THE UNIVERSITY *of* EDINBURGH

The Relationship Between Disfluencies, Associations, and Inferences in Speech Comprehension

Esperanza R. Badaya

Supervisors: Prof. Martin Corley & Dr. Hannah Rohde

THE UNIVERSITY *of* EDINBURGH

Doctor of Philosophy

Department of Psychology

School of Philosophy, Psychology, and Language Sciences

The University of Edinburgh

2023

Declaration

I hereby declare that:

(1) this thesis is my own composition,

(2) the work reported in this thesis has been carried out by myself, and that the work of others is acknowledged in the text, and

(3) the work presented in this thesis has not been submitted for any other degree or professional qualification.

Esperanza R. Badaya

Abstract

When producing speech spontaneously, speakers do more than produce words: They hesitate, correct themselves, and fill their pauses with *um* and *er* - a range of phenomena referred to as *disfluencies*. In turn, the presence of these disfluencies can affect speech comprehension: Filled pauses, such (i.e., *um* or *uh*), have been widely attested to affect how listeners process and interpret speech. For example, the presence of a filled pause has been shown to bias comprehenders' expectations of what will follow them (e.g., discourse-new entities, Arnold et al., 2004; hard-to-describe objects, Arnold et al., 2007) and their evaluations of both the speaker and the message (e.g., uncertainty, Brennan & Williams, 1995; deception, Arciuli et al., 2010).

This thesis investigates whether the online processing of disfluent speech and the interpretation of meaning can be accounted for by similar mechanisms. Previous research has shown that filled pauses are produced in predictable patterns and that their production offers a window into the speaker's mental state. Consequently, the biases exerted by disfluencies can be accounted for by comprehenders' passive learning of the distribution of filled pauses, and by a form of social reasoning about the causes for the speaker to experience trouble in speech production. These two mechanisms have contrasting predictions regarding the flexibility and the costs associated with comprehending disfluent speech. We took prediction of upcoming lexical items and interpretation of deceit as the test bed for these questions. The series of investigations reported here took a novel approach by

comparing language comprehension in first and second language when speech is produced by first- or second-language speakers.

Part I explores a proposed process for efficient speech comprehension: prediction. In Experiments 1 and 2, we replicated and extended Bosker et al. (2014) eye-tracking studies wherein native listeners displayed anticipatory eye movements towards low-frequency items upon encountering native, but not non-native, disfluencies. In two experiments, we explored whether the presence of a filled pause led native and non-native listeners to anticipate a low-frequency word and whether this was dependent on the speaker's identity, i.e. if it was a native or a non-native speaker. We found clear effects in a time window of analysis reflecting word recognition. For native comprehenders, the presence of a disfluency aided the recognition of a low-frequency word, regardless of the speaker's linguistic background. In contrast, a disfluency produced by a native speaker increased the recognition of a low-frequency word in non-native listeners, while filled pauses produced by a non-native speaker benefited the recognition of both high- and low-frequency words. These results suggest that the addition of time without propositional content is a likely, but not a sufficient, explanation for the benefits due to the presence of disfluencies. Instead, comprehension of elements accompanying disfluencies is particularly beneficial when these items contextually co-occur with disfluencies.

Part II of this thesis explores how comprehenders interpret disfluent utterances. In Experiments 3 and 4 we investigated how listeners interpret disfluent speech as deceitful as a function of the speaker's and the listener's linguistic background. We replicated and extended Loy et al.'s (2017) eye-tracking studies in which native listeners were more likely to interpret disfluent utterances as deceitful, which was reflected in an early bias in their eye movements. Across two eye-tracking experiments, we found that utterances containing a filled pause were more likely to be interpreted as deceptive. Importantly, the emergence of this bias occurred early in the time course of comprehension, as evidenced by eye movements. Further, listeners were insensitive to the presence of alternative causes for the speaker to be disfluent (e.g., producing speech in their second language), nor

did the task demands (e.g., comprehending speech in their second language) impact the emergence of this disfluency-as-deception bias. The speed with which the effect emerged, alongside its invariance, suggests that the bias has its roots in comprehenders' stereotypes about the sound of deceit, i.e., an association between disfluency and deception.

