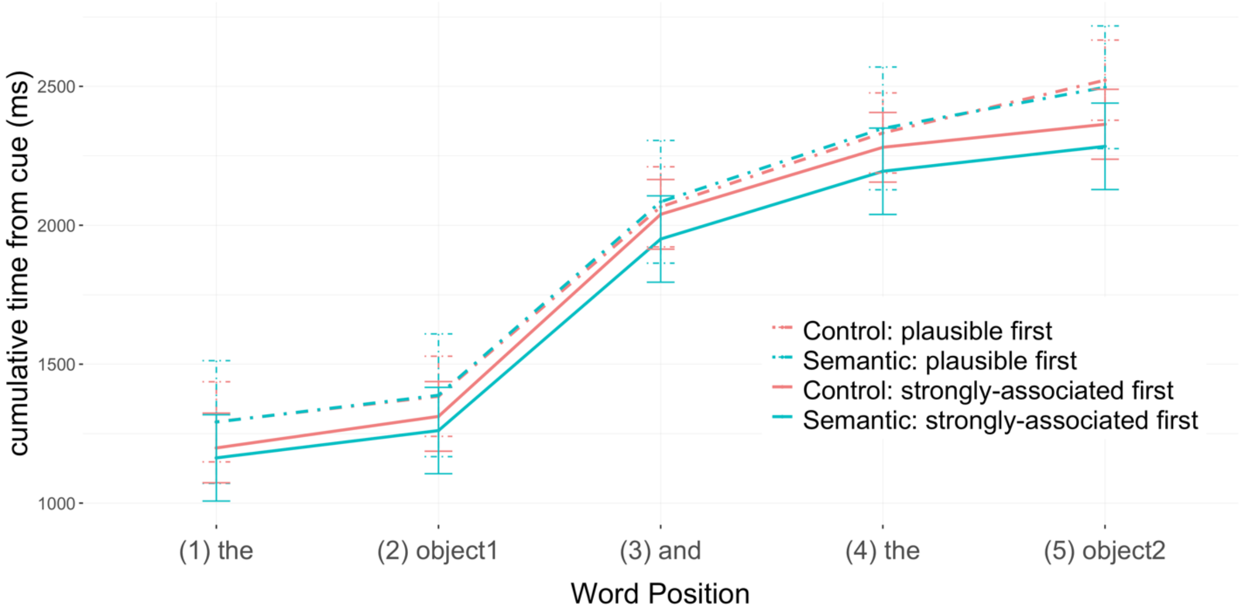

Figure 2.8. Model predictions for onsets in Exp 3. Data are divided into trials where participants named the strongly-associated target first (e.g., oven; solid lines) or the plausible target first (e.g., apple; dashed lines), in the Semantic condition (blue) or the Control condition (pink). Error bars represent ±1 SE.

Experiment 3 Discussion

Results from Experiment 3 align with our findings in Experiments 1 and 2: participants were more likely to name the strongly-associated object first in the Semantic condition, suggesting they used information from the preceding question to begin planning their responses early – affecting word order choices. This finding supports our argument that visual salience alone cannot explain our results in Experiments 1 and 2, as there were no duplicated images in Experiment 3 that could affect participants' behavior.

Note also that in Experiments 1 and 2, the Overlap trials allowed participants to predict with complete accuracy what one of the targets was going to be (the overlapping target), while uncertainty remained around the second target. In Experiment 3, participants did not have complete certainty about any of the targets, but appear to have used semantic-based predictions to begin planning the most likely target first (the strongly-associated target). In both cases we find that speakers plan their utterances using any available information, indicating a robust effect of message uncertainty on utterance formulation, across these variations in uncertainty type. Moreover, the results of Experiment 3 suggest that decisions about word order and early planning do not necessarily require complete certainty about a specific message component for utterance planning to begin, but rather could be based on speakers' predictions. Presumably, a strong enough prediction is worth the risk of an error in utterance planning (i.e., in case a planned word is not the correct target). Although we were underpowered to analyze changes in productions over multiple trials, future studies could examine how participants' strategies evolve over the course of the experiment, e.g., whether they rely more strongly on their predictions for early planning, learn to refine their predictions, and more.

While the word order effect was robust across experiments, results from the latency analysis in Experiment 3 were inconclusive. The numerical pattern aligned with our hypothesis – shortest speech initiation latencies were found for trials in the Semantic condition where the strongly-associated object was named first – but the difference was not statistically significant. It is possible that we had too few observations to detect the expected interaction effect (between Condition and Order) after data exclusions. Owing to constraints for finding item sets to align with the Semantic question condition, Experiment 3 had fewer items than in Experiments 1 and 2. Moreover, data exclusions were more common due to varying sentence structures and inaccurate object naming. The higher exclusion rate is likely because of the online format of our experiment, which had no familiarization phase for the object names and no interaction with research personnel. All of these factors could decrease the motivation to respond quickly and accurately, and decrease adherence to experiment instructions more generally. Future experiments might maximize the manipulation strength by using a confederate interlocutor or a deadline that encourages participants to begin speaking more quickly, or exert more control on participants' sentence structures in order to maximize power for detecting an effect. Nonetheless, the results of the word order analysis were robust, supporting our hypothesis that participants use early semantic information to begin planning their responses even when only partial message information is available.

Chapter Two General Discussion

We developed two novel methods to manipulate the uncertainty of producers' messages in a picture-naming task. Results showed that word order choices varied with degree of message uncertainty: producers prioritized message components likely to be needed, resulting in early placement of these elements in the utterance. These findings suggest that speakers make early utterance decisions based on available information, allowing them to begin speaking sooner, and then continue planning the rest of the utterance when more information becomes available. Our results were also comparable across spoken and typed modalities, suggesting similar strategies for planning under uncertainty, regardless of output modality.

Although this initial investigation was necessarily more constrained than natural conversations, it is likely that similar patterns will emerge in conversation. Speakers often begin planning their response to an interlocutor before the interlocutor has finished speaking (Bögels, 2020; Bögels et al., 2015; Corps et al., 2018), suggesting that sometimes utterance planning precedes knowledge of the full information needed to reply. To explain how speakers begin early planning in conversation, turn-taking researchers typically point to predictive comprehension, which suggests that comprehenders can use context cues to predict what the upcoming words or messages are going to be (Kuperberg & Jaeger, 2016). Using these predictions, speakers can begin planning their own response even before the interlocutor has finished their turn (Corps et al., 2018; Levinson, 2016; Levinson & Torreira, 2015).

But predictions are, by definition, uncertain. Depending on the degree of constraint in the sentence, the degree of the listener's confidence in their prediction may vary (Klimovich-Gray et al., 2019; Lewis & Bastiaansen, 2015; Luke & Christianson, 2016). Predictions can also be somewhat vague, i.e., the listener might have a general idea of what the speaker is going to say but still be unsure how exactly the utterance will turn out. Even trained coders often disagree on when exactly in an incoming question the answer becomes clear (Bögels, 2020). This result suggests that response planning based on predictions must also carry some uncertainty; if the speaker does not know exactly what they are responding to, they cannot know exactly what to plan in their response utterance, and production strategies might vary.

Our study expands on these ideas in several ways. First, prior turn-taking work focused on stimuli with high message predictability, e.g., general knowledge questions with short answers, manipulating whether the required answer became clear mid-question (allowing early planning) or only at the end (Bögels et al., 2015). In Experiments 1 and 2 presented here, timing was carefully controlled, while we manipulated *how much* information participants had about their message, and consequently, *how much* of their utterance they could plan. This type of message uncertainty resembles conversational contexts where speakers begin planning their response when they only have partial information about what they are responding to. In Experiment 3 we introduced early semantic cues indicating *how likely* certain components are to be included in the message, and found that speakers prioritize those likely components in their response. Using semantically-constraining questions in Experiment 3 allowed a more ecologically valid context for our experiment, as natural conversation similarly contains a semantic context that constrains speakers' messages. Notably, although our participants were told to respond promptly, they had plenty of time to produce their utterances and no penalty for slow responses. The fact that they still attempted early planning, even under uncertainty, might suggest that this production strategy is well-practiced from their own language experience, becoming natural even without obvious time pressure.

Another novel aspect of our findings is that uncertainty affected participants' utterance forms, as they chose word orders that allowed early planning, potentially over alternatives that would have prioritized the listener's needs. Recall that participants were told that another participant would later listen to their recordings to identify the targets. From an audience-design perspective emphasizing getting the information to the listener as early as possible, naming the *non-overlapping* object first in Experiments 1 and 2 would be the most efficient word order for the listener to hear, because a non-overlapping object immediately identifies the target pair and allows the listener to complete the task goal. Similarly, naming the *less likely* object in Experiment 3 (the plausible object, but not the strongly-associated object) would be more informative to the listener, because the listener could be quite confident that the strongly-associated object would be a target, but more uncertain about which of the plausible objects would be the other target. But participants still named the overlapping (or strongly-associated) object first, suggesting that early planning for the speaker was prioritized over early informativity for a future listener, consistent with studies suggesting that there are limits to the degree to which producers accommodate listeners' needs (Horton & Keysar, 1996). Indeed, utterance form choices often reflect speakers' use the flexibilities of language to mitigate production demands (MacDonald, 2013), and planning under uncertainty is a particular type of production demand.

Given the current results, natural further steps include testing participants in conversational settings with interlocutors for a more ecologically valid context, and varying the degree or type of message uncertainty to see how planning biases are affected. Another step is to begin exploring the neural correlates of production under message uncertainty, especially given that effects of competition in early planning can sometimes be detected in the neural signal but not in behavioral measures such as reaction time (Piai et al., 2020). More generally, more research on production under uncertainty would not only address a common everyday situation that is under-studied in the laboratory, but could also shed light on the interaction between various production processes, including early planning, message formulation, and word order choices.

Chapter Three: The time course of event description under message uncertainty

In Chapter Two I demonstrated that in situations of message uncertainty, speakers tend to choose word orders that maximize production efficiency – prioritizing the more certain components and placing them earlier in the utterance, allowing production to begin sooner as the remaining information unfolds. However, not all sentence structures and messages allow for such flexibility in word order. What other strategies can speakers use to plan under message uncertainty when the word order is fixed? How will they affect speakers' utterance plans?

In Chapter Three, I present three experiments investigating the time course of speakers' utterances in a motion event-description paradigm. The task resembles contexts of sports narration described earlier (Popov, 2019), where speakers must describe an unfolding event despite not knowing how it will end. The paradigm is also similar to the experiments in Chapter Two in that a key component of the message is only revealed late in the trial. Other message components are available from the outset, however, allowing for some early planning. Crucially, I use a cross-linguistic comparison of Spanish and English to show how the permitted word order of each language affects the time course of event description under message uncertainty.

Methodologically, Chapter Three allows examining a more naturalistic situation compared to Chapter Two experiments, as participants simply describe an unfolding event rather than naming cued pictures. Moreover, participants will produce more complex sentences (including a subject, verb, object) rather than simple conjoined noun phrases. Finally, we also introduce more human interaction in these experiments, increasing the salience of social considerations in utterance production.

Experiment 1

In Experiment 1 participants view a video of an unfolding motion event, where an animate entity moves from a source location to a goal location on the opposite end of the screen. Figure 3.1 shows an example display[3], with a butterfly (the *actor*) that moves from a bench (the *source*) to a slide (the *goal*). However, because there are two potential goal locations on the screen, the target goal is not known at the start of the trial. The trial then unfolds according to one of two conditions: In the Direct condition, the actor moves in a direct path towards the target, making the target goal immediately known once the movement begins. In the Uncertain condition, the actor moves in a path that is equally between the two goals, creating temporary ambiguity about the target goal. Only toward the end of the trial, the actor turns toward the target goal and away from the competitor, thus completing the event and disambiguating the message information.

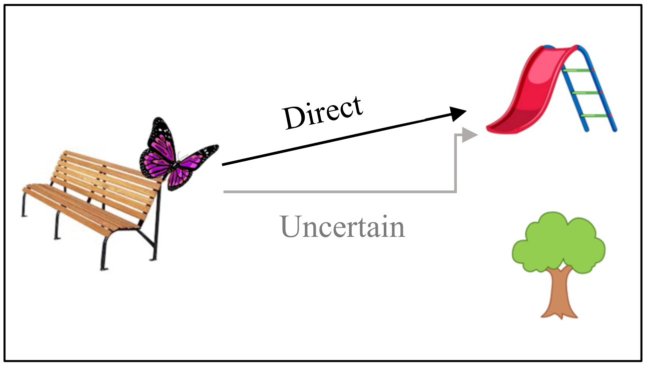

Figure 3.1. Example of a trial display in Experiment 1. Arrows and labels are for illustration only, demonstrating the trajectory for the Direct condition (black arrow) and the Uncertain condition (gray arrow).

---

[3] I created this particular example display to use consistently throughout Chapter Three, but see Appendix 1 for full stimuli lists.

We hypothesized that in both conditions, speakers will use any available information to plan their utterances. In the Direct condition, speakers might plan and even produce the entire utterance even before the actor reaches the target goal. In the Uncertain condition, a speaker might begin describing the scene ("the butterfly is moving…") even before they know where the butterfly is moving to, and continue planning the rest incrementally, as more information becomes available. This incremental planning should affect the time course of production, causing longer utterance durations and more disfluencies in the Uncertain compared to the Direct condition, because participants will still be planning their utterance while production is ongoing and the event is unfolding.

Speech initiation time might not differ greatly between conditions, as both conditions allow for some early planning of the utterance even before the event is completed ("the butterfly is moving to…") – given that the actor and the verb are apparent from the start of the trial. However, we expected a trade-off between speech initiation times and disfluencies and/or durations: if speakers wait for the entire message to become disambiguated, they might begin speaking later but can avoid disfluencies. Alternatively, if speakers begin speaking as soon as they have minimal message information, their speech initiation times will be short but their utterances might include more disfluencies. Finally, participants also have the option of mentioning the source in their response – that is, the location the actor leaves from (e.g., *the butterfly moves* *from the bench* *to the*…). If speakers want to begin speaking early but maintain fluency in their utterance, they might be more inclined to mention the source in the Uncertain condition, buying them more time until the goal is disambiguated.

Taken together, although our approach in Experiment 1 is rather exploratory, we do expect to see differences in speech patterns according to the message uncertainty condition –

Direct versus Uncertain. Importantly, speech patterns could indicate whether speakers begin early planning even in the Uncertain condition, where part of the message is only revealed at the end of the trial. If so, this would suggest a tendency to begin utterance planning even under uncertainty, taking advantage of incremental planning to allow early production.

Method

Experiment 1 was pre-registered prior to examination of the data. The experiment was hosted on a lab server and carried out remotely, with data recorded through participants' microphones and sent directly to the server.

Participants

60 undergraduate students, all native English speakers, were recruited in exchange for course credit. 35 additional participants were removed from analyses due to corrupt audio files (n = 14), server error (n = 8), failure to follow instructions (n = 6), majority of recordings cut off (n = 4), and excessive background noise (n = 3). During data analysis we discovered two additional participants who never used a verb in their utterance (e.g., *the butterfly to the bench*), and five participants who never used an article before the goal (e.g., *the butterfly moves to bench*). Because these were critical points of analysis, we removed data from those participants from any further processing.

Stimuli

Images from the MultiPic database (Duñabeitia et al., 2017) were used to create 50 displays of four images each: one actor and three objects. Actor images depicted either animate entities or vehicles. For each display item, each of the three object images was randomly categorized as either the source, the target goal, or the competitor goal. On screen, the goals and the source locations appeared at an equal distance from the screen margin on their respective sides (right versus left, counterbalanced). The source was placed at the horizontal center of the screen. The two goals appeared on the opposite side of the source, at an equal distance from the source and an equal distance from the horizontal center (top versus bottom, counterbalanced).

Procedure

Participants completed the experiment online in a Zoom session with a research assistant (RA). The RA greeted the participant and introduced themself as the participant's "experiment partner". The RA explained that once the experiment begins, both of their cameras will be off, and only the participant's microphone will remain on. Next, the RA sent the participant a link to the experiment hosted on our lab server. Once the participant accessed the experiment in their browser, the RA turned off their own microphone but remained in the Zoom session.

Participants first completed a familiarization phase for the 200 images used in the experiment. Participants viewed each image with its appropriate label written underneath, and pressed the Enter key to continue to the next image. After the familiarization phase, participants read the task guidelines instructing them to describe the simple motion events they would view. Participants were told that on every trial, their experiment partner (the RA) would have to identify in real time which scene they were describing. An example of a scene and description sentence were provided within the instructions.

After reading the instructions, participants were told to inform their experiment partner that they are ready to begin the practice trial. Then, once the RA turned on their microphone gave the participant permission to continue, the practice trial began. When participants produced the target utterance, the RA said “okay, got it!” to give them feedback and create a collaborative game environment.

Finally, the experimental phase consisted of 50 trials. Participants were told that their experiment partner’s microphone will be off for the rest of the experiment, so they will not be able to hear any feedback. The entire experiment lasted approximately 25 minutes.

Each trial in the experimental phase began with the display of the four images. After one second of preview, the actor began moving toward the target goal, in either a Direct or an Uncertain trajectory. The movement lasted five seconds in both conditions. In the Uncertain condition, movement in a straight path in between the targets lasted four seconds, and the turn towards the goal lasted one second. In the Direct condition, the actor moved in a direct diagonal path towards the goal for the entire five seconds. The actor then remained at the target goal for an additional two seconds before the trial timed out. In total, each trial lasted 8 seconds; see Figure 3.2 for the trial timeline. Once the trial was completed, a new screen appeared with a “next” button. When participants clicked on it, a fixation cross appeared for one second, and then the next trial began.