Overall, the findings of this experimental work support an account where the effects of filled pauses are better conceptualised as a routine. Experience with language creates a 'heuristic' whereby filled pauses are contextually associated with language production difficulties, which constrains processing in a relatively cost-free manner. The emergence of this heuristic may fall under a general capacity of comprehenders to monitor their interlocutor (e.g., epistemic vigilance) which evaluates both their competence and their honesty. Further studies should explore the combination of other verbal and non-verbal cues associated with speaker confidence to investigate whether comprehension of filled pauses is indeed a reflection of the routinisation of social cognition in language comprehension.

Lay Summary

Spontaneous speech is hardly perfect: Speakers tend to correct themselves, repeat words, or fill their pauses with '*um*' and '*uh*'. In contrast to what would be expected, the presence of these elements - what we will refer to as *disfluencies* - has been shown to aid, rather than hinder, speech comprehension: For example, listeners tend to expect their interlocutor to say some things over others when they encounter a disfluency. Further, the presence of a disfluency can even drive listeners to interpret the sentence beyond from what it literally says: For example, a disfluent sentence can lead listeners to believe that the speaker is lying.

How does this happen? On the one hand, disfluencies occur in predictable patterns: They are more likely to occur before a speaker refers to a rare item, or they use an uncommon word. On the other hand, disfluencies are more likely to occur when the speaker struggles to produce speech. These two features map onto two well-known abilities of human animals: The ability to detect statistical regularities (i.e., associations), and to reason about others (i.e., inferences). In this thesis, we explored what feature could account best for how we comprehend disfluent speech. We investigated the effects of disfluencies on listeners' anticipations of what will come after the disfluency, and the effects of disfluency in the interpretation of deceit as test beds to explore this question. To do so, we explored what happens when speech is produced by a native or a non-native speaker: The latter group may produce more disfluencies by virtue of producing speech in their second language, and thus their disfluencies may not be taken as natives' are.

We also contrasted how disfluent speech is comprehended in first- and second-language comprehension: The latter group is arguably in a more difficult situation, and thus their comprehension of disfluent speech may differ.

In Part I, we investigated whether listeners expect an uncommon word following a disfluency. We measured listeners' eye movements as they observed a scene and followed a speaker's instructions. The speaker, who could be a native or a non-native English speaker, referred to one (out of two) objects on the screen either fluently or disfluently (i.e., *Click on*/*Click on thee uh*). Importantly, the two objects on the screen were pairs of objects who are referred to by common and uncommon words (e.g., a train and an accordion). We found that for native listeners, the presence of a disfluency aided in comprehension of rarer words, regardless of the speaker's linguistic background - but this was not due to listeners anticipating such elements. In contrast, non-native listeners recognised rarer words better when a native speaker produced a filled pause while they recognised both common and uncommon words better when a non-native speaker was disfluent. We interpret these findings as reflecting individuals' having learnt an association between disfluencies and what can cause problems for a speaker to be disfluent.

In Part II, we took the same elements to explore whether listeners interpret disfluent speech as deceptive. We measured listeners' eye movements as they attended to speech produced by a potentially deceiving speaker, who referred to one (out of two) objects on a screen as the potential location of a treasure. We found that native and non-native listeners were more likely to take disfluent instructions as deceitful, regardless of who uttered them: Listeners were more likely to click on the object the speaker did not refer to as the actual location of the treasure when the speaker was disfluent. Importantly, listeners started to fixate on such objects shortly after they were produced. This pattern did not differ between speakers and listeners. We take these findings as evidence that, instead of reasoning about why the speaker is disfluent, listeners interpret disfluent speech following a stereotype of how deception sounds - which, coincidentally, happens to include disfluencies.

Overall, this thesis aligns with the idea that comprehension of disfluent speech follows listeners' associations between disfluencies and the speaker's trouble in speech production - even when this association is not accurate. There is a possibility that these associations emerge as a consequence of a repeated inference, in order to ease comprehension of spontaneous speech. Future research should explore the consequences of this assumption.

Acknowledgements

Doing this doctorate is definitely a supposedly fun thing I will never do again[1]. However, I (allegedly) made it and here is a very, very, long list of people who have made it possible. I will try my best to name all (is the least I can do, I am not getting anyone a drink), but if I forget to name anyone: I unapologetically blame my sleep-deprived brain.

I must start with my supervisors, Martin Corley and Hannah Rohde. From providing feedback at lightning speed these last three months, to patiently listening to me ramble for hours without making an actual point and dealing with my 'tomorrow! I'll send it tomorrow', Martin and Hannah have provided a great deal of support and wisdom. Martin, thank you for your advice and support throughout these years, dealing with my crippling anxiety, and for sharing your top tips for writing/analysing/good tapas in Edinburgh. Hannah, thank you for your always insightful feedback and your very special skill of phrasing it as if that is what I have meant all along (when it clearly wasn't). I do look up to both of you as researchers and supervisors, and I hope my future endeavours show, if even in the slightest, what great mentorship I've had.

Following the university-related mandatory thank-you's, the PG Office (and in particular Katie Keltie and Toni Noble) has incredibly been helpful - especially when I had to navigate this very kafkaesque system. Ronny Wieland and Simon Smith have dealt with all my emails, my forgetting passwords, and trying over-the-top complicated stuff at the eye-tracking lab. Thanks to the wonderful stats team we have (and have had) at the department: Dr. Anastasia Ushakova, Dr. Josiah King, Dr. Emma Waterston,

[1] I am just being dramatic for the performance. It was alright.

and Dr. Umberto Noè; Dr. Anita Tobar-Henríquez for offering me the chance to work with her and her advice to navigate the post-PhD world; Dr. Patrick Sturt and Dr. Dan Mirman for their advice during my annual reviews, and for always approaching them with kindness and empathy. I would also like to thank the members of the Cognitive Neuroscience Section at the Carlos III Health Institute because that is where I took my first steps as a researcher: Thanks to Laura, David, Pili, Sabela, Zahra, José, Paco, Javi, and Manolo for everything I learnt from you and supporting me when I was applying to Edinburgh. During my studies, I have also had the chance to tutor a lot of bright students, out of which I must mention Milla, Caroline, and Lea: the three of you alone gave me more headaches than my thesis, but I am also immensely proud of you and all of your achievements. Finally, I am grateful for several PPLS Research Grants that funded all the experiments in this thesis as well as my attendance at several conferences.

I am also grateful to the Scottish Graduate School for Social Sciences for the Saltire Emerging Research award, which funded my five-month stay at Ghent University (and also made the transition from student to non-student smoother and less traumatic). These thanks extend to Prof. Rob Hartsuiker for taking me into his lab and for all the fruitful conversations we had, and to all the Hartsuiker lab members: Anna, Merel, Mariia, Nathan, Zaynab, Aurelie, and Binger. During this stay, Prof. Stef Grondelaers tested the limits of human imagination with all the stimuli we created together. Finally, a special shout-out has to go to Irene Winther, who constantly took me outside of my comfort zone, and boy, how much fun I had!

Before I move on to the much dreaded emotional, non-academic related acknowledgments, I would like to thank my examiners, Dr. Christina Kim, Dr. Christoph Scheepers, and Dr. Chris Cummins: Not only for making the viva such a great experience (only now I can acknowledge that I really enjoyed it), but because your feedback has undoubtedly contributed to this thesis (any errors - especially typos - remain my own).