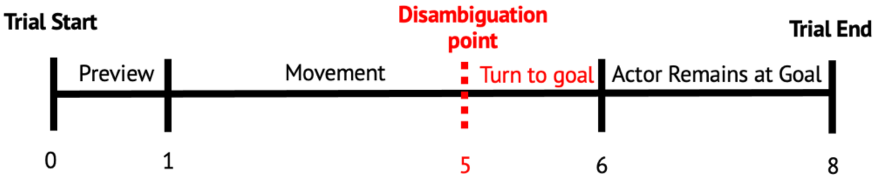

Figure 3.2. Trial timeline. The disambiguation point (marked in red) was only relevant in the Uncertain condition. Trials began with one second of preview before the actor started moving. In the Direct condition, the actor moved directly in a diagonal path toward the goal target and

reached it after 5 seconds of movement. In the Uncertain condition, the actor moved in a straight line between the two goals for 4 seconds, then turned towards the target goal (disambiguation point; marked in red) for an additional 1 second of movement until arriving at the target goal. In both conditions, the actor then remained at the target for an additional 2 seconds until the trial timed out.

Participants were randomly assigned to one of two lists, each consisting of 25 trials in the Direct condition and 25 trials in the Uncertain condition. The condition in which each item appeared was be counterbalanced across lists. The experiment script pseudorandomized movement direction (left versus right) and placement of target (top versus bottom) for each trial, with each of the four possible combinations occurring on a random quarter of the trials. Trial order was randomized per participant.

Analyses and Results

Preprocessing

Speech files were transcribed by trained research assistants. They first transcribed the entire response given by the participant per each trial. Next, transcribers annotated seven components of each response: 1) article before the actor, 2) actor name, 3) verb, 4) prepositional phrase, 5) article before the goal, 6) goal, 7) source.

In contrast with the Experiments described in Chapter Two, here our aim was to include the variability in lexical items and sentence structures produced by participants, given the more naturalistic descriptions and longer, more complex utterances. We therefore did not discard trials where participants used different labels than intended, and we included all variations of verbs and prepositional phrases (e.g., the butterfly moved to the slide, the butterfly is going toward the slide, etc.). Although we had initially planned to analyze whether participants mentioned the source location in their response (e.g., the butterfly moved *from the bench* to the slide) as a function of condition, data inspection showed little within-participant variability on this factor (Appendix 2). We therefore included Source as a control variable in the statistical models but did not analyze it as a predictor of interest. The near-lack of within-participant variance was not surprising; structural priming effects are well-attested in the literature of language production (Branigan, 2007; Jacobs et al., 2019; Smith & Wheeldon, 2001) – participants who produce one sentence structure are likely to self-prime themselves to continue producing it throughout the experiment.

Montreal Forced Aligner was used to extract onsets and offsets of each word produced in the utterance, including spaces between words. Because utterances varied in structure and number of words, we could not know for certain what word position each sentence component was produced in (i.e., whether the actor was the first or the second word in the utterance; "the butterfly moved…" vs "butterfly moved…", "the pink butterfly moved", etc.). To match between word positions and relevant components, our preprocessing script drew on the transcribers' annotation of each component, by searching for the annotated component within the forced alignment output. For example, to find the speech onset of the actor name, the script first searched the transcription to locate what word the participant used to name the actor, then searched the forced alignment output to find that word (agnostic of which serial position it is in), and then returned the onset time for the actor name.

For each utterance we calculated the onsets and durations of six main chunks: 1) speech initiation latency, 2) Actor Noun Phrase (Actor NP), including the determiner and the noun, 3)

Verb phrase, 4) Prepositional phrase (PP), 5) Article before the goal, and 6) Goal noun. Each chunk was calculated by subtracting the offset of the component from the offset of the previous component; Figure 3.3 illustrates the chunking scheme. Any spaces between components were included with the later chunk. The entire utterance duration (Total Duration) was calculated by subtracting the speech latency onset from the offset of the goal component. Note that given the variability in the number of words produced per utterance and per chunk, and given our interest in phrase durations rather than speech onsets per word, the analysis in Experiment 1 examines chunk durations instead of word onsets (which were used in Chapter Two). Before performing statistical analyses, we removed trials with missing components: utterances that did not mention the actor (2%), verb (4%), prepositional phrase (4%), article (7%), and goal object (9%). Additionally, we removed trials that failed forced alignment (5%; typically due to excessive noise) and utterances that we were unable to chunk due to excessive participant description (n=18). There were too few observations in each of these categories to allow for meaningful statistical analyses between conditions.

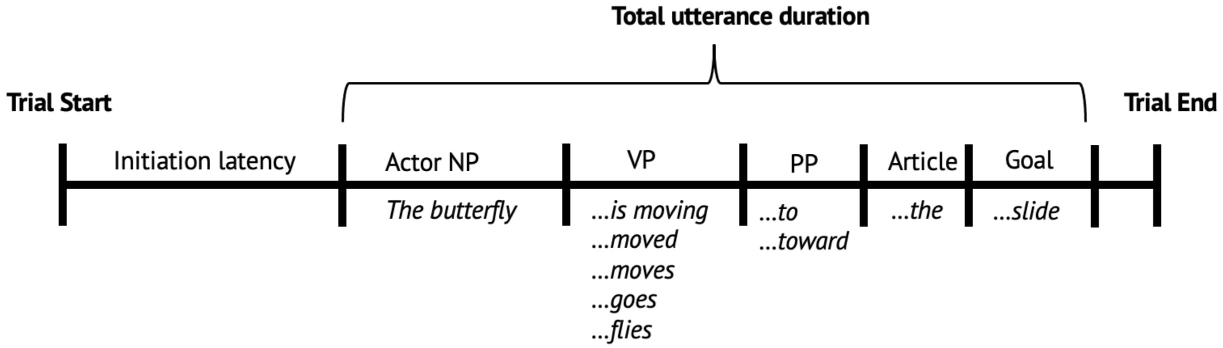

Figure 3.3. Chunking scheme in Experiment 1, illustrated with an example of various phrases that could be used for the same trial. Three leading dots indicate that any spaces between

components were included with the later chunk. Note that spaces do not necessarily correspond to silent pauses; spaces could also include additional words that participants inserted between components.

Throughout data transcription, research assistants manually coded for each utterance whether it included a disfluency or not (*e.g., the butterfly moved to the um slide, th-the butterfly moved to the slide*), and whether the recording was cut off or not (e.g., *the butterfly moved to the sl–*). Disfluencies were coded as 0 or 1, and the particular type of disfluency was not specified – i.e., filler, repetition, correction mid-word, etc. Transcribers were instructed not to code pausing or word lengthening (e.g., "theeeee butterfly") as a disfluency, given the subjectivity of this judgment, and that lengthening would be captured in our durations analyses.

Because the manually coded disfluencies varied in nature (e.g., addition of several words versus repetition of only one phoneme), they could cause wide variation in duration that is difficult to control for. A subset of these disfluencies, however, were fillers (e.g., "um", "uh"). Fillers are much more uniform in nature (single-syllable phrases) and were included in our hypotheses when designing the experiment, so we wanted to retain them in similarity to word lengthening and silent pauses. We therefore used an experimental script to code for each utterance whether the participant produced a filler. We then created a separate category for these fillers versus all other manually coded disfluencies. Utterances with fillers were included in analyses, in similarity to lengthened productions and silent pauses. All other, uncategorized, disfluencies were removed from our main analyses (1%). Finally, we removed utterances where the last word was cut off due to the trial timing out (2% in the Uncertain condition, 1% in the Direct condition).

After these steps of data cleaning, one participant was left with less than 25% of trials in at least one condition and was removed from further analyses. Of the remaining 60 participants, the average percent of retained trials was 78% in the Uncertain condition (SD = 15%, range 36%-100%) and 83% in the Direct condition (SD = 14%, 32%-100%).

Durations Analysis

We first examined the effect of condition on the duration of each chunk in participants' utterances. Figure 3.4 displays the averaged chunk durations in each condition, on a timeline of the trial procedure. Figure 3.4 also distinguishes between components and spaces between components, but note again that for duration analyses, spaces were included with the following chunk.

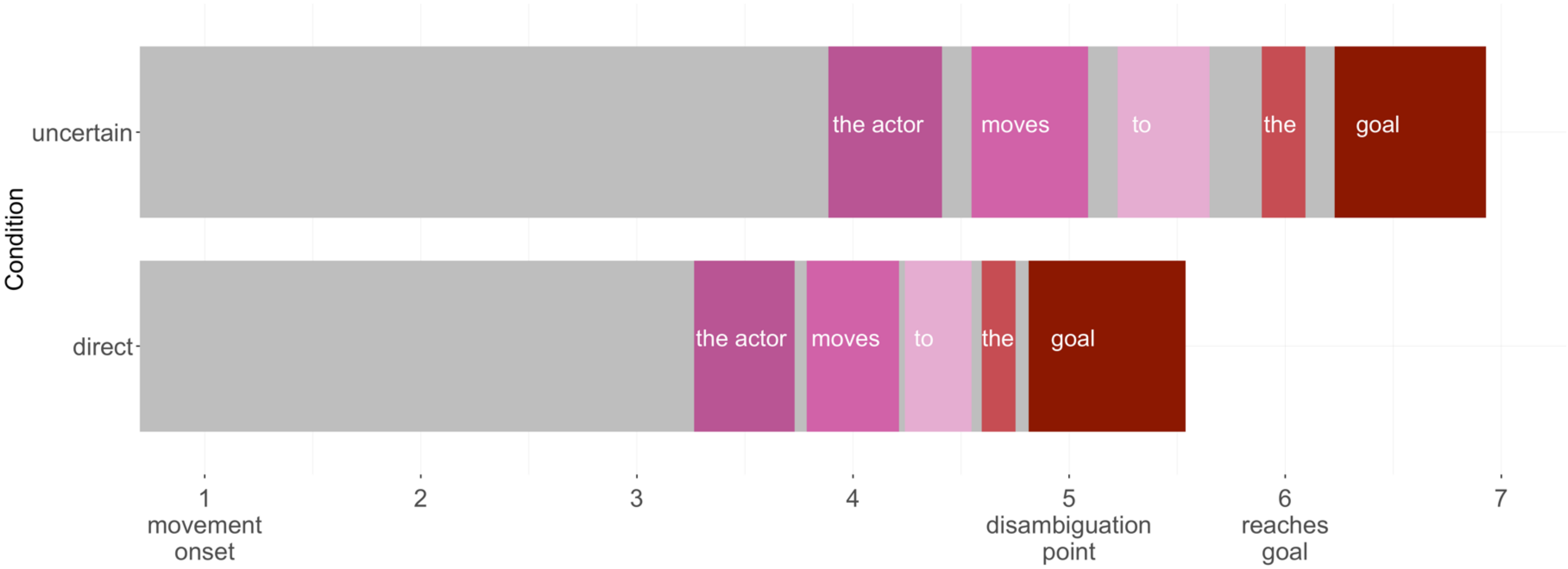

Figure 3.4. Average chunk durations illustrated on the trial timeline, for the Uncertain (top panel) and Direct (bottom panel) conditions. Each component is displayed in a different color with overlayed text to identify the component. Gray bars indicate initiation latency (time from trial start until Actor NP onset) and spaces between components. For simplicity, only trials where

no source was mentioned are displayed here. Note that trials began at 0 seconds and lasted until 8 seconds, but those time points are not displayed for graphing purposes.

For each chunk (Latency, Actor NP, Verb, PP, Article, Goal), we ran a separate linear mixed effects model regressing Duration on Condition (Uncertain vs Direct; coded [-.5,.5]). We further controlled for participants' mention of source, which would naturally lengthen their utterances, by adding a predictor for Source (No-source vs Source; center-coded [-.5, .5]). The random effects structure included by-subject and by-item random intercepts, and by-subject and by-item random slopes for Condition. Significance values are reported using Satterthwaite's method t-tests. Figure 3.5 displays the difference between conditions at each chunk position.

Results show that durations were longer in the Uncertain condition compared to the Direct condition, across all utterance chunks before the goal: speech initiation latency ($b = 0.12$, SE = 0.02, $t = 5.47$, $p < .0001$), Actor NP ($b = 0.1$, SE = 0.02, $t = 4.92$, $p < .0001$), Verb Phrase ($b = 0.18$, SE = 0.04, $t = 5.006$, $p < .0001$), Prepositional Phrase ($b = 0.404$, SE = 0.06, $t = 7.12$, $p < .0001$), and Article ($b = 0.43$, SE = 0.08, $t = 5.37$, $p < .0001$). However, there was no significant difference at the Goal position itself ($b = 0.02$, SE = 0.03, $t = 0.6$, $p > .1$).
The total utterance duration, from noun phrase onset until goal object offset, was significantly longer in Uncertain compared to Direct trials; $b = 0.25$, SE = 0.03, $t = 8.12$, $p < .0001$.

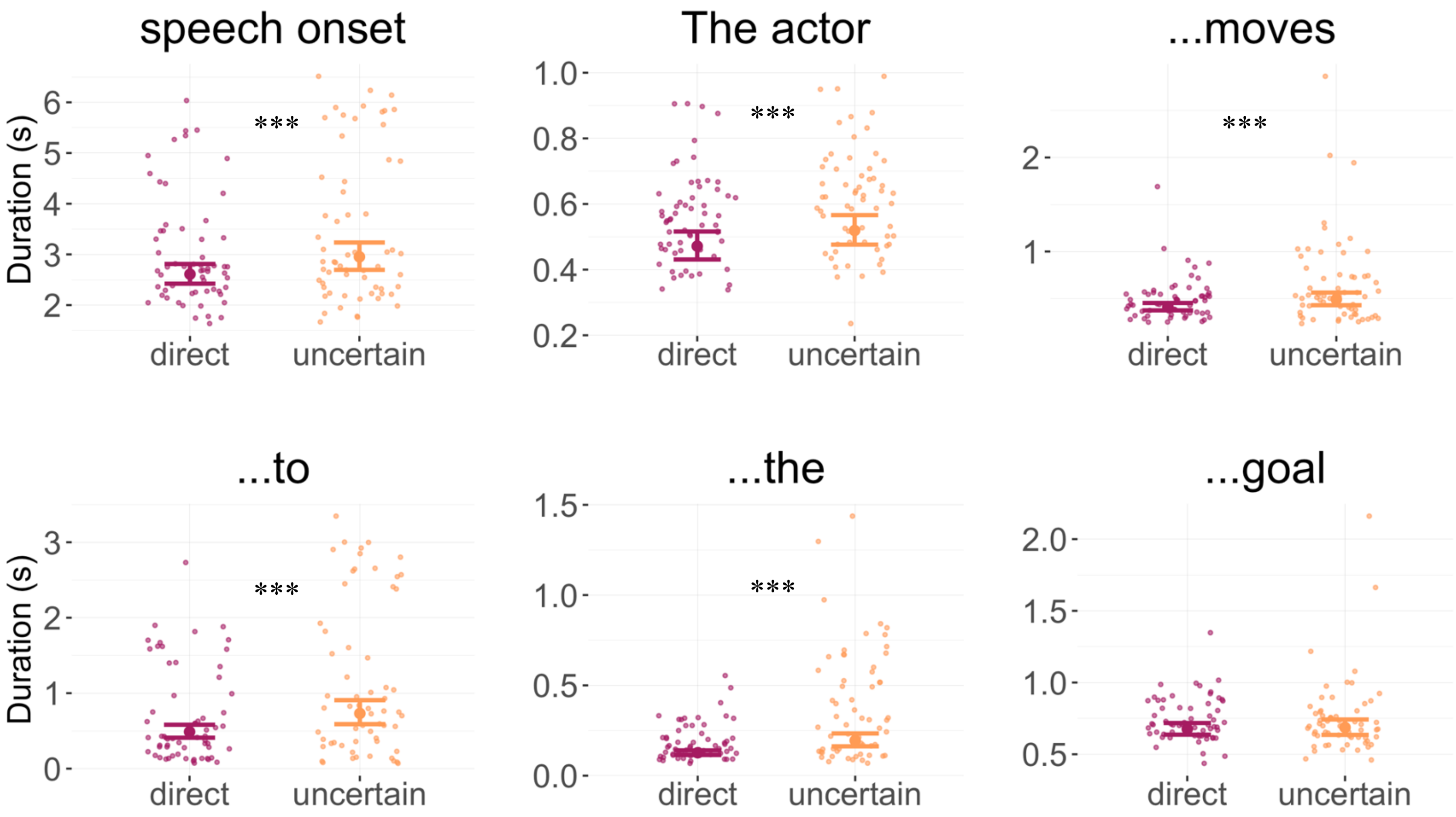

Figure 3.5. Model predictions for the duration of each chunk in the Direct (purple, left bars) and Uncertain (orange, right bars) conditions. For graphing purposes, predictions were held at the average level of Source, which was controlled for in the statistical models. Points represent raw individual participant means. Note the different scales on the y-axis. *Speech onset* = Latency, *the actor* = Actor NP, *...moves* = Verb phrase, *...to* = PP, *…the* = Article , *...goal* = Goal Name. ****p* < .0001, ***p* < .01, **p* < .05.

<u>*Disfluencies*</u>

Fillers

To examine whether the rate of fillers was affected by condition, we ran a generalized linear mixed effects model regressing Filler [0,1] on Condition, Language, and their interaction; controlling for Source. The random effects structure included by-subject and by-item intercepts and slopes for Condition. Results showed no significant effect of the predictors on Filler production *p*s > .1, suggesting participants' production of fillers was not influenced by our manipulation. There was a low rate of fillers overall (4%) and we did not run any additional analyses on them.

Pauses

For each utterance chunk, we extracted the duration of the silent pause preceding it. Pauses longer than 100ms were considered meaningful pause (see Hieke et al., 1983; Swets et al., 2021) and assigned a value of 1 (pause), all others were assigned 0 (no pause). We then ran a generalized linear mixed effects model regressing Pause [0,1] on Condition, chunk Position (Actor, Verb, PP, Article, Goal object), and their interaction. This omnibus model showed a significant interaction between Condition and Position χ^2 (4) = 49. 05), $p < .0001$; see Figure 3.6.

Follow-up comparisons showed that participants produced more pauses in the Uncertain condition compared to the Direct condition in positions before the Verb (b = -0.42, SE = 0.19, $z = -2.32$, $p < .05$), the PP (b = -1.08, SE = 0.17, $z = -6.49$, $p < .0001$), the Article (b = -1.66, SE = 0.21, $z = -7.79$, $p < .0001$), and the Goal (b = -0.54, SE = 0.2, $z = -2.57$, $p < .05$). For the Actor NP, the pause was calculated as the difference between the initial article offset and the actor onset (between "the" and "butterfly"), and there was no significant difference between conditions (b = -0.18, SE = 0.27, $z = -0.66$, $p > .1$). This finding aligns with the general preference to plan the article and noun components of the phrase together (Allum & Wheeldon, 2007; see also Chapter Two). Moreover, any hesitation was likely included in the speech initiation latency that immediately preceded the actor noun phrase. Note, however, that there was a significant effect of Condition on pauses before the goal name, even though the goal was a similar noun phrase as the actor; comprising a noun preceded by an article. This might suggest that early in the utterance, participants maintain their tendency to plan the noun and the article together. However, they are willing to pause between the NP components if needed – i.e., if they've reached the critical point (the goal) but the information is not disambiguated yet. In that case, they can produce "the" to reduce some of the long silence. If the goal information is not immediately disambiguated, this will result in a pause between the article and the goal noun.

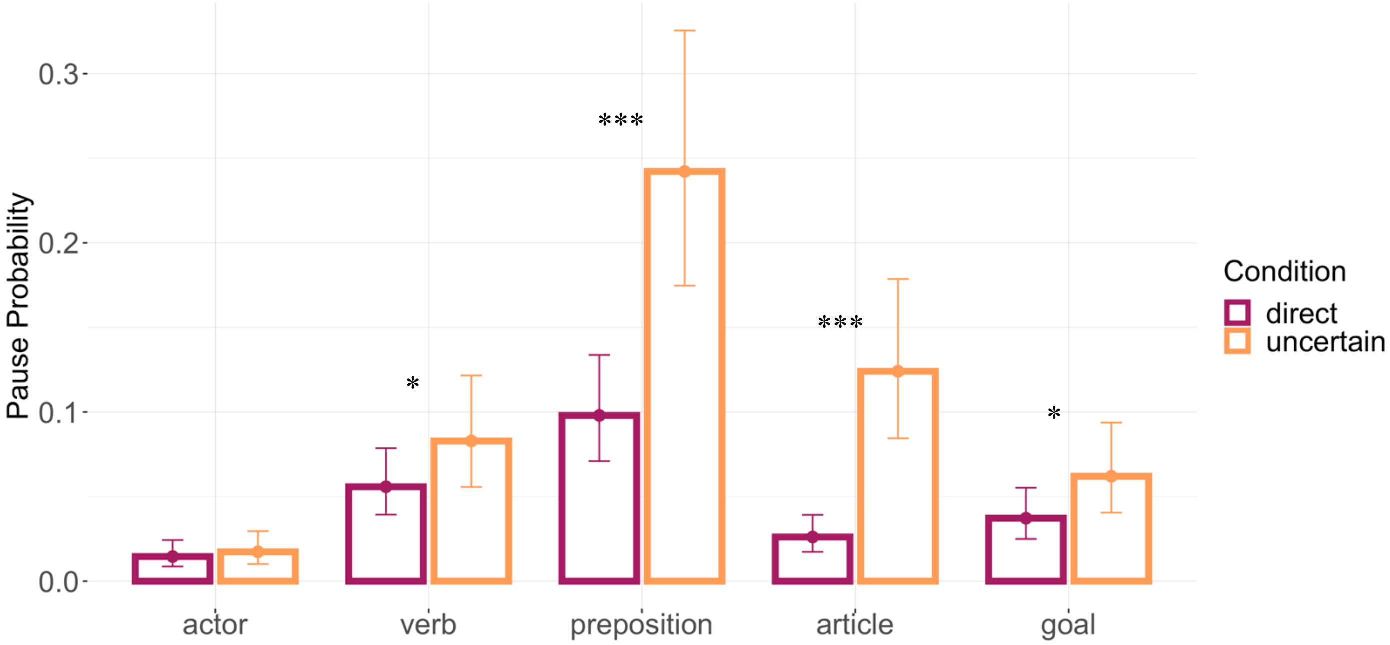

Figure 3.6. Model predictions for the probability of pauses at each chunk position, separated by the Direct (purple, left bars) and Uncertain (orange, right bars) conditions. Predictions were held at the average level of Source. Error bars represent 95% CI. ****p* < .0001, ***p* < .01, **p* < .05.

Tradeoff between speech initiation latency and utterance duration

We next examined the expected tradeoff between speech initiation time and total utterance duration. We ran a linear mixed effects model regressing Total Duration (logged) on Latency (logged), Condition, and their interaction; while controlling for Source. We also included by-subject and by-item random intercepts, and by-subject and by-item random slopes for condition.

Results show that shorter initiation latencies predicted longer utterance durations (b = -0.48, SE = 0.02, $t = -31.09$, $p < .0001$), suggesting that when participants began speaking sooner, their utterances were longer – they needed to plan more incrementally during production. This effect was qualified by a significant interaction between Condition and Latency (b = -0.43, SE = 0.03, $t = -15.81$, $p < .0001$), and follow-up comparisons suggest that the association between Latency and Total Duration was stronger in the Uncertain condition (linear trend = -0.22, SE = 0.007, 95% CI [-0.23, -0.21]) compared to the Direct condition (linear trend = -0.082, SE = 0.007, 95% CI [-0.095, -0.07]). Figure 3.7 displays this relationship.

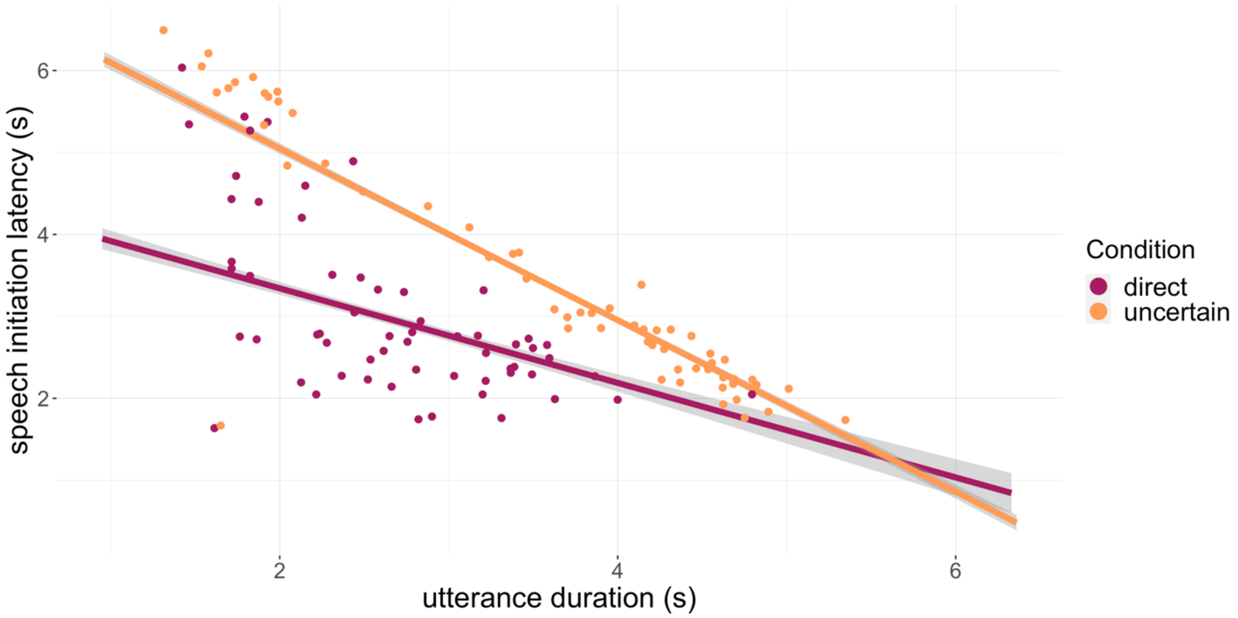

Figure 3.7. Raw scatter plot and fitted line of the average correlation between speech initiation latency and total utterance duration, for the Direct (purple) and Uncertain (orange) condition. Points represent individual participant means. Error bands represent 95% CI.

Experiment 1 Discussion

In a motion event-description task, we manipulated whether the goal of the movement was known from the start of the trial (Direct) or only disambiguated once the movement was nearly complete (Uncertain). We found that participants were slower to begin speaking and produced utterances of longer durations in the Uncertain condition. Longer durations were found for each of the utterance components that we examined up until the goal component itself: actor noun phrase, verb, prepositional phrase, and article before the goal. Moreover, there was a tradeoff between speech initiation latency and total utterance duration: the sooner participants began speaking, the longer their utterance durations were, suggesting they were planning upcoming parts of their utterance while producing the earlier parts. Finally, silent pauses were more likely in the Uncertain condition compared to the Direct condition, but we did not find a difference in the rate of fillers between conditions.

The results first support theories of moderate (non-radical), or strategic, incrementality in production (Ferreira & Swets, 2002; Swets et al., 2021). Differences in durations between conditions were found already at speech initiation time, suggesting that speakers did assess the entire message before beginning to speak, and delayed their response in the Uncertain condition when a component of the message was still undetermined – even though that component was only needed later in the utterance. On the other hand, participants did begin producing their utterances in the Uncertain condition before the disambiguation point, indicating that they relied on incremental planning and were willing to risk getting stuck mid-utterance (or lengthening their utterance duration) if the goal was not disambiguated in time. Although later speech initiation times typically indicate more advance planning, here we find a combination of the two – speakers were later to begin speaking in the Uncertain condition, but produced longer utterances.

Why were utterance durations longer in the Uncertain condition? One possibility is the increase in cognitive demand. Perhaps participants were slower because they were paying careful attention as the event unfolded to reveal the goal in the Uncertain condition (a visual attention demand), or because they were simultaneously planning their goal response while producing the earlier portions of the utterance (a linguistic demand). That is, under this account the lengthening of words is a consequence of the increased cognitive demand of incremental planning under message uncertainty – the need to attend to unfolding events, to plan while speaking, and/or a combination of those. Alternatively, perhaps participants lengthened their word durations in order to minimize long periods of silence or fillers, while buying time for the goal to be revealed.

This can be seen as a more strategic approach; participants begin speaking before the goal is disambiguated to ensure they have enough time to produce the utterance, but then lengthen word durations to conform to turn-taking and prosodic norms – perhaps even signaling to their experiment partner that they are waiting for the message information to be completed.

These options are not mutually exclusive, however, and it is also possible that the cause of lengthening changes across word positions as the critical point (the goal) approaches. Note also that shorter speech initiation latencies led to longer utterance durations in both conditions. This suggests some degree of incrementality in the Direct condition too; the less time participants spent planning prior to speech initiation, the more they needed to plan while production was occurring, thus increasing the cognitive demand and utterance durations. Indeed, the complex multi-phrase utterances in our experiment are exactly where incremental planning has been found useful, allowing to plan smaller chunks at a time in order to decrease the cognitive effort of maintaining large utterance portions (Brown-Schmidt & Konopka, 2015) – a factor that is shared across the Direct and Uncertain conditions. The fact that the association between speech initiation latency and utterance duration was stronger in the Uncertain condition compared to the Direct condition might suggest that the difference in strength between conditions could instead be attributed to a strategic difference in planning under message uncertainty, rather than a consequence of incremental planning more generally.

Contrary to our hypothesis, we did not find a difference between conditions in the rate of filler production, which was rather low overall (4%). It is possible that because we used a uniform message structure (motion event; *the x moved to the y*) across trials, there was no "key plot point" in which a different message was being planned, as in previous studies that found an association between message uncertainty and filler production (Fraundorf & Watson, 2013). That is, in our design it was only a component of the message that was missing, and always of the same thematic role and in the same serial position. As a result, perhaps participants did not need to use fillers to signal message uncertainty – the uncertainty point was predictable across trials, and pauses and lengthening were preferred instead. The fillers we did find might therefore be more indicative of picture naming difficulties, which were equally likely to occur across conditions.

On the other hand, we did find an increase in silent pauses in the Uncertain condition compared to the Direct condition, starting already before the verb phrase. This again suggests that components of the entire message were considered early on in the utterance, including uncertainty about the later goal. However, the effect almost tripled between the verb phrase and the prepositional phrase, which might suggest the preposition position was a bottleneck position. Note that the components produced up until that point ("the actor moves") are necessarily true and certain from the moment movement begins. But the goal – or even the existence of any destination goal rather than just a movement event – is only certain once the goal is disambiguated. Speakers might therefore prefer to wait for the goal before committing to producing even the prepositional phrase. This finding is particularly interesting given that our trials were uniform in their message structure and always included a goal destination. We might instead expect speakers to fluently produce the entire utterance up until the goal, with the rate of pauses before the goal showing the largest increase. Our results instead suggest a preference to plan the prepositional phrase and the goal noun phrase together, perhaps reflecting speakers' long-term experience of planning these components together.

Figure 3.4 also shows that the goal disambiguation point aligned, on average, with production of the verb. Thus an alternative explanation to the bottleneck of pauses at the prepositional phrase is that this was simply a consequence of the goal being disambiguated at precisely that point in time, where the preposition was the next component to be produced, and was delayed as participants retrieved the goal information. This explanation appears circular, however, given that participants could have begun producing their utterances much earlier (at the very least, matching the earlier initiation time in the Direct condition), in which case the pauses would likely occur sooner. Instead, it seems more likely that participants had particular sentence positions and phrase boundaries where they were comfortable pausing or lengthening components, and their utterance production was timed in accordance – starting already from speech initiation time and throughout production.

To conclude, we find evidence of moderate incremental planning in the face of message uncertainty, affecting the time course of utterance production. Participants delay speech initiation when there is message uncertainty, suggesting there is some consideration of the entire message prior to beginning to speak. However, speakers begin describing the event even before the goal target is disambiguated, causing longer utterance durations and more silent pauses, with effects increasing nearer to the goal.

In contrast with the experiments described in Chapter Two, the current paradigm does not allow much flexibility in word order that participants could use to delay the production of the uncertain components. In fact, the canonical position of the goal at the end of the sentence structure in English is already the ideal situation for speakers, given that the goal is the uncertain component of the event that is only disambiguated at the end. But what if the uncertain component needs to be produced at an earlier position in the utterance? How does message uncertainty interact with grammatical word order during utterance planning? If the effects of message uncertainty depend on the permitted word order of a language, will speakers of languages with different word orders vary in their strategies for production under uncertainty? In Experiment 2 we attempt to answer these questions using a cross-linguistic version of the event-description paradigm.

Experiment 2A: cross-linguistic comparison of word order

In Experiment 2A we use a cross-linguistic approach to investigate how the permitted word orders of English and Spanish affect planning under message uncertainty. The paradigm uses a similar motion event-description task as in Experiment 1, except that the two goal locations are the same object in different colors. Adequate descriptions of the event must therefore specify the color of the target goal, and not only its name. Figure 3.8 shows an example of a display, where the event description would be “the butterfly moves to the *red* slide”.

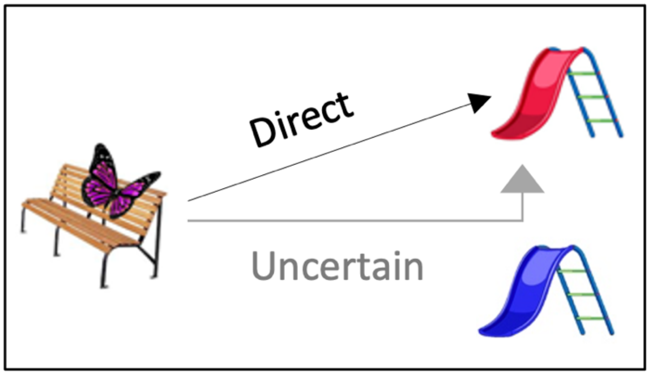

Figure 3.8. Example of a trial display in Experiments 2A. Arrows and labels are for illustration only, demonstrating the trajectory for the Direct condition (black) versus the Uncertain condition (gray).

In this design, although there is initial uncertainty about the goal identity in the Uncertain condition, the noun component (*slide*) of the goal noun phrase “red slide” can be planned in advance because it is shared by both goal options. From a linearization perspective, however, grammar rules of English require that the color – the *uncertain* component – be spoken first, before the noun. This presents an interesting case for planning under message uncertainty, because speakers could plan the noun in advance, but they cannot produce it before the color information is disambiguated. Moreover, if speakers want to plan what they can in advance, this would require separate planning of the noun from the adjective, even though the “red slide” is likely a unified concept and the adjective is part of the noun phrase. In sum, the asymmetry between what can be *planned* first versus what can be *produced* first creates an interesting context for production under uncertainty, and it is unclear whether early planning would be beneficial.

In contrast, Spanish word order typically places the noun before the adjective (e.g., *tobogán rojo, ‘slide-red’*). When producing Spanish, speakers in the Uncertain condition can produce the certain component (*slide*) first, allowing more time to plan the uncertain component (*red*). Thus differences in speech patterns between the Uncertain and Direct conditions should be relatively small in Spanish compared to English, or at the very least emerge later in the utterance. This would suggest that because rules of linearization differ cross-linguistically, different languages have different affordances for planning in situations of message uncertainty.

Method

Experiment 2A was pre-registered prior to examination of the data.

Participants

30 native English speakers and 30 native Spanish speakers participated in exchange for course credit or 10 USD. Data from an additional 11 participants were excluded from analyses due to corrupt audio files (n = 7) or server error (n = 4). During data transcription we discovered one additional participant who never used a verb, and three participants who used excessive descriptions across all utterances (e.g., *the butterfly moves to the left in between the tree and the slide and then down to the slide*). Data from these participants were removed from any further processing.

Participants' native language was self-reported, and defined as the language that they spoke at home before the age of five. Participants were recruited via an introductory Psychology course in exchange for course credit. Given the limited number of native Spanish speakers in the participant pool, we also recruited participants by advertising through email lists, social media, and flyers around campus. These participants were compensated with 10 USD. Native Spanish speaking participants were all students or employees at English-speaking universities in the USA, who reported high-to-native English proficiency.