PPLS is a great place to be (in addition to the people mentioned before) because the following individuals are (or were) in it: the Drs-to-be (or already Drs.) Greta,

Riccardo, Carolane, Tanvi, Abby, Loris, Sheeren, Emma, Alistair, Federica, Renzo, Lasse, Sam, Alex, Wei, Karim, Lena, Aida, Carine, Irene, and Bonan (and the addition of my fellow co-tutors, not my friends just yet, Yav and Ross). I am very grateful that I get to share out-of-office spaces with most of you. Special mentions go to Greta, I have to confess the *caffetones* were just an excuse: spending time with you (even if we don't climb) has kept me sane and is one of my favourite things to do in Edinburgh (and abroad! we are so european). My lads, Tanvi and Abby, have filled my days with entertainment, joy, one too many bevvies, and sometimes, public embarrassment (I'm looking at you Abby). Tanvi has told me she's buying new friends to replace me once I leave the UK – maybe instead of friends you could buy your own snacks and stop eating mine? (but also: we both know I would not have survived these months without your patience, your love, and your food). Carolane, most of the people on this list have to thank you for forcing me out of my shell when we first met (even when I was, how to put it, not the nicest I can be). Emma and Alistair, definitely not thank you for taking me to Spanish events to drink bad wine and beer, but thank you for sharing your "purely out of hobby and not of addition" shelf with gin and beer. Riccardo, thank you for keeping me sane during the pandemic with our Blackford Hill hikes, book exchanges, and, most importantly: your lasagna. I miss you terribly.

In the almost six years I have spent in Edinburgh I have managed to find incredible people who have inspired me (or forced me at times) to become a better person. Karim and Hannah McCall: thank you comrades for sharing your love with me, celebrating our victories, and creating such a caring, loving space. Adam, thank you for making life more bearable by having coffee with me every single Saturday. Nadine, thank you for rescuing me from the office and from myself on numerous occasions, always using food as an excuse. Thanks to Olivia for our Monday lunches (and Josh for being an excellent baker!), for sharing your wisdom with me, and generally, for being such a role model. Bérengère, thank you for the training sessions, the 101 Bakery sessions, the fancy events sessions, the writing sessions, the movie sessions, the we-are-looking-for-a-place-to-work-from sessions, and most importantly: for (your) cooking-daal and Dishoom sessions. A

la pesada de Yolanda: mazo gracias por los viajecitos en el Cli-Yoli, pasar la cuarentena conmigo viendo la serie de Berto, y por siempre traerme membrillito. I believe I speak on behalf of the Clerk St. gang when I thank whoever ordered a Chris Hemsworth calendar for Christmas that got sent to our flat by mistake and never came to pick it up - he has looked over us from the kitchen wall and made our morning coffees much more interesting. My deepest gratitude also goes to Sara, Niklas, Nahuel, Susan, Fra, and Idris. Last, but not least, I would like to thank the Langendam-McLurg family: Thanks to Abi, Floris, and particularly, thanks to Bear and Saskia for five amazing years of Thursdays filled with adventures!

A huge part of my friends have put up with me remotely. Starting with Martin Skala, who has dealt with me since our Erasmus days. Look at me, I'm going to be Dr. Calimero! This will not stop me from being the angsty teenager you know, but now I'm also going to be super entitled and ask you to only refer to me as Doctor. I think this sets the score as Calimero 100000000000 – Martin 1 - checkmate, hon. To spare me from embarrassment, here's a bit of code-switching to thank all my friends from home. Rocío, gracias por seguir recogiéndome en el banquito de la biblioteca, repetirme las tonterías que digo, pedir más tintos de la cuenta, y sobre todo: ayudarme a elegir un tono de rojo para teñirme, y luego reirte porque no ha salido como quería. También le tengo que dar las gracias a Sergio: No quepo en mí para ver todo lo que vas a llegar a ser, garbancito. Víctor, siempre me echas la bronca, me pones las pilas, me cuentas cosas graciosas, me das de comer, me coges el teléfono y me llevas a desayunar. Creo que los que te rodeamos sobrevivimos gracias a ti. Raquel, gracias por siempre hacer planes cuando estoy, por contarme todos los chismorreos, por escucharme pacientemente, y siempre compartir tu entusiasmo por las cosas. En la misma línea, Pabs, gracias por ser la Mejor Persona Del Mundo, por llenarme el móvil de memes y mensajes de voz donde solo dices *'eeeh'* (y también por entender que no voy a escuchar tus audios pero que te voy a obligar a que escuches los míos donde sólo digo *'uuum'*). Marcos, gracias por siempre coger el teléfono, conocer y respetar mis tiempos, y explicarme las cosas de física que no entiendo. Almu: Gracias por venir a verme, por hacer que vaya, por coordinar tu agenda para que pueda