Stimuli

A total of 40 display items were created, each consisting of an actor, a source, and two goals. Stimuli creation was similar to Experiment 1, except that the two goal locations were the same object image in different colors (e.g., a blue slide and a red slide). One of these images was in its original color as created in the MultiPic database, and the other image was a modified version that we created by changing only the color of the original image using free online software. That is, the two images used as goals were identical except for their color. Participants were randomly assigned to one of two lists, each consisting of 20 trials in the Direct condition and 20 trials in the Uncertain condition. The goal color was randomly pre-determined and kept constant across lists; the only manipulation was whether it appeared in the Direct or in the Uncertain condition.

Procedure

During the experimental session, trials of Experiment 2A were interleaved with trials of Experiment 2B (see next section), which did not include identical goals of different colors. Participants were therefore instructed that if a trial contained two of the same object in different colors, they must specify the color so that their experiment partner (the RA) can correctly identify the goal. However, if the trial did not contain two of the same object in different colors, they should not mention the color of the object. Participants performed one practice trial of each type (Experiment 2A and Experiment 2B) before beginning the experimental phase. The entire experiment, interleaving trials from Experiment 2A and Experiment 2B, lasted approximately 35 minutes.

Trial procedure was identical to Experiment 1, except that each trial lasted 10 seconds: 1 second of preview, 5 seconds of movement, and 4 seconds that the actor remained at the target goal (in Experiment 1 the actor remained at the target for only 2 seconds). The experimental phase consisted of 20 trials in the Direct condition and 20 trials in the Uncertain condition. The condition in which each item appeared was counterbalanced across two lists.

The stimuli list and procedure were identical in the Spanish and the English version of the experiment. Each version of the experiment was carried out entirely in its target language; including instructions, consent forms, familiarization phase, and interactions with our research assistants. Research assistants were native speakers of the language in which they were facilitating the experimental session.

Analyses and Results

Preprocessing

Preprocessing was identical to Experiment 1, except as noted here. Utterances were divided into six main chunks: 1) speech initiation latency, 2) Actor noun phrase, 3) Verb phrase , 4) Prepositional phrase + article before the goal, 5) Goal noun, and 6) goal Color. Note that in contrast with Experiment 1, here we combined the prepositional phrase with the article before the goal noun. This is because in Spanish, when the masculine definite article *el* (the_{masc}) is preceded by the preposition *a* (to), the two are combined into the contraction *al* (to the). To keep our analysis chunks uniform across grammatical genders and languages, we simply chunked the preposition together with the subsequent article for all cases.

Before performing statistical analyses, we removed trials with missing components, as displayed in Table 1. Additionally, we removed utterances that failed forced alignment (5%). We also removed utterances that included the color post-nominally in English (n = 5) or pre-nominally in Spanish (n = 1), utterances where participants provided excessive description (n = 2), and utterances where transcribers noted that the trial was cut off as the participant was producing the last word (n = 17). There were too few observations in each of these categories to allow for meaningful statistical analyses between conditions.

Table 1. *Proportion of excluded trials in Experiment 2A due to missing components, displayed by Language and Condition.*

	English		Spanish	
	Direct	Uncertain	Direct	Uncertain
Actor	.01 (.09)	.01 (.09)	.07 (.25)	.07 (.26)
Verb	.01 (.08)	.01 (.1)	.06 (.24)	.06 (.23)
Preposition + Article	.01 (.11)	.01 (.12)	.08 (.27)	.08 (.28)
Goal Name	.04 (.19)	.06 (.24)	.15 (.36)	.16 (.37)
Goal Color	.06 (.23)	.06 (.23)	.2 (.4)	.21 (.41)

Note. Values represent mean proportion of overall observations prior to data cleaning. Standard deviations appear in parentheses.

Utterances with fillers were included in analyses, in similarity to lengthened productions and pauses. All other, uncategorized, disfluencies (5%) were removed from our main analyses. After these steps of data cleaning, four participants (three Spanish speakers, one English speaker) were left with less than 25% of trials in at least one condition and were removed from further analyses. Of the remaining 56 participants (27 Spanish speakers; 29 English speakers), the average percent of retained trials was 78% (SD = 17%, range 25%-100%); comparable across conditions.

Durations Analysis

We first examined the effect of condition on the duration of each chunk in participants' utterances. Figure 3.9 displays the averaged chunk durations in each condition, on a timeline of the trial procedure, further distinguishing the spaces between chunks. Note again that for analysis purposes, any spaces between components were included with the later component.

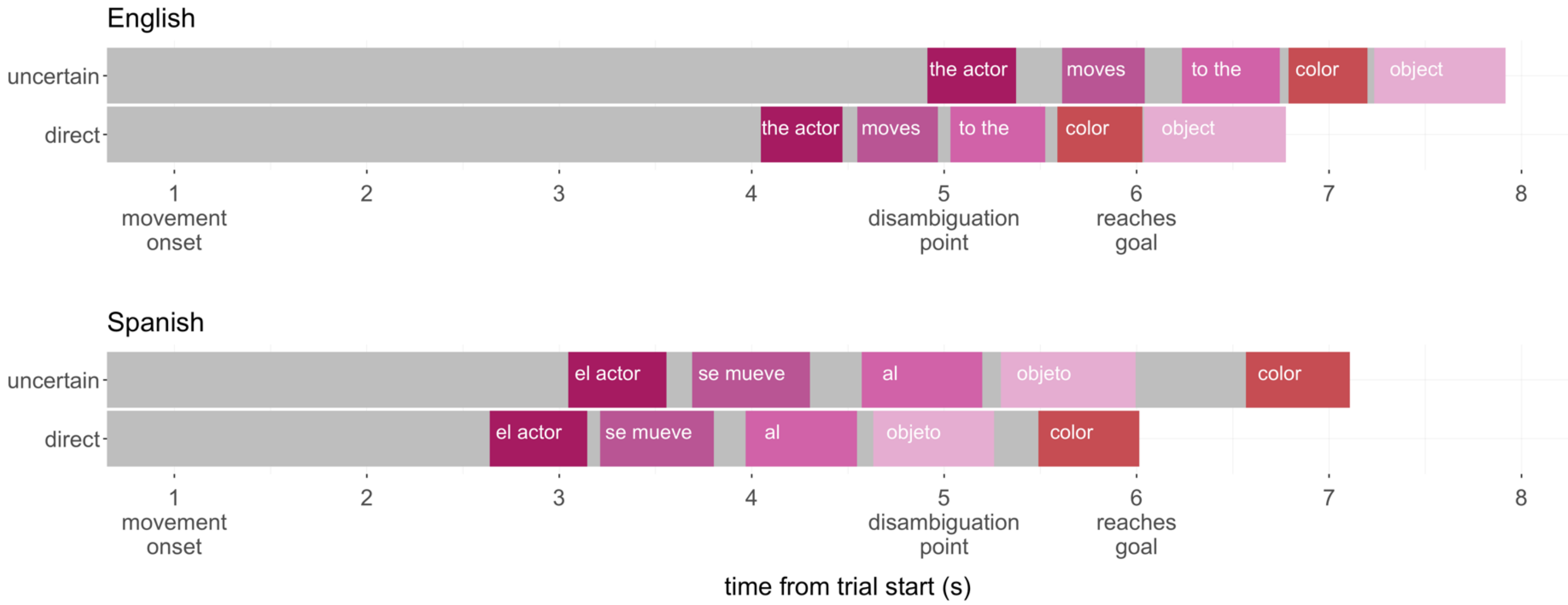

Figure 3.9. Average chunk durations illustrated on the trial timeline, as a function of Language (Spanish, English) and Condition (Uncertain, Direct). Each component is displayed in a different color with overlayed text to identify the component; note the different word order in English versus Spanish. Gray bars indicate initiation latency (time from trial start until Actor NP onset) and spaces between components. For simplicity, only trials where no source was mentioned are displayed here. Trials began at 0 seconds and ended at 10 seconds but these time points are not displayed.

We ran a linear mixed effects model separately for each chunk, regressing Duration on Condition (Direct vs Uncertain; center-coded [-.5, .5]), Language (English vs Spanish; center-coded [-.5, .5]), and their interaction. We further controlled for participants' mention of source, which would naturally lengthen their utterances, by adding a predictor for Source (No-source vs Source; center-coded [-.5, .5]). The random effects structure included by-subject and by-item random intercepts, by-subject random slopes for Condition and Language, and by-item random slopes for Condition, Language, and their interaction. To allow for convergence, we simplified by-item random effects (Barr et al., 2013) of the following models to include only the interaction between Condition and Language: Actor NP, Verb, PP. Significance values are reported using Satterthwaite's method t-tests produced from the *LmerTest* package. Figure 3.10 displays the difference between conditions at each chunk position.

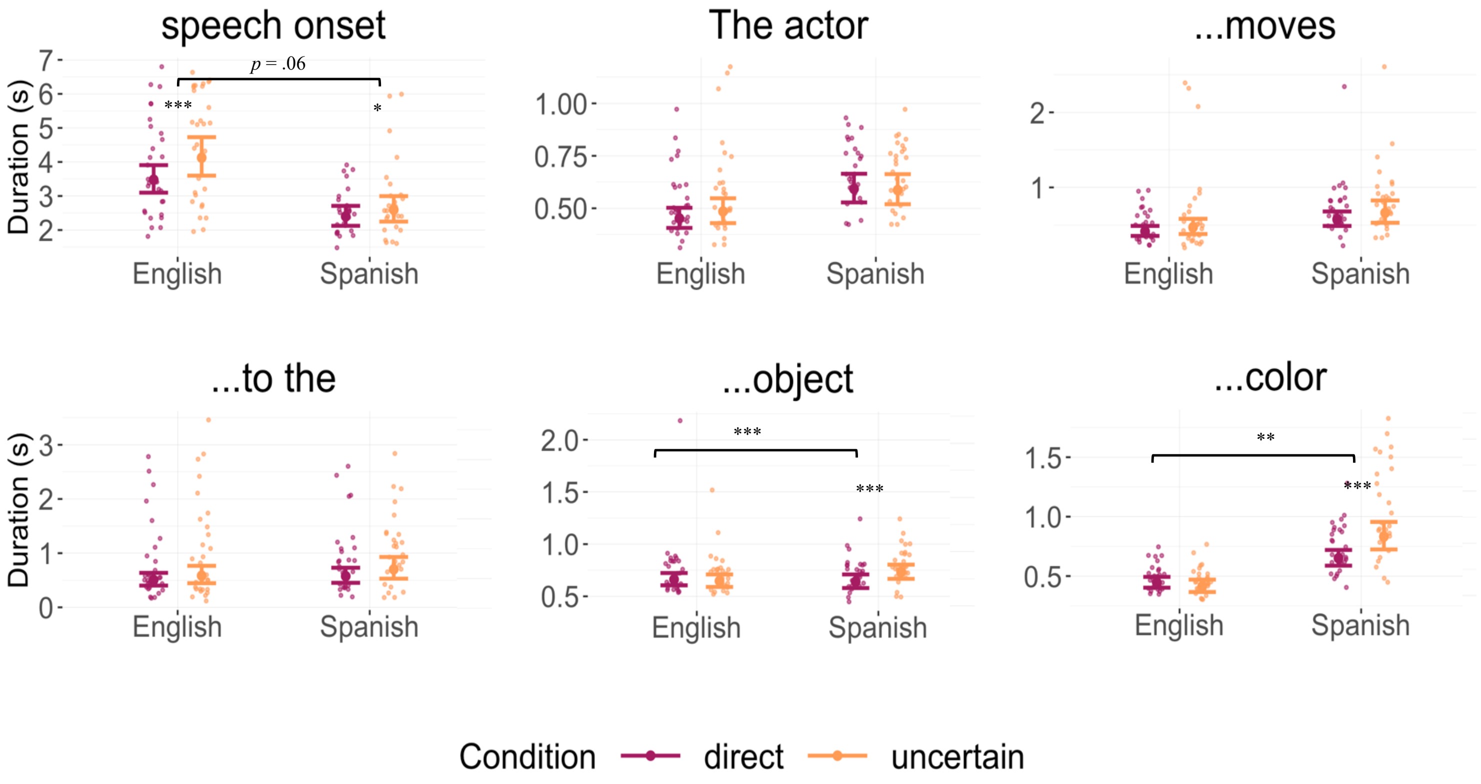

Figure 3.10. Model predictions for the duration (in seconds) of each chunk as a function of Language (English, Spanish) and Condition (Direct, Uncertain) in Experiment 2A. Predictions were held at the average level of Source, which was controlled for in the statistical models. Points represent raw individual participant means. Note the different scales on the y-axes, and recall that in English the color word was produced before the object word. *** $p < .0001$, ** $p < .01$, *$p < .05$.

Speech initiation latency

Results show a significant effect of Condition on speech initiation latency, $b = 0.12$, SE = 0.03, $t = 4.89$, $p < .0001$; speakers had longer durations in the Uncertain condition compared to the Direct condition. There was also a significant effect of Language ($b = -0.42$, SE = 0.09, $t = -4.76$, $p < .0001$), suggesting that speakers began speaking sooner in Spanish compared to English. Moreover, there was a marginally significant interaction between Condition and Language, ($b = -0.09$, SE = 0.05, $t = -1.88$, $p = .06$), driven by a larger effect of Condition in English ($b = -0.17$, SE = 0.03, $t = -5.03$, $p < .0001$) compared to Spanish ($b = -0.08$, SE = 0.04, $t = -2.13$, $p = .038$).

Actor noun phrase

Durations of the actor noun phrase were significantly longer in Spanish compared to English productions, $b = 0.23$, SE = 0.07, $t = 3.25$, $p < .01$. There was no significant effect of Condition and no interaction between Condition and language (ps > .1).

Verb phrase

Durations of the verb phrase were significantly longer in Uncertain compared to Direct trials; $b = 0.13$, SE = 0.04, $t = 3.4$, $p < .01$. Durations of the verb phrase were also significantly longer in Spanish compared to English productions; $b = 0.33$, SE = 0.13, $t = 2.57$, $p < .05$. There was no significant interaction between Condition and language ($p > .1$).

Preposition + article

Durations of the prepositional phrase (with the following article) were significantly longer in Uncertain compared to Direct trials; $b = 0.17$, SE = 0.05, $t = 3.23$, $p < .01$. There was no significant effect of Language and no interaction between Condition and language (*p*s > .1).

Goal color

Durations of the color phrase were significantly longer in Uncertain compared to Direct trials; $b = 0.09$, SE = 0.04, $t = 2.23$, $p < .05$. Durations of the color phrase were also significantly longer in Spanish compared to English productions; $b = 0.54$, SE = 0.07, $t = 8.04$, $p < .0001$. Importantly, there was a significant interaction between Condition and Language ($b = 0.32$, SE = 0.08, $t = 3.97$, $p < 0.001$), driven by a significant effect of Condition in Spanish ($b = -0.25$, SE = 0.06, $t = -4.06$, $p < 0.001$) but not in English ($b = 0.07$, SE = 0.05, $t = 1.38$, $p > .1$).

Goal object name

Durations of the goal object phrase were significantly longer in Uncertain compared to Direct trials; $b = 0.06$, SE = 0.02, $t = 2.65$, $p < .01$. There was no significant effect of Language on durations of the object goal phrase, $p > .1$. Importantly, there was a significant interaction between Condition and Language ($b = 0.16$, SE = 0.05, $t = 3.39$, $p < .01$), driven by a significant effect of Condition in Spanish ($b = -0.13$, SE = 0.03, $t = -3.93$, $p < 0.001$) but not in English ($b = 0.02$, SE = 0.03, $t = 0.8$, $p > .1$).

Total utterance durations

The total duration of utterances, from noun phrase onset until either goal object offset (English) or goal color offset (Spanish), was significantly longer in Uncertain compared to Direct trials; $b = 0.495$, SE = 0.101, $t = 4.74$, $p < .0001$. Durations were also significantly longer in Spanish compared to English; $b = 0.92$, SE = 0.24, $t = 3.82$, $p < .001$. There was no significant interaction between Condition and language ($p > .1$).

Disfluencies

Fillers

To examine whether the rate of fillers was affected by condition, we ran a generalized linear mixed effects model regressing Filler [0,1] on Condition, Language, and their interaction; controlling for Source. The random effects structure included by-subject and by-item intercepts, a by-subject slope for Condition, and by-item slopes for Condition, Language, and their interaction. Results showed no significant effect of the predictors on Filler production (*p*s > .1), suggesting participants' production of fillers was not influenced by our manipulation.

Pauses

We ran a generalized linear mixed effects model regressing Pause [0,1] on Condition, Language, chunk Position (Actor, Verb, PP, Goal Color, Goal object), and their interactions. This omnibus model showed a significant three-way interaction between Condition, Language, and Position, $\chi^2(4) = 29.26$, $p < .0001$. Follow-up comparisons showed that English speakers produced more pauses in the Uncertain condition compared to the Direct condition in positions before the Verb ($b = -0.47$, SE = 0.206, $z = -2.3$, $p < .05$) and the PP ($b = -0.81$, SE = 0.21, $z = -3.79$, $p < .001$), but not before other positions (*p*s < .05). In contrast, Spanish speakers produced more pauses in the Uncertain condition compared to the Direct condition only before the Goal Color ($b = -0.89$, SE = 0.16, $z = -5.66$, $p < .0001$), and marginally more before the PP ($b = -0.33$, SE = 0.17, $z = -1.9$, $p = 0.06$). That is, the Uncertain condition increased the rate of pauses for both English and Spanish speakers, but it occurred earlier in the utterance for English speakers. Figure 3.11 displays these results.

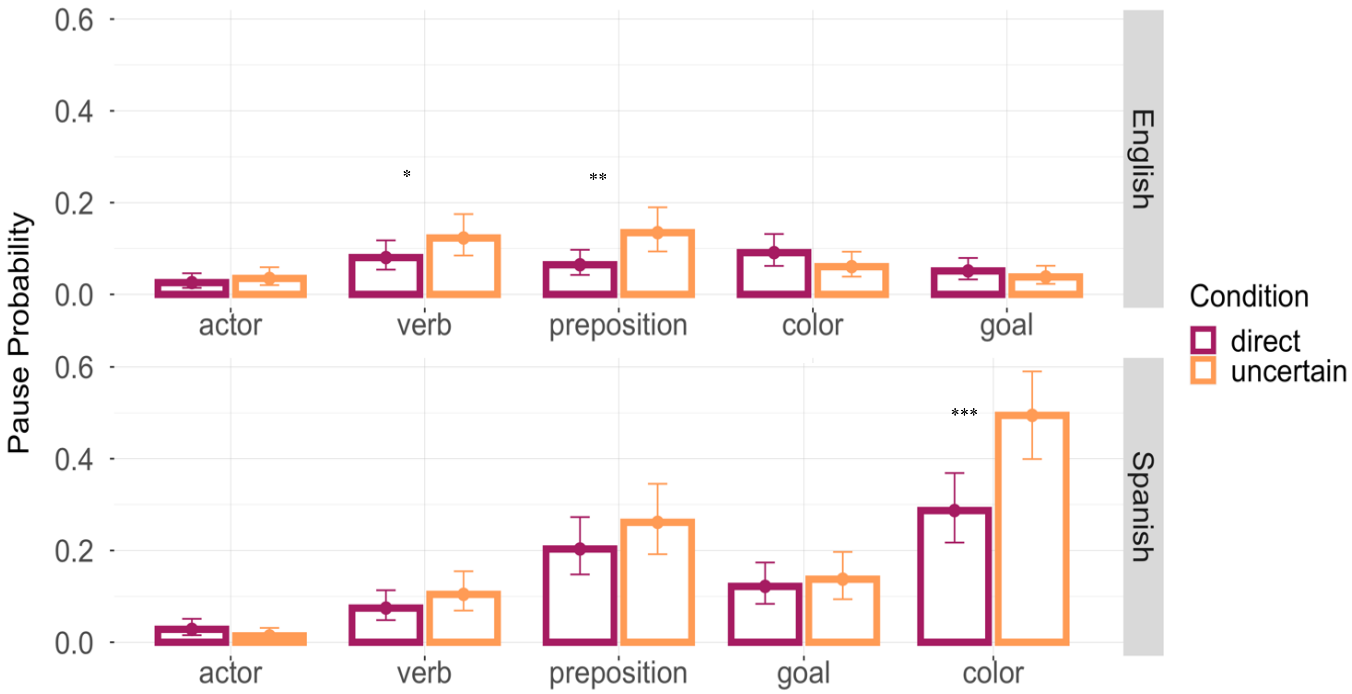

Figure 3.11. Model predictions for the probability of pauses at each chunk position in Experiment 2A, as a function of Language and Condition. Predictions were held at the average level of Source. Note the different word orders for Spanish versus English. Error bars represent 95% CI. *** $p < .0001$, ** $p < .01$, *$p < .05$.

Correlation Analysis

We next examined the expected tradeoff between speech initiation time and total utterance duration. We ran a linear mixed effects model regressing Total Duration (logged) on Latency (logged), Condition, Language, and their three-way interaction; while controlling for Source. We also included by-subject and by-item random intercepts, and random slopes for condition (by-subject), and the interaction between condition and language (by-item).

Results show that shorter initiation latencies predicted longer utterance durations ($b = -0.36$, $SE = 0.02$, $t = -18.68$, $p < .0001$), indicating that when participants began speaking sooner, their utterances were longer – suggesting incremental planning during production. Importantly, there was a significant three-way interaction between Condition, Language, and Latency ($b = 0.25$, $SE = 0.06$, $t = 4.05$, $p < .0001$). Follow-up comparisons show that the association between Latency and Total Duration was stronger in the Uncertain condition compared to the Direct condition both in English ($b = .14$, $SE = 0.01$, $t = 11.26$, $p < .0001$) and in Spanish ($b = 0.07$, $SE = 0.01$, $t = 5.15$, $p < .0001$), but the effect of condition was larger in English compared to Spanish, $b = 0.07$, $SE = 0.02$, $t = 3.8$, $p < .001$. Figure 3.12 displays this relationship.

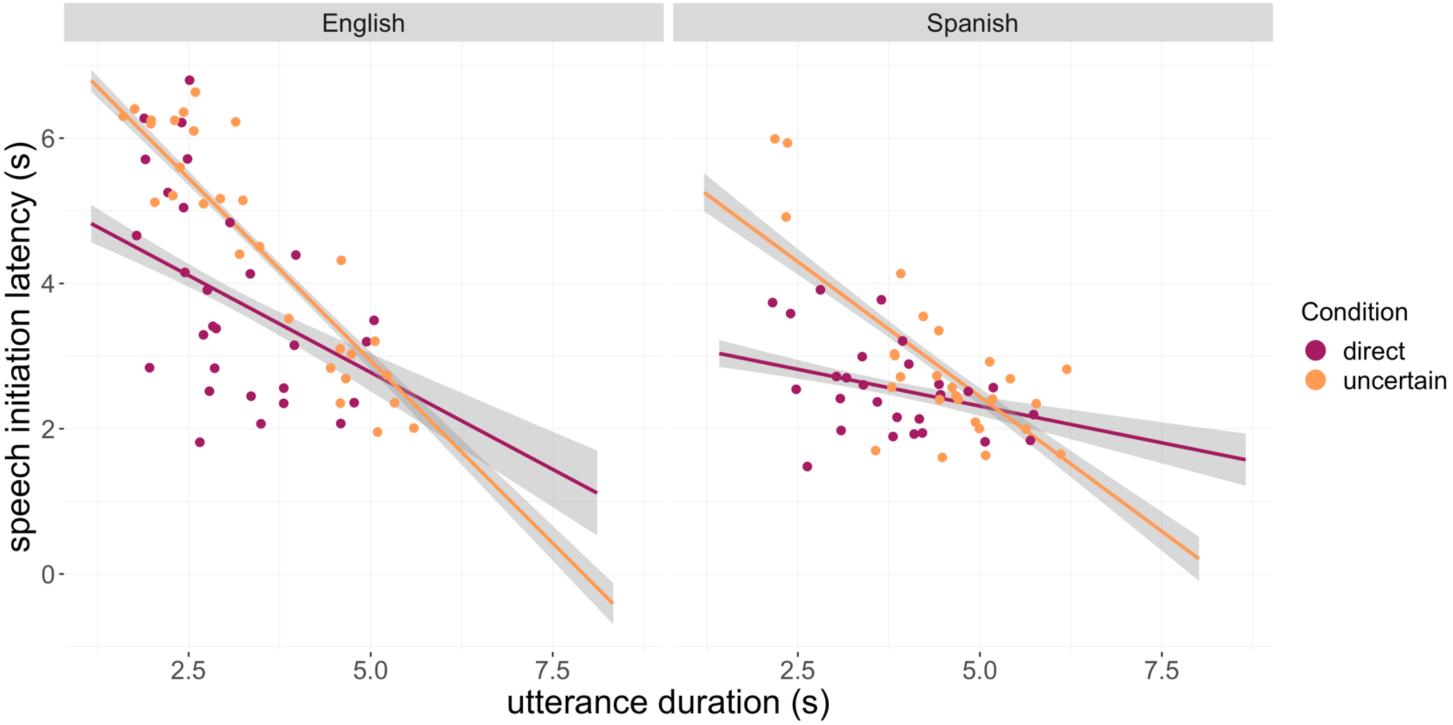

Figure 3.12. Raw scatter plots and fitted lines of the average correlation between speech initiation latency and total utterance duration, for the Direct (purple) and Uncertain (orange) conditions, in English (left panel) and Spanish (right panel). Points represent individual participant means. Error bands represent 95% CI.

Experiment 2A Discussion

The results of Experiment 2A largely replicate those of Experiment 1: we find longer speech initiation latencies, longer utterance durations, and a higher rate of pauses in the Uncertain condition compared to the Direct condition. This suggests more reliance on incremental production in Uncertain trials, as participants begin speaking before the goal information is disambiguated. However, the focus of our hypotheses was on the cross-linguistic comparison. How do effects differ between Spanish (post-nominal goal modifier) and English (pre-nominal goal modifier)? Specifically, which word positions are most affected by the uncertainty manipulation in each language?

Our results show that effects of uncertainty tended to be larger at earlier positions for English speakers: a marginally larger effect of condition on speech initiation latencies and a higher rate of pauses before the verb and preposition. In contrast, the effect of condition for Spanish speakers was larger in later positions, with longer durations at the goal name and goal color, and more pauses before the goal color (the critical disambiguating information). Together, these findings suggest that the post-nominal modifier position allowed Spanish speakers more time to plan their utterances incrementally (less advance planning), with larger effects of condition emerging as the critical goal point approached. But English speakers had to consider the critical point at earlier positions (more advance planning), given the pre-nominal modifier.

Interestingly, the overall duration of the utterance did not yield an interaction between Condition and Language, suggesting that differences that emerged throughout the utterance tended to even out across the entire utterance duration. This aligns with the correlation displayed in Figure 3.12 – if participants produced shorter utterance durations for later speech initiation times, this could even out the overall utterance duration. Note also that the effect of condition on this correlation was stronger in English compared to Spanish, perhaps because English speakers needed to be more strategic about their utterance durations, making up for their later speech initiation time by speeding up their utterance durations before the trial times out.

Notably, Spanish speakers tended to begin speaking earlier than English speakers across conditions. This effect was unexpected for the Direct condition, where speakers of both languages have the entire message information from the start of the trial. In a recent study on incremental sentence production with contrastive modified noun subjects (e.g., *the three-legged cat moves above the train*), Swets et al. (2021) similarly found that speakers of French begin production sooner than speakers of English and German. The authors attributed this effect to post-nominal modifiers in French compared to pre-nominal modifiers in English and German, allowing French speakers to rely more on incremental production in this context. Moreover, speed of processing (as measured by a letter comparison task) predicted the scope of planning for French speakers but not for English and German speakers. Swets et al. speculate that the syntactic choices available for a particular language influence the degree of incremental planning for a given context, and these varying production strategies might rely on different cognitive support systems (e.g., speed of processing in French but not in English and German).

Our results in Spanish align with the Swets et al. (2021) finding, showing earlier speech initiation for post-nominal modification. However, there are a number of alternative causes to consider for this effect. First, it is difficult to rule out the possibility that these main effects of speech initiation time are sample-dependent, given our separate participant samples for English and Spanish speakers (although Swets et al. also used separate samples). Perhaps future studies could explore this using a within-subjects design with native Spanish-English bilinguals completing the same task in each language. Second, recall that Experiment 2A trials were interleaved with Experiment 2B trials. Experiment 2B trials were all of the Uncertain type, such that approximately 76% of the trials in the experimental session were Uncertain. This could have caused participants to rely more strongly on their strategies for Uncertain trials (more advance planning or more incremental planning), carrying over to the Direct trials too – resulting in later speech initiation times in English across conditions.

In this regard, it is important to emphasize the flexibility of language planning strategies for any given context. As Swets et al. (2021) state, it is not the case that speakers of a particular language will *always* plan more or less incrementally, but strategies rather depend on the particular linguistic context and the languages' affordances (e.g., word order). That is, our current design may have provided more motivation for earlier speech initiation and incremental planning in Spanish compared to English, but this tendency could be reversed in other grammatical contexts. In Experiment 2B, we use a manipulation that reduces the benefit of incremental planning for Spanish speakers, and examine how their planning strategies are affected.

Experiment 2B: Cross-linguistic comparison using grammatical gender

In Experiment 2B we use a similar cross-linguistic approach as in Experiment 2A, except that the critical difference between languages is in grammatical gender markings. In Spanish, every noun is arbitrarily assigned a feminine or masculine grammatical gender, stored with it in the mental lexicon (Harris, 1991). Furthermore, articles, pronouns, and adjectives must agree in gender with the nouns they are associated with – e.g., *la* bicicleta [the$_{\text{fem}}$ bicycle$_{\text{fem}}$], *el* tobogán [the$_{\text{masc}}$ slide$_{\text{masc}}$]. Thus in order to produce the definite article (*el* or *la*) that precedes a noun, speakers must know the grammatical gender of the noun itself. In English, on the other hand, there is no grammatical gender system, and the same definite article is used (*the*) regardless of the noun it accompanies.

The need to specify grammatical gender on the article limits how much of the utterance a Spanish speaker can produce before knowing the noun itself. In our event-description paradigm, if the potential goals are of different grammatical genders, the speaker must know the target goal before they can plan (or produce) the associated article. Alternatively, if both goals are of the same grammatical gender, the speaker can plan the article even before the target goal is disambiguated, because the article will be identical in either case. Thus in this paradigm Spanish speakers should be more affected by the message uncertainty when goals differ by gender, as speakers are more limited in how much of the utterance they can plan in advance.

We expected to find longer word durations and a higher rate of disfluencies in the Different-gender compared to the Same-gender condition in Spanish, particularly in word positions prior to the article location. Speech patterns in the Same-gender condition in Spanish should be similar to English speakers' utterances, because in both cases the article does not depend on the uncertain noun component, and speakers could take advantage of this to plan the article even before the goal is disambiguated. Such results would suggest that the effects of message uncertainty depend on the interaction between the particular grammatical features of a language and the message context, and that a component as small as the definite article can affect speaker's utterance planning and the resulting speech patterns.

Method

The Method was similar to Experiment 2A, but the main manipulation was whether the two possible goals were of the same or different grammatical gender in Spanish. The two goal locations were different objects, and participants were not required to specify their color. Because the relevant comparison here is Same-gender versus Different-gender trials, all trials use an uncertain trajectory to the goal target – only disambiguating the target goal at the end of the trial. Trial procedure was identical to Experiment 2A, and recall that trials of Experiment 2A and

Experiment 2B were interleaved within the same experimental session and participant information remains the same.

Stimuli

A total of 46 display items were created, each consisting of an actor, a source, and two goals. Stimuli creation was similar to Experiment 1, except that the two goal locations differed in Spanish grammatical gender. Figure 3.13 shows examples of Same-gender and Different-gender trials. In Same-gender trials, the definite article in Spanish does not differ between the two targets, similar to English:

1. (a) La mariposa se mueve al$_{\text{masc}}$ tobogán$_{\text{masc}}$
   (b) La mariposa se mueve al$_{\text{masc}}$ árbol$_{\text{masc}}$
   (c) The butterfly moves to the slide
   (d) The butterfly moves to the tree

Recall that in Spanish, the contraction *al* is comprised of the preposition *a* (to) and the masculine definite article *el* (the$_{\text{masc}}$).

In Different-gender trials, the article depends on the target in Spanish, but not in English:

2. (a) La mariposa se mueve al$_{\text{masc}}$ tobogán$_{\text{masc}}$
   (b) La mariposa se mueve a la$_{\text{fem}}$ bicicleta$_{\text{fem}}$
   (c) The butterfly moves to the slide
   (d) The butterfly moves to the bicycle

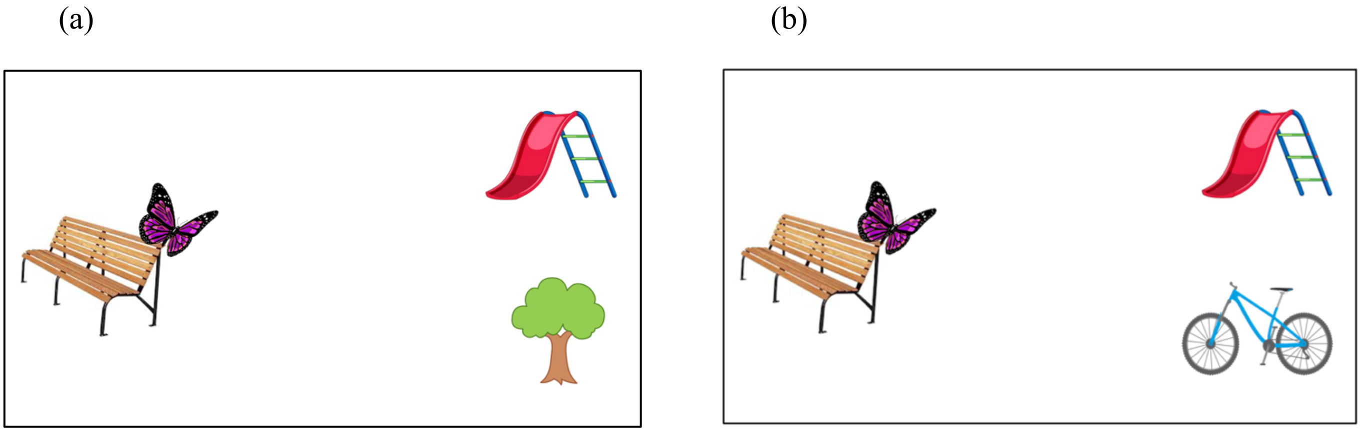

Figure 3.13. Example of trial displays in either (a) same-gender or (b) different-gender conditions of Experiment 2B.

Participants were randomly assigned to one of two lists, each consisting of 23 trials in the Same-gender condition and 23 trials in the Different-gender condition. Note that the condition of an item is determined by the grammatical genders of its goal and competitor objects, such that items are nested within condition (i.e., a given item display cannot appear in both the Same-gender and the Different-gender conditions). However, the two lists counterbalanced which of the two goal objects was the target. Although grammatical genders were meaningless for the English speakers, lists were assigned in the same way so that the same items were used across languages. For analysis purposes, however, all stimuli in English were considered of the "Same-gender" condition given that the article (*the*) is equivalent across goal nouns.

Analyses and Results

Preprocessing

Preprocessing was identical to Experiment 2A, except as noted here. Utterances were divided into five main components: 1) speech initiation latency, 2) Actor NP, 3) Verb phrase, 4) PP + article before the goal, 5) Goal noun. Before performing statistical analyses, we removed trials with missing components, as displayed in Table 2. Additionally, we removed utterances that failed forced alignment, typically due to excessive noise (5%). We also removed utterances where participants provided excessive description (n = 7), and utterances where transcribers noted that the trial was cut off as the participant was producing the last word (n = 24). There were too few observations in each of these categories to allow for meaningful statistical analyses between conditions.

Table 2. *Proportion of excluded trials in Experiment 2B due to missing components, displayed by Language and Condition.*

	English	Spanish	
	Same	Different	Same
Actor	.01 (.08)	.07 (.26)	.05 (.22)
Verb	.01 (.08)	.05 (.23)	.05 (.21)
Preposition + Article	.01 (.11)	.08 (.27)	.07 (.26)
Goal Name	.02 (.19)	.15 (.35)	.14 (.35)

Throughout data transcription it became evident that participants sometimes used synonyms of the goal object, which could be of a different grammatical gender than the label we used to design the trials. Although we provided participants with labels in the familiarization phase, it is not surprising that they occasionally used other labels during the trials, given differences in dialects or random variation in synonym use. To account for this, instead of using our pre-designed conditions, an experimental script coded for each goal produced whether it was masculine or feminine (based on the article preceding it). We then compared this gender to the gender of the competitor. If they matched, we consider this a Same-gender trial; if they do not match, we consider it a different-Gender trial (even after correction, each occurred approximately 50% of the time). Note that we could not know for certain which label participants had in mind for the competitor goal, but rather relied on previous norming (Duñabeitia et al., 2017).

Utterances with fillers were included in analyses, in similarity to lengthened productions and pauses. All other, uncategorized, disfluencies (5%) were removed from our main analyses. After these steps of data cleaning, one participant (Spanish speaker) was left with less than 25% of trials in at least one condition and was removed from further analyses, leaving a sample of 29 Spanish speakers and 30 English speakers.

Durations Analysis

Figure 3.14 displays the averaged chunk durations in each condition, on a timeline of the trial procedure. Analyses were similar to Experiment 2A except as follows. First, there was no Color chunk to analyze, so only five chunks were analyzed. Second, Condition was a nested predictor, only relevant within the Spanish stimuli. For English stimuli, all stimuli were considered of the "Same-gender" condition, given that the article (*the*) is equivalent across goal nouns.

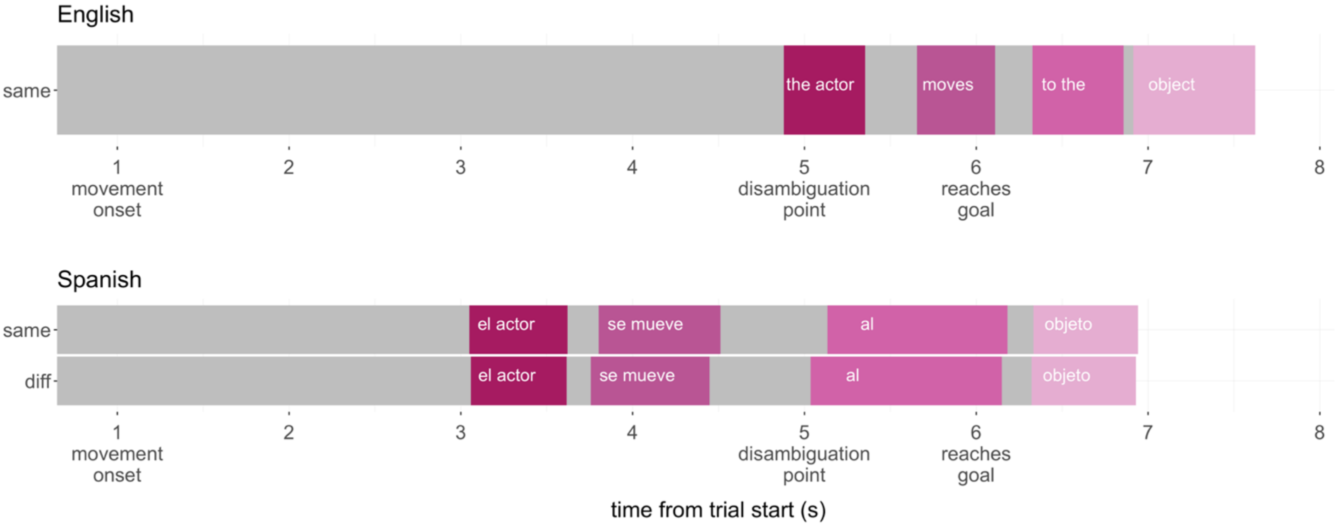

Figure 3.14. Average chunk durations illustrated on the trial timeline, as a function of Language (English, Spanish) and Condition (Same, Different). Each component is displayed in a different color with overlayed text to identify the component. Gray bars indicate initiation latency (time from trial start until Actor NP onset) and spaces between components. For simplicity, only trials where no source was mentioned are displayed here. Note that trials began at 0 seconds and ended at 10 seconds but these time points are not displayed.

We ran linear mixed-effects models regressing Duration on Condition (Different vs Same; center-coded [-.5, .5]) nested within Language (English vs Spanish; center-coded [-.5, .5]). This is equivalent to including fixed effects of Language and the interaction between

Language and Condition. To allow for convergence, we simplified by-item random effects (Barr et al., 2013) of the following models to include only the interaction between Condition and Language: Verb, PP, Goal. Figure 3.15 displays the difference between conditions at each chunk position.

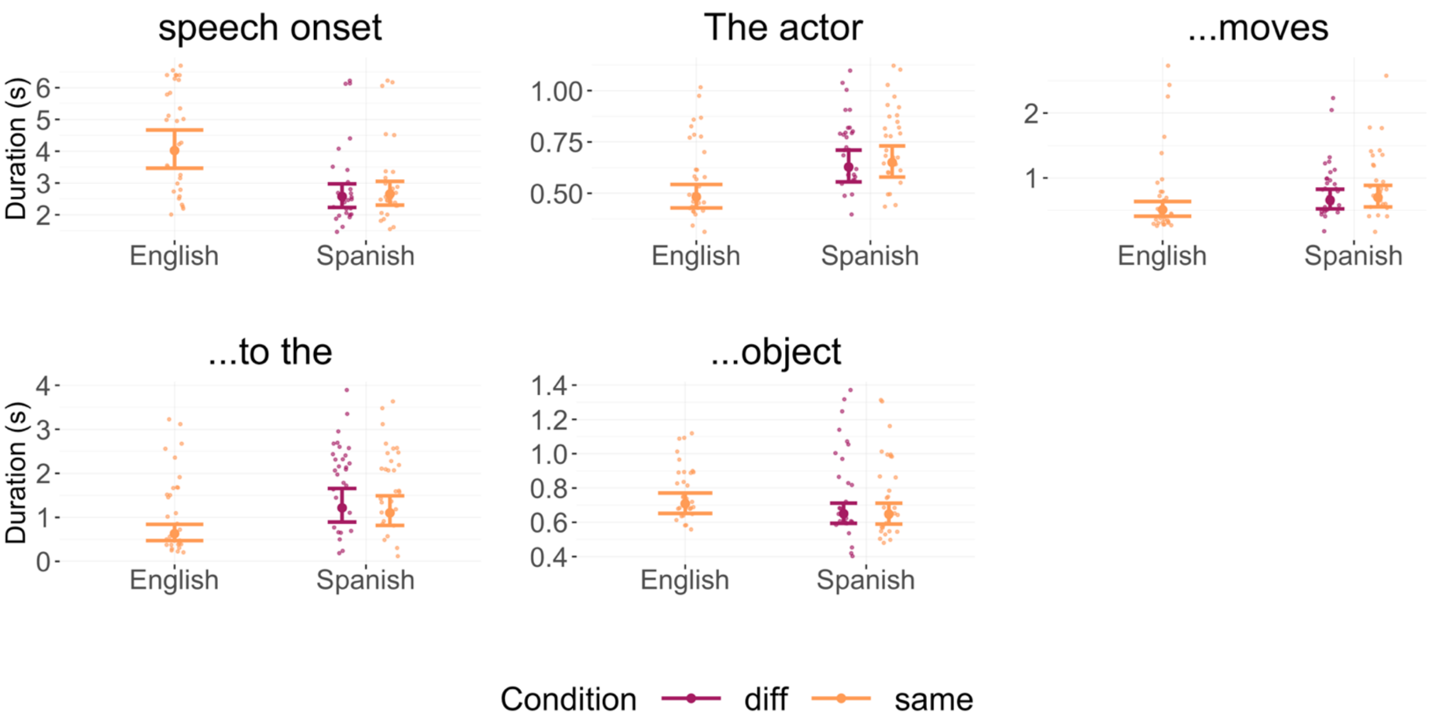

Figure 3.15. Model predictions for the duration of each chunk as a function of Language (English, Spanish) and Condition (Same, Different) in Experiment 2B. Predictions were held at the average level of Source, which was controlled for in the statistical models. Points represent raw individual participant means. Note the different scales on the y-axes.

Speech initiation latency

Speakers were faster to begin speaking in Spanish compared to English ($b = -0.45$, SE = 0.1, $t = -4.43$, $p < .0001$). There was no significant interaction between Condition and Language, $p > 0.1$.

Actor noun phrase

Durations of the actor noun phrase were significantly longer in Spanish compared to English productions, $b = 0.26$ SE = 0.08, $t = 3.29$, $p < .01$. There was no significant interaction between Condition and Language, $p > 0.1$.

Verb phrase

There were no significant effects of Language or the interaction between Language and Condition (*p*s > .1) on durations.

Prepositional phrase

Durations of the prepositional phrase were significantly longer in Spanish compared to English trials; $b = 0.66$, SE = 0.21 $t = 3.19$, $p < .01$. There was also a significant interaction between Condition and Language ($b = -0.2$, SE = 0.1, $t = -2.01$, $p = .04$). This was driven by a marginally significant effect of longer durations in the Different condition compared to the Same condition in Spanish ($b = 0.09$, SE = 0.05, $t = 1.96$, $p = 0.06$).

Goal object name

There were no significant effects of Language or the interaction between Language and Condition (*p*s > .1) on durations.

Total utterance durations

The total duration of utterances, from noun phrase onset until goal object offset, was significantly longer in Spanish compared to English; $b = 1.03$, SE = 0.3, $t = 3.42$, $p < .01$. There was no significant interaction between Condition and language ($p > .1$), suggesting no effect of gender Condition on the duration of Spanish utterances.

Disfluencies

Fillers

To examine whether the rate of fillers was affected by condition, we ran a generalized linear mixed effects model regressing Filler [0,1] on Language and Condition nested within Language; controlling for Source. The random effects structure included by-subject and by-item intercepts, a by-subject slope for Condition, and by-item slopes for Condition, Language, and their interaction. Results showed a small effect of Language on fillers, with more fillers produced in Spanish compared to English ($b = 1.15$, SE = 0.56 , $z = 2.04$, $p = .042$). No other predictors had a significant effect on filler production (ps > .1), suggesting participants' production of fillers was not influenced by our manipulation.

Pauses

We ran a generalized linear mixed effects model regressing Pause [0,1] on Condition nested within Language, chunk Position (Actor, Verb, PP, Goal object), and their interactions. This omnibus model showed a significant interaction between Language and Position (χ^2 (3) = 111.44, $p < .0001$, driven by more pauses for Spanish speakers compared to English speakers at the PP position ($b = -1.59$, SE = 0.32, $z = -5.04$, $p < .0001$) and the Goal position ($b = -0.95$, SE = 0.34, $z = -2.84$, $p < .001$), but not at other positions (ps > .1). Importantly, there was no three-way interaction, suggesting no effect of Condition within Spanish speakers ($p > .1$). Figure 3.16 displays these results.

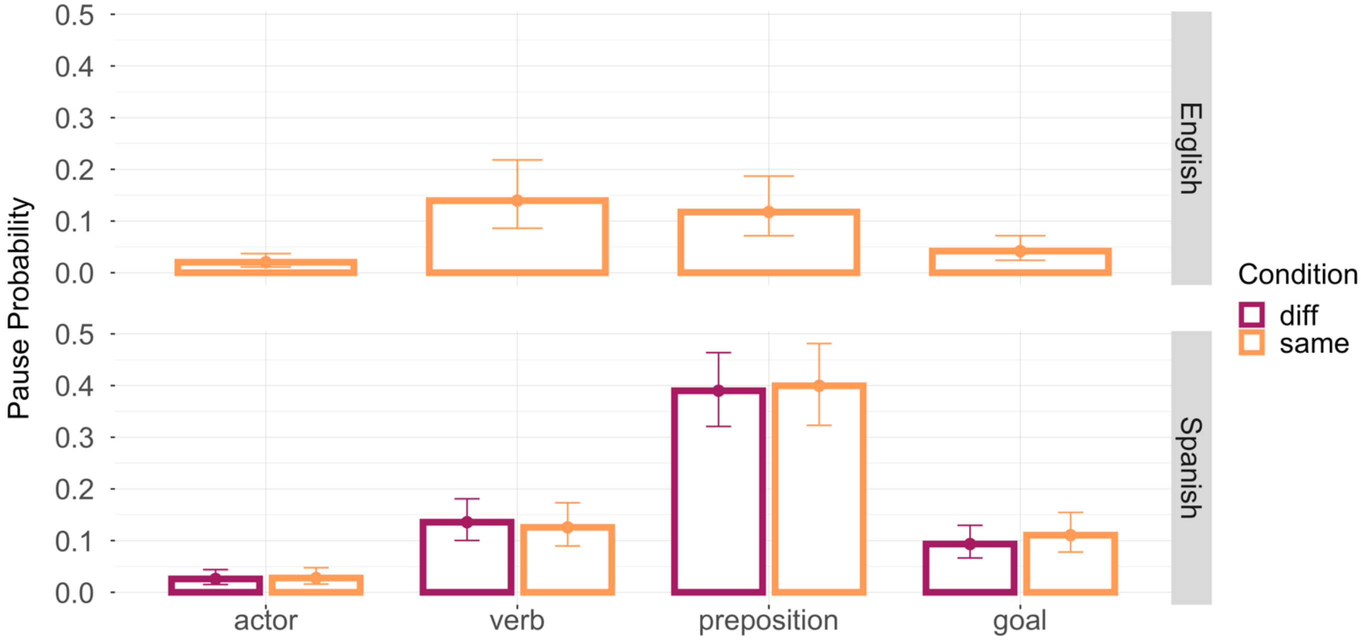

Figure 3.16. Model predictions for the probability of pauses at each chunk position in Experiment 2B, as a function of Language and Condition. Predictions were held at the average level of Source. Error bars represent 95% CI.

Correlation Analysis

We next examined the expected tradeoff between speech initiation time and total utterance duration. We ran a linear mixed effects model regressing Total Duration (logged) on Latency (logged), Condition nested within Language, and their interaction; while controlling for Source. We also included by-subject and by-item random intercepts, and random slopes for Condition nested in Language (by-subject) and Language (by-subject and by-item). Shorter initiation latencies predicted longer utterance durations (b = -0.66 , SE = 0.02 , t = -38.67, p <

.0001), indicating that when participants began speaking sooner, their utterances were longer – suggesting incremental planning during production. However, there was no interaction between Latency, Language, and Condition (*p* > .1); suggesting the relationship was similar across Languages and Condition. Figure 3.17 displays this effect.

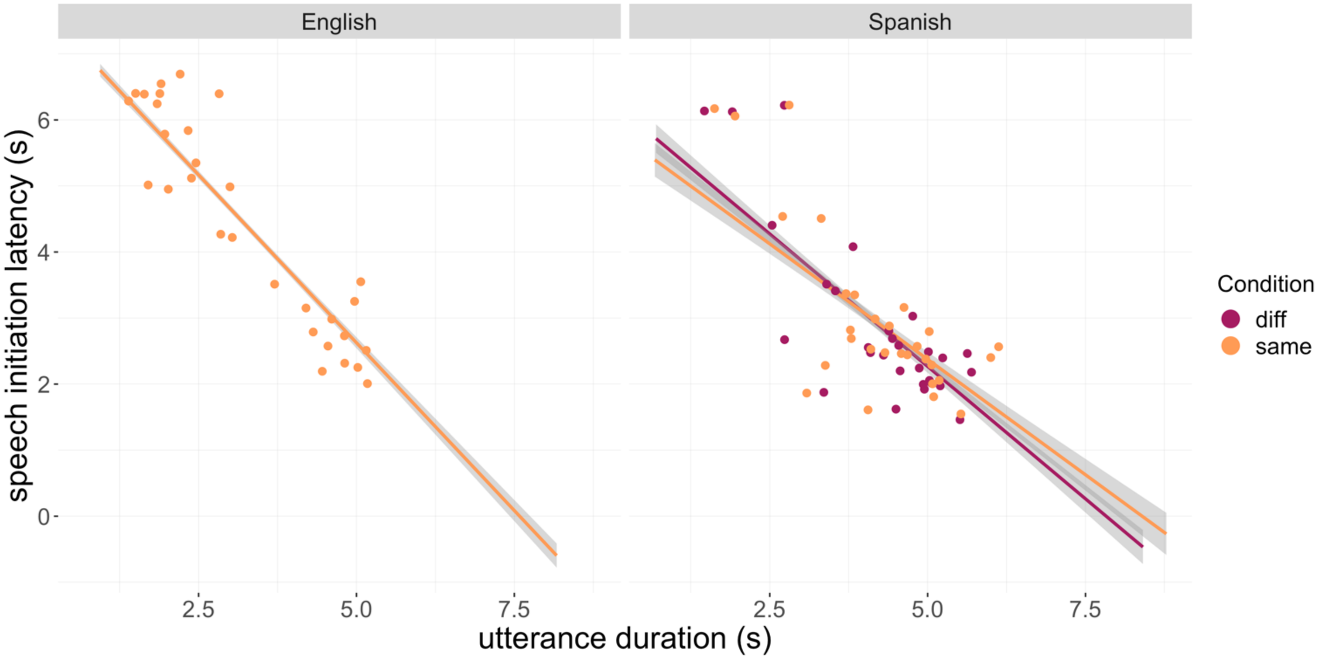

Figure 3.17. Raw scatter plots and fitted lines of the average correlation between speech initiation latency and total utterance duration in Experiment 2B, for Different-gender (purple) and Same-gender (orange) utterances in English (left panel) and Spanish (right panel). Points represent individual participant means. Error bands represent 95% CI.

Experiment 2B Discussion

Results of Experiment 2B do not support our hypothesis that speakers will use any shared information across messages – even just the grammatical gender of an article – to produce their utterance incrementally. Instead, durations were equivalent across the Same-gender and Different-gender conditions of our Spanish stimuli. In comparing our Spanish speakers with English speakers, we again find that Spanish speakers were faster to begin speaking, although they also had more pauses throughout their utterance. That is, Spanish speakers began speaking sooner but were then slowed down by incremental planning during the utterance itself. Figure 3.15 and Figure 3.16 show that the prepositional phrase was the core position of this slowing down – with longer chunk durations and more pauses in Spanish compared to English. Slowing down at this position is expected, given that it precedes the uncertain goal information, and the form of the Spanish article (included in the preposition chunk) differs based on the gender of the goal which is still uncertain (recall that in Experiment 2B all trials were of the uncertain type).

The lack of a difference between gender conditions suggests that Spanish speakers did not take advantage of the similar articles to plan at least an additional word or two before the goal was disambiguated, but rather planned the article together with the goal target once it was disambiguated. There are a number of reasons this might be the case. First, it is possible that participants refrained from retrieving and encoding the two goal options and their grammatical genders until the goal was disambiguated. The time and cognitive cost of encoding both options in order to plan the gender-marked article may not have been worth the effort: producing the article only buys speakers a minimal amount of time, which is likely smaller than the time it would take to retrieve both target options, compared their genders, and plan the gender-marked article if it is equivalent across options. Perhaps having more time before target disambiguation would make this strategy more effective, and then we would see a difference between gender conditions. Note that in English, if participants want to plan the article that is shared across options (*the*), this requires less effort – as there is no need for retrieving the goal options and comparing their genders in advance.

Second, participants show a tendency to pause before the prepositional phrase, as expected given the phrase boundary at that point (see also results of Experiments 1 and 2A). But if they are already pausing at that position, they might simply wait at that point until the goal is disambiguated, allowing them to produce the following phrases fluently without separating between the article and the noun. This could be more of a stylistic or prosodic preference for keeping the article and noun together (see Chapter Two), unless there is good enough reason for separating them – which, as described previously, there likely was not in our task.

Finally, we cannot rule out the possibility that our gender manipulation was not strong or consistent enough. Recall that our stimuli images could have multiple synonymous label options that differ in grammatical gender. Although we adjusted our analyses based on the gender that participants used for the target goal, we could only rely on previous norming for the gender of the competitor goal. While we did provide participants with the intended labels in the familiarization phase, it is possible that they had a different label and gender in mind during the trial, which would flip our condition assignment. Relatedly, it is also possible that because the article is a minor short component, and further mixed in with the prepositional phrase chunk, effects are too small to be detected and instead washed out in the entire utterance timing. Note that there was a marginally significant effect at the preposition chunk, in the expected direction – longer durations in the Different-gender condition. But I am wary of interpreting this result, given the marginality and given that this was the *only* result to differ by condition. However, it might be indicative of a small underlying effect, and perhaps warrants a replication attempt.

Taken together, we do not find evidence that speakers take advantage of just *any* grammatical properties that are shared between message components in order to plan ahead in their utterance. There is a conceptual difference, however, between Experiment 2B and the previous experiments described here. In both Experiment 1 and Experiment 2A, and in Chapter Two, it was a *message* component that could be prioritized and planned earlier. In Experiment 2B, however, it was only a grammatical feature – and related article – of the message that could be planned ahead. The salience of the message component and the more complex planning required for it might make speakers rely more on strategies for incremental planning for the *message* components themselves, but less so for grammatical properties and associated articles. Previous research on comprehension does show that components as small as the grammatical gender of prenominal articles or adjectives can influence predictive processing (Gussow et al., 2019; Lew-Williams & Fernald, 2007; Van Berkum et al., 2005), which is also seen as strategic, proactive comprehension. But the effort and cost of these predictions in comprehension are likely much smaller than that of production. More generally, production is more difficult and effortful than comprehension, and this will affect cues that speakers use in utterance planning.

Chapter Three General Discussion

In three motion event-description tasks, we manipulated whether information about the goal target was clear from the start of the trial, or only disambiguated at the end. In Experiment 1 we found that English speakers were slower to begin speaking when the goal target of a motion event was uncertain until the event was completed. Speakers also had longer utterance durations and more pauses, suggesting they were planning upcoming portions of the utterance simultaneously while speaking, as more of the message information unfolded. In Experiment 2A we found that languages that differ in their permitted word order could encourage different planning strategies in situations of uncertainty. This depends on when the uncertain component must be produced – i.e., at which serial position in the utterance. When the uncertain component was at a later position in the sentence (post-nominal color in Spanish), it encouraged more incremental planning; speakers could begin their utterance and continue planning more as the information unfolded. However, when the uncertain component was at an earlier position (pre-nominal color in English), we found evidence of more advance planning, presumably in order to minimize mid-utterance lengthening and pauses. Finally, in Experiment 2B we did not find that the grammatical gender of potential goal targets (matching or non-matching gender) affects speakers' planning of the gender-marked article preceding the goal. This might suggest that the effort required for comparing potential goals and their grammatical gender is not worth the relatively small benefit of planning (and producing) the article sooner.

Our results support the strong emphasis on flexibility in incremental utterance planning (Ferreira & Swets, 2002; Swets et al., 2021). The language production system can adopt various strategies – advance planning, incremental planning, word lengthening, pauses – in order to plan efficiently for a given context. Efficient planning could translate into earlier speech initiation onset, which is oftentimes preferred in conversation. But if the message itself is still uncertain, early speech initiation might not provide much of an advantage in overall utterance times, i.e., in completing the communication of the message itself. Moreover, it is more cognitively demanding to plan-and-produce simultaneously, while also attending to the message information that is still unfolding. This suggests a necessary balance between advance planning and incrementality, encouraging different production strategies across contexts and task goals. Our results indicate that this balance is influenced by both message properties and grammatical properties of the language in which the utterance is spoken: uncertain messages delay and lengthen production, but the locus of these effects will depend on the language.

Interestingly, we consistently find that in our sample English speakers were later to begin speaking compared to Spanish speakers, regardless of uncertainty condition. Although differences between Spanish and English were not expected in the Direct condition (no message uncertainty) or in Experiment 2B (no differences in word order), they may have carried over as a planning strategy throughout the experiment (see also Swets et al., 2021). Alternatively, the consistent difference in speech initiation times could also reflect differences in speakers' individual language skills affecting their planning strategies. Specifically, our English speaking sample consisted of predominantly monolingual speakers, while our Spanish speaking sample consisted of Spanish-English bilingual speakers. It is possible that the Spanish speakers began speaking earlier because lexical retrieval is typically slower for bilinguals (Gollan et al., 2005), and they wanted to allow themselves more time throughout the utterance to retrieve upcoming words. I did not have enough power to conduct an individual differences analysis, but this could be explored in future studies. Importantly, these sample differences do not affect conclusions from our uncertainty manipulation, given that we looked at the interaction between condition and language rather than interpreting main effects.

Finally, another factor to affect speech timing is task demands, including the timeline of the trial. Recall that in Experiment 1 English speakers had a total of 8 seconds per trial, while in Experiments 2A and 2B they had 10 seconds per trial. The only difference between these timelines was at very end, as the actor in Experiment 1 remained at the goal for two seconds before the trial timed out; while in Experiments 2A and 2B it remained at the goal for four seconds. Aside from this difference, the Uncertain trials in Experiment 1 were effectively equivalent to the trials in Experiment 2B (which were all of the uncertain type and did not include contrastive colors; though the particular stimuli set did differ). A qualitative comparison of the speech initiation time shows that English speakers were faster to begin speaking in Experiment 1 compared to Experiment 2B by almost one second. This could be explained by the shorter timeline of Experiment 1, encouraging English speakers to begin speaking sooner in order to complete their utterance before the trial times out. Note that even in Experiment 2B English speakers tended to complete their utterances before 8 seconds were up (despite having 10 seconds total) – suggesting that they could have starting speaking at equal initiation times in both experiments and still completed the task successfully. But the comparison suggests that a sooner deadline encourages an earlier initiation time, again displaying the sensitivity of strategies for incremental planning to all relevant task factors – including the risk of not meeting the deadline.

Taken together, message uncertainty affects production strategies and the resulting speech patterns, and these effects depend on grammatical properties of the language being spoken. Other decision-making task demands – including risks and benefits, deadlines and efficiency – also appear to influence speakers' production strategies and decisions about when and how to produce their utterance. Uncertainty about the message increases the burden of these demands and the need to strategize utterance planning, which speakers readily do. These results emphasize the need to incorporate message uncertainty into current models of language production, rather than positing that speech *always* begins with a predetermined message.

Chapter Four: Conclusions and future directions

Summary of findings

In this dissertation, I set out to investigate utterance planning under message uncertainty – asking when and how speakers plan their utterances when they are still uncertain of the message to communicate. I started in Chapter One by introducing message uncertainty and reviewing related literature, arguing that message uncertainty might be a quite common context of production, yet it is not accounted for in production models and experimental work. The goal throughout the rest of the dissertation was to begin bringing message uncertainty into the lab and see how it affects utterance planning.

In Chapter Two, I developed a novel picture-naming paradigm to manipulate message uncertainty. Across three experiments I showed that when given a cue about a component of the message (one of two pictures to be named), producers tend to plan the more certain components of a message earlier, and can continue planning the rest incrementally, even after production has begun. This provides a benefit in initiation times, as producers can begin their planning and production without having to wait for the full message to be disambiguated. We found this in both spoken (Experiment 1) and typed (Experiment 2) responses, and even when there was no certainty about any of the message components but only a higher likelihood of one of them being the target based on semantic cues (Experiment 3). These results suggest that speakers make word order decisions that maximize production efficiency under message uncertainty, prioritizing more certain components and planning the rest incrementally.

In Chapter Three, I developed a motion event-description paradigm to investigate the time course and speech patterns of speakers' utterances under message uncertainty. In this paradigm speakers did not have flexibility in word order but instead had to rely on incrementality for planning under message uncertainty. Results from Experiment 1 show that when speakers are uncertain about the goal of the motion event in an unfolding scene, they are later to begin speaking compared to when the goal is apparent from the start. This suggests that speakers are not radically incremental (see also Ferreira & Swets, 2002), but instead allow at least some time for the event to proceed – and the goal information to be closer to disambiguation – before beginning utterance planning. On the other hand, speakers do begin planning even before the goal is disambiguated, suggesting some degree of incremental planning. This incrementality was also apparent in their speech patterns; longer utterance durations and more pauses in cases of message uncertainty.

Experiment 2A replicated this finding while further showing that the degree of incrementality depends on the word order of the language being spoken. When the critical message component (color of the goal object) was at a later serial position in the utterance (post-nominal modifier in Spanish), speakers could plan more incrementally, and showed longer durations at later positions in the utterance. In contrast, when the uncertain message information needed to be spoken at an earlier position (pre-nominal modifier in English), we found more advance planning and longer durations at earlier positions. result from Experiment 2B suggest that the cost of incremental planning under message uncertainty might need to be worth the "effort" – we did not find that Spanish speakers plan a prenominal article when the subsequent noun is uncertain, even if the article is the same across message options. This could be because the article is such a small component (and there is a strong preference to plan it together with the noun as one phrase), making it not worth the cognitive effort of planning the article early and incrementally. Instead, speakers appeared to prefer advance planning of the entire noun phrase.

Together, our results carry implications for language production models and especially theories of incremental planning. While it is generally agreed upon that language production can begin with a message fragment (Levelt, 1989), this is often in the context of a complex message that needs to be broken down into components for easier message-to-utterance processing. But here we show that uncertainty about the identity of message components leads speakers to rely on various strategies for language production; affecting their word order decisions and the time course of sentence production.

Methodological limitations

The main motivations behind this experimental work came from real-life contexts of language production, including turn-taking in conversation and narration of live events. However, for experimental control purposes, our tasks were lab-based (or web-based) and presented a rather artificial context for production. We did attempt to introduce important properties of conversation in context – e.g., telling participants that their recordings would later be used for another participant (Chapter 2) or having an RA give them feedback and leave the impression that the RA was actively listening throughout the experiment (Chapter 3); but this is not comparable to real social interactions. Future studies could improve on these attempts by increasing dialogue and social interaction during the experiment, and/or by diversifying the sentence structures and semantic content elicited by the stimuli. When previous findings from controlled experiments show reliable effects, it is a good opportunity to begin branching out into more natural contexts and examining the limits of the lab-based effects.

Relatedly, while our studies examined very particular types of message uncertainty, there are many more that emerge in natural production. In the studies presented here, the majority of the message was known already, with only a component (corresponding to a word) missing. But as noted in Chapter One, there could also be cases where a speaker is debating between multiple messages, retrieving information from memory, developing a vague idea, and many other examples. The implications of these contexts for speech patterns might be different than those I discussed here. For example, I did not study errors in production because the paradigms I used were rather simple and speech errors were rare. But understanding different contexts of message uncertainty and their implications for utterance planning will be important for generalizing any theory of language production under message uncertainty.

On decision-making in language production

The flexibility of the language production system makes a decision-making framework crucial for understanding speakers' planning under message uncertainty. Currently there is wide agreement on a highly flexible production system (Brown-Schmidt & Konopka, 2015; Ferreira & Swets, 2002; Swets et al., 2021), as found in our results too, with production strategies and decisions that are highly context-dependent. But we now need to explicitly include these context pressures and decision-making variables in models of language production processing (both theoretical and computational), allowing a more precise account of how production decisions are made.

There is already work in this direction in the context of single word naming. For example, *evidence accumulation* has been discussed and modeled for decision-making in lexical selection tasks (Anders et al., 2015). The suggestion is that reaction times in word naming depend on the activation of various competing targets, with a threshold determining how much activation a given target needs to accumulate before it is spoken – arguably, how much certainty it needs.

Relatedly, *signal detection theory* has been applied to model speech errors in single word naming, allowing to predict naming errors as a function of conflict between semantically similar words (Nozari & Hepner, 2019). Flexibility in production strategies is instantiated by shifts in the threshold, or criterion, for speech – how much certainty is needed for word production to begin? For example, the threshold differs across contexts (low semantic competition versus high semantic competition), it shifts when participants are told to avoid errors, and it depends on speakers' individual differences. In fact, shifts in the threshold could be used explain differences in aphasia patients' performance profiles (e.g., fluent but error-prone versus near-mute). When these patients have difficulty in evidence accumulation (impaired lexico-semantic representations), uncertainty is high. This could result in a production error or in no response at all, depending on the patient's individual threshold; with implications for the optimal treatment plan (Nozari & Hepner, 2019).

Similar decision-making considerations are crucial for understanding production under *message* uncertainty. As discussed previously, a speaker must decide how much uncertainty they are willing to tolerate when they begin speaking, and/or whether it is worth waiting for more information to unfold before beginning utterance planning. There is naturally a cost to early planning under message uncertainty – the need to plan and produce simultaneously, while still deciphering new message information. Similarly, there are risks involved – the risk of getting stuck mid-utterance or even producing an error; either because the speaker expected a different message outcome (e.g., a different goal location) or because they were considering various messages that interfered with each other (Harley, 1984). On the other hand, there is an efficiency benefit to beginning speech early, and sometimes speakers are even required to begin their utterance before the message is certain.

So, how is the threshold determined for speech initiation in cases of message uncertainty? The answer could depend on properties of the message itself (e.g., prominence or importance of the uncertain component) or the language (e.g., early production of a short word might not be worth a high cognitive cost). However, there are also higher-level context effects to consider. For example, perhaps the threshold for certainty would be higher in a high-stakes conversation such as a job interview, where speakers would prefer to spend more time thinking through their answer before uttering even parts of a response; in fear of either getting stuck and disfluent, or of uttering something that they would need to take back.

Another consideration is the timing available for production. There is some indication in our results that a sooner deadline encourages speakers to begin production sooner, even if they have plenty of time to complete their utterance (see Chapter Three Discussion). In that case, it is the risk of missing the opportunity that encourages earlier speech and more incremental planning. But that risk might differ across contexts too. The risk in our experiment was arguably rather small; participants were probably motivated to do well but there was no penalty for missing the time window for production and no external reward for earlier production. Introducing explicit penalties and rewards would likely bias speakers' behaviors, however.

To conclude, if we want to give answers that are more precise than "it depends" for any question about language production strategies, we will need to investigate speakers' behaviors across contexts and languages, and build up an account that can explain production decisions in various contexts. Modeling these contexts and decision-making pressures would allow a better understanding of how utterance planning decisions are made in the context of message uncertainty (but not only), and could produce more precise predictions for future research.

Conclusion

Message uncertainty poses particular demands on the language production system, but this highly flexible and strategic system can cope with the challenge: across six experiments I found that speakers adapt their utterances in the face of uncertainty, using the flexibility of word order and durations to plan their utterances early and communicate their messages efficiently. Despite the rather conclusive results, perhaps the biggest takeaway is that there are many more questions and work to be done. A highly flexible system suggests several other contexts to explore – across task demands, goals, languages, and speakers – working towards a unified account that can explain the variety of strategies and outcomes, and predict their occurrence.

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

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

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

Anders, R., Riès, S., van Maanen, L., & Alario, F.-X. (2015). Evidence accumulation as a model for lexical selection. *Cognitive Psychology*, *82*, 57–73. https://doi.org/10.1016/j.cogpsych.2015.07.002

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*, *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*. https://www.frontiersin.org/article/10.3389/fpsyg.2020.00189

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research*, *32*(1), 25–36.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Balzer-Siber, M. (2015). *Functional and Stylistic Features of Sports Announcer Talk: A Discourse Analysis of the Register of Major League Soccer Television Broadcasts* [PhD Thesis]. East Tennessee State University.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00328

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

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., Herron, D., Ching Lu, C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., … Tzeng, O. (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.

Beaty, R., Frieler, K., Norgaard, M., Merseal, H., MacDonald, M., & Weiss, D. (2020). *Spontaneous Melodic Productions of Expert Musicians Contain Sequencing Biases Seen in Language Production*. https://doi.org/10.31234/osf.io/qdh32

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, 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. (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, 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

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. (2007). Syntactic Priming. *Language and Linguistics Compass*, *1*(1–2), 1–16. https://doi.org/10.1111/j.1749-818X.2006.00001.x

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. (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

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

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 Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

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. *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*. COLING 1996. https://aclanthology.org/C96-1052

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2017). MultiPic: A standardized set of 750 drawings with norms for six European languages. *The Quarterly Journal of Experimental Psychology*, *0*(ja), 1–24. https://doi.org/10.1080/17470218.2017.1310261

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*. *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

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*, 1220–1234. https://doi.org/10.3758/BF03193224

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. (2023). Language production under message uncertainty: When, how, and why we speak before we think. In K. D. Federmeier & J. L. Montag (Eds.), *Psychology of*

Learning and Motivation: Speaking, Writing, and Communicating (Vol. 78). Academic Press. https://doi.org/10.1016/bs.plm.2023.02.005

Gussow, A. E., Kapnoula, E. C., & Molinaro, N. (2019). Any leftovers from a discarded prediction? Evidence from eye-movements during sentence comprehension. *Language, Cognition and Neuroscience*, *34*(8), 1041–1058. https://doi.org/10.1080/23273798.2019.1617887

Gussow, A., & MacDonald, M. (2021). Language Production Under Uncertainty: Advance Planning and Incrementality. *34th 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

Heller, D., Skovbroten, K., & Tanenhaus, M. K. (2009). Experimental evidence for speakers sensitivity to common vs. Privileged ground in the production of names. *Proc. PRE-CogSci*.

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

Hieke, A. E., Kowal, S., & O’Connell, D. C. (1983). The Trouble with “Articulatory” Pauses. *Language and Speech*, *26*(3), 203–214. https://doi.org/10.1177/002383098302600302

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91–117. https://doi.org/10.1016/0010-0277(96)81418-1

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

Jacobs, C. L., Cho, S.-J., & Watson, D. G. (2019). Self-Priming in Production: Evidence for a Hybrid Model of Syntactic Priming. *Cognitive Science*, *43*(7), e12749. https://doi.org/10.1111/cogs.12749

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., 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

Klimovich-Gray, A., Tyler, L. K., Randall, B., Kocagoncu, E., Devereux, B., & Marslen-Wilson, W. D. (2019). Balancing Prediction and Sensory Input in Speech Comprehension: The Spatiotemporal Dynamics of Word Recognition in Context. *The Journal of Neuroscience*, *39*(3), 519–527. https://doi.org/10.1523/JNEUROSCI.3573-17.2018

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

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–131). New York, NY: Wiley.

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 MIT Press. *Cambridge, MA*.

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.

Lewis, A. G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex*, *68*, 155–168. https://doi.org/10.1016/j.cortex.2015.02.014

Lew-Williams, C., & Fernald, A. (2007). Young Children Learning Spanish Make Rapid Use of Grammatical Gender in Spoken Word Recognition. *Psychological Science*, *18*(3), 193–198. https://doi.org/10.1111/j.1467-9280.2007.01871.x

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

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

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

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Interspeech*, *2017*, 498–502.

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

Nozari, N., & Hepner, C. R. (2019). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive Neuropsychology*, *36*, 193–207. https://doi.org/10.1080/02643294.2018.1476335

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

Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H., & Yuan, J. (2014). FAVE (Forced Alignment and Vowel Extraction) Program Suite v1. 2.2. *DOI: Https://Doi. Org/10.5281/Zenodo*, *22281*.

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

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

Snyder, K. M., & Logan, G. D. (2014). The problem of serial order in skilled typing. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1697–1717. https://doi.org/10.1037/a0037199

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., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *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

Swets, B., Fuchs, S., Krivokapić, J., & Petrone, C. (2021). A Cross-Linguistic Study of Individual Differences in Speech Planning. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.655516

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 Berkum, J. J. A., 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*(3), 443–467. https://doi.org/10.1037/0278-7393.31.3.443

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

Westfall, J. (2016). *PANGEA: Power ANalysis for GEneral Anova designs (working paper)*.

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

Appendices

Appendix 1. Chapter 3 Stimuli lists.

Experiment 1

Item ID	Source	Goal	Competitor	Actor
1	nest	park	river	bird
2	path	bench	door	runner
3	house	pyramid	church	car
4	farm	lake	well	horse
5	bedroom	fridge	mailbox	prince
6	wave	lighthouse	iceberg	boat
7	plant	forest	beach	toad
8	bed	window	meat	cat
9	branch	windmill	statue	eagle
10	fence	bone	tree	dog
11	wheat	tractor	scarecrow	farmer
12	rain	acorn	leaf	frog
13	honey	daisy	clover	bee
14	bathroom	backpack	oven	girl
15	wheelbarrow	tomato	cucumber	gardener
16	lab	ambulance	pharmacy	doctor

17	mountain	sun	logs	bear
18	xbox	court	pool	boy
19	doll	stairs	rug	baby
20	campsite	cannon	mansion	taxi
21	submarine	island	treasure chest	pirate
22	gold	throne	castle	queen
23	graveyard	statue of liberty	great wall of china	delivery man
24	cloud	airport	city	helicopter
25	drum	microphone	piano	singer
26	mushroom	stump	rose	fairy
27	outlet	money	diamond	burglar
28	anchor	ship	pearl	mermaid
29	printer	bell	computer	teacher
30	flower	rock	cactus	scorpion
31	clarinet	harp	saxophone	man
32	lamp post	sunflower	fountain	crow
33	olive	lettuce	fruit	mouse
34	table	bible	cross	priest
35	cabin	fireplace	broom	witch
36	bridge	desert	mine	train
37	gym	market	stadium	woman
38	mirror	crown	dress	princess

39	grapes	onion	pineapple	beetle
40	trophy	watch	ball	turkey
41	pepper	cupcake	avocado	squirrel
42	harbor	notebook	paint brush	artist
43	truck	hose	rope	fireman
44	asparagus	pumpkin	strawberry	cow
45	coconut	apple	cheese	monkey
46	chair	shield	fire	knight
47	television	bike	radio	policeman
48	spaceship	moon	planet	astronaut
49	basket	pear	corn	lizard
50	present	hammer	screwdriver	elf

Experiment 2A

ID	**source**	**goal**	**target_color**	**competitor_color**	**actor**	**source_ES**	**goal_ES**	**target_color_ES**	**competitor_color_ES**	**actor_ES**
1	bench	umbrella	blue	red	monkey	banco	paraguas	azul	rojo	mono
2	skateboard	watch	brown	blue	robot	patineta	reloj	marrón	azul	robot
3	piano	barrell	red	green	waiter	piano	barril	rojo	verde	mesero
4	castle	door	green	yellow	donkey	castillo	puerta	verde	amarillo	burro
5	orange	fountain	grey	blue	bird	naranja	fuente	gris	azul	libélula
6	tunnel	car	blue	red	man	túnel	carro	azul	rojo	hombre
7	doll	present	blue	pink	woman	muñeca	regalo	azul	rosado	mujer
8	fireplace	bible	grey	green	nun	fuego	biblia	gris	verde	monja
9	chair	hairbrush	brown	black	butterfly	silla	cepillo	marrón	negro	mariposa
10	cheese	candy	pink	yellow	mouse	queso	dulce	rosado	amarillo	ratón
11	drawer	suitcase	blue	red	frog	cajón	maleta	azul	rojo	rana
12	bin	bell	yellow	black	mosquito	basura	campana	amarillo	negro	mosquito

13	puppet	balloon	blue	red	monster	marioneta	globo	azul	rojo	monstruo
14	stump	fork	brown	green	lumberjack	tocón	tenedor	marrón	verde	leñador
15	helmet	boot	red	purple	eagle	casco	bota	rojo	morado	Águila
16	dart board	bowling pin	blue	yellow	man	tablero de dardos	bolo	azul	amarillo	hombre
17	chair	T-shirt	pink	blue	dog	silla	camiseta	rosado	azul	perro
18	roof	football	pink	brown	pelican	techo	balón	rosado	marrón	pelícano
19	biscuit	soap	pink	green	ant	galleta	jabón	rosado	verde	hormiga
20	fence	brick	brown	blue	sheep	cerca	ladrillo	marrón	azul	oveja
21	blanket	box	brown	red	elephant	manta	caja	marrón	rojo	elefante
22	stool	rope	brown	pink	chicken	silla	soga	marrón	rosado	pollo
23	stairs	tie	green	pink	soldier	escaleras	corbata	verde	rosado	soldado
24	milk	pot	red	brown	beetle	leche	maceta	rojo	marrón	escarabajo
25	puddle	shell	grey	brown	snail	charco	cascarón	gris	marrón	caracol
26	flower	bottle	green	blue	seal	flor	botella	verde	azul	foca

27	washing machine	towel	purple	blue	bee	lavadora	toalla	morado	azul	abeja
28	treasure chest	candle	grey	orange	prince	tesoro	vela	gris	naranja	príncipe
29	bath	broom	brown	pink	wizard	bañera	escoba	marrón	rosado	mago
30	daisy	hat	black	grey	wasp	margarita	sombrero	negro	gris	avispa
31	notebook	sponge	yellow	blue	fairy	cuaderno	esponja	amarillo	azul	hada
32	fish tank	crutches	black	brown	girl	pecera	muletas	negro	marrón	niña
33	sofa	teapot	blue	green	nurse	sofá	tetera	azul	verde	enfermera
34	table	keyboard	orange	grey	knight	mesa	teclado	naranja	gris	caballero
35	broccoli	feather	red	blue	bear	brócoli	pluma	rojo	azul	oso
36	lighthouse	television	purple	brown	runner	faro	televisión	morado	marrón	corredor
37	sledge	lawnmower	green	orange	fox	trineo	cortacésped	verde	naranja	zorro
38	cactus	walking stick	brown	orange	farmer	cactus	bastón	marrón	naranja	granjero

39	fire	shield	grey	purple	gladiator	fuego	escudo	gris	morado	gladiador
40	net	lace	yellow	blue	pig	neto	cordón	amarillo	azul	cerdo

Experiment 2B

ID	**source_ES**	**goal_ES**	**comp_ES**	**actor_ES**	**goal gen**	**comp gen**	**source_EN**	**goal_EN**	**comp_EN**	**actor_EN**
1	nido	parque	río	mujer	M	M	nest	park	river	woman
2	camino	banco	puerta	corredor	M	F	path	bench	door	runner
3	pozo	pirámide	altar	coche	M	M	well	pyramid	church	car
4	lago	casa	granja	caballo	F	F	lake	house	farm	horse
5	habitación	faro	buzón	príncipe	M	M	bedroom	lighthouse	mailbox	prince
6	planta	bosque	playa	sapo	M	F	plant	forest	beach	toad
7	cama	ventana	carne	gato	F	M	bed	window	meat	cat
8	rama	molino	estatua	águila	M	F	branch	windmill	statue	eagle
9	valla	hueso	pino	perro	M	M	fence	bone	tree	dog

10	espantapájaros	tractor	trigo	pastor	M	M	scarecrow	tractor	wheat	farmer
11	lluvia	bellota	hoja	rana	F	F	rain	acorn	leaf	frog
12	miel	margarita	trébol	abeja	F	M	honey	daisy	clover	bee
13	baño	mochila	horno	niña	F	M	bathroom	backpack	oven	girl
14	carretilla	tomate	pepino	jardinero	M	M	wheelbarrow	tomato	cucumber	gardener
15	laboratorio	ambulancia	farmacia	médico	F	F	lab	ambulance	pharmacy	doctor
16	montaña	sol	leña	oso	M	F	mountain	sun	logs	bear
17	consola	frontón	piscina	niño	M	F	xbox	court	pool	boy
18	muñeca	escalera	alfombra	bebé	F	F	doll	stairs	rug	baby
19	camping	cañón	mansión	taxi	M	M	campsite	cannon	mansion	taxi
20	submarino	isla	tesoro	pirata	F	M	submarine	island	treasure chest	pirate
21	lingote	trono	torre	reina	M	M	gold	throne	castle	queen

22	nube	aeropuerto	ciudad	helicóptero	M	F	cloud	airport	city	helicopter
23	tambor	micrófono	piano	cantante	M	M	drum	microphone	piano	singer
24	seta	tronco	rosa	hada	M	F	mushroom	stump	rose	fairy
25	enchufe	dinero	diamante	ladrón	M	M	outlet	money	diamond	burglar
26	ancla	barco	perla	sirena	M	F	anchor	boat	pearl	mermaid
27	impresora	timbre	ordenador	profesor	M	M	printer	bell	computer	teacher
28	flor	piedra	cactus	escorpión	F	M	flower	rock	cactus	scorpion
29	clarinete	arpa	saxofón	duende	F	M	clarinet	harp	saxophone	elf
30	farola	girasol	fuente	cuervo	M	F	lamp post	sunflower	fountain	crow
31	pastel	aceituna	fruta	ratón	F	F	cupcake	olive	fruit	mouse
32	mesa	biblia	tumba	cura	F	F	table	bible	cross	priest
33	cabaña	chimenea	escoba	bruja	F	F	cabin	fireplace	broom	witch
34	puente	desierto	mina	tren	M	F	bridge	desert	mine	train
35	espejo	corona	vestido	princesa	F	M	mirror	crown	dress	princess

36	uvas	cebolla	piña	cucaracha	F	F	grapes	onion	pineapple	beetle
37	trofeo	reloj	pelota	pavo	M	F	trophy	watch	ball	turkey
38	pimiento	pera	aguacate	ardilla	F	M	pepper	pear	avocado	squirrel
39	puerto	cuaderno	pincel	pintor	M	M	harbour	notebook	paint brush	artist
40	cuerda	manguera	camión	bombero	F	M	rope	hose	truck	fireman
41	espárrago	calabaza	fresa	vaca	F	F	asparagus	pumpkin	strawberry	cow
42	coco	manzana	queso	mono	F	M	coconut	apple	cheese	monkey
43	silla	escudo	fuego	guerrero	M	M	chair	shield	fireplace	knight
44	televisión	bicicleta	destornillador	policía	F	M	television	bike	screwdriver	policeman
45	regalo	luna	martillo	astronauta	F	M	present	moon	hammer	astronaut
46	cesta	lechuga	maíz	lagartija	F	M	basket	lettuce	corn	lizard

Appendix 2. Within-subject variability on Source naming in Experiment 1 of Chapter 3

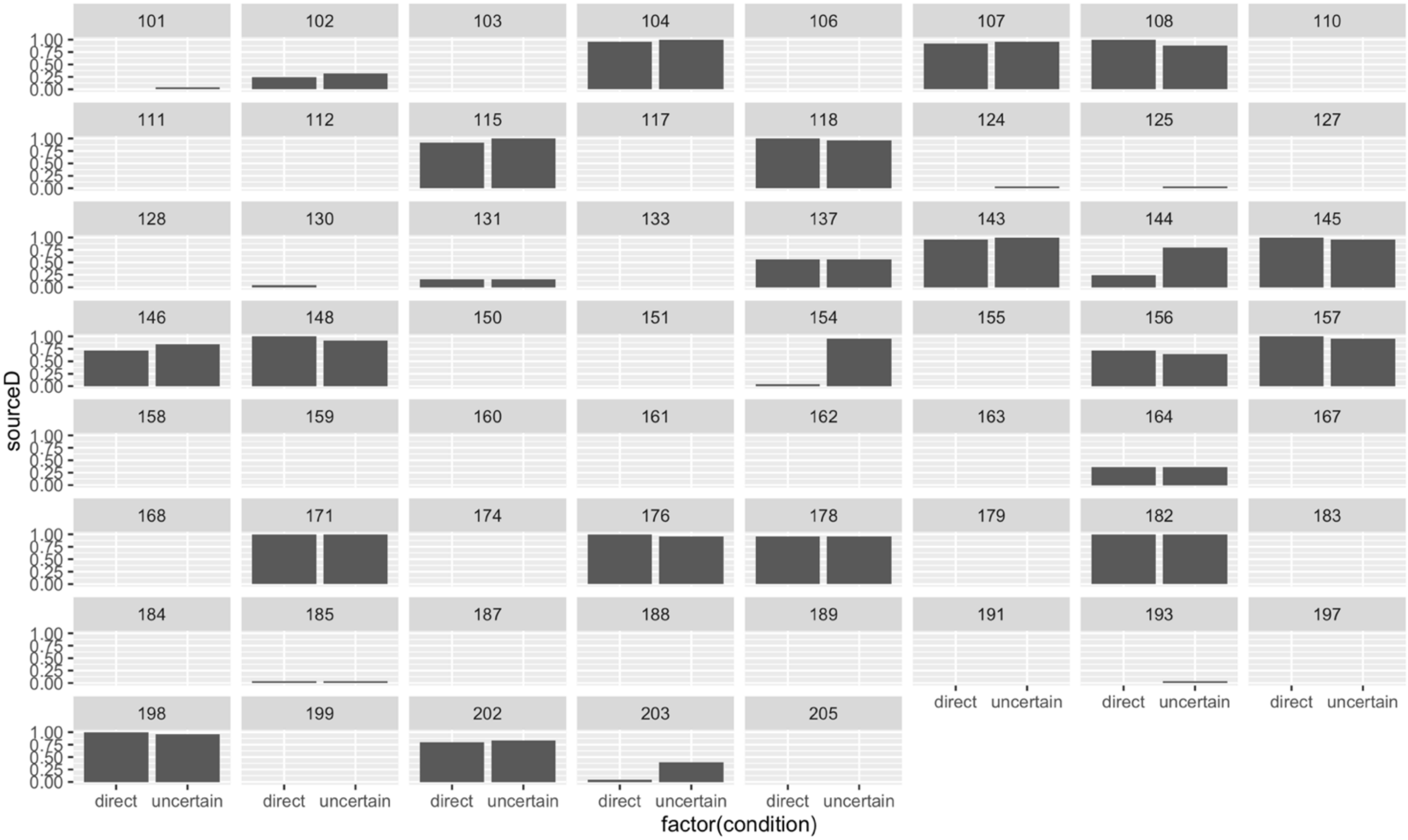

Figure A2.1. Proportion of trials (y-axis) in which each subject mentioned the source in the Direct and the Uncertain conditions in Experiment 1 of Chapter 3. Each grid represents a participants; numbers are participant IDs. The majority of participants either always mention the source (proportion equals 1) or never mentioned the source (proportion equals 0).