ocupar tu casa cuando vuelvo a Madrid, por hacer de psicóloga, y por todo lo que esta por venir. Carmen: Gracias por guardarme un trocito de pastel, por los mensajes aleatorios, por los viajes, por las canciones, y tu compasión. Tengo mucho que aprender de vosotras dos. Gracias también a Héctor y Jaime, mis lairds, por guardarme siempre una silla en vuestra mesa; a Bea(triz), a José David, a Indre (o Indruten), a Miguel S., Helena, y Pablo con uve. Volver a Madrid se me hace menos bola gracias a vosotros. Finalmente, gracias a Miguel Ángel Castellanos: Gracias por siempre darme muy malos consejos, echarme la bronca, preguntarme por la tesis cuando ya te he dicho que no quiero hablar de eso, ser el mayor marujo del mundo, y por siempre responderme con mas sensatez de la que estoy dispuesta a tolerar. Te debo muchos vermús.

Los últimos agradecimientos son para los que han hecho que yo esté aquí. Espero que este trabajo, que seguro ninguno de vosotros va a leer, os haga sentir orgullosos de mí (o al menos os compense por todos los esfuerzos que habéis hecho individualmente y en grupo). Mamá, gracias por recordarme que tengo que ser mas comprensiva conmigo misma, por llevarme a comprar libros, y haber alimentado mi curiosidad desde siempre (aunque al final no haya acabado siendo arqueóloga, o jueza, o arquitecta). Tío, gracias por ser la voz de la razón, ponerte estricto cuando ves que me voy por las ramas, por tus momentos pedantes y por hacer de padre cuando no te tocaba. Tía, gracias por siempre chinchar al tío y a los primos conmigo, darme consejos de ropa, y recibirme en vuestra casa cuando lo he necesitado. Santiago y Héctor, sé que soy una cafre que no contesta al móvil, pero prometo compensaros llevándoos de cerves: Gracias por siempre contestar a mis stories de Instagram (aunque sea para meteros conmigo). Aunque no os lo diga, estoy muy orgullosa de vosotros y de las personas en las que os estáis convirtiendo. Finalmente, gracias a mis abuelos, Esperanza y Antonio: me gusta imaginarme que estaríais orgullosos de mí (y que me llevaríais de vermús y me compraríais un helado para celebrarlo).

As the saying goes, it does take a village. So long, and thanks for all the fish!

Contents

Abstract **i**

Lay Summary **ii**

Acknowledgements **v**

1 Introduction **1**

1.1 Thesis overview 5

2 What we know about disfluencies **9**

2.1 What are disfluencies? 9

2.2 When are disfluencies produced? 12

2.2.1 The fillers-as-symptoms account 13

2.2.2 The fillers-as-signals account 16

2.2.3 Interim discussion 20

2.3 How can filled pauses affect speech comprehension? 22

2.3.1 Methodology 23

2.4 Conclusion 27

3 Prediction in speech comprehension **30**

3.1 Predictive processing in spoken language: The role of manner of delivery 33

3.1.1 Disfluency in predictive processing 34

3.2 Accounts of prediction in language comprehension 44

3.2.1 Speaker's linguistic background: Non-native speakers 47

3.2.2 Listener's linguistic background: Non-native comprehenders . . . 51
3.3 Conclusion and Chapter aims . . . 54

4 Disfluency as Difficulty **56**
4.1 The disfluency bias . . . 58
4.2 Experiment 1: The role of speaker identity . . . 63
4.2.1 Methods . . . 64
4.2.2 Results . . . 68
4.2.3 Discussion . . . 84
4.3 Experiment 2: The role of listener identity . . . 89
4.3.1 Methods . . . 93
4.3.2 Results . . . 96
4.3.3 Discussion . . . 106
4.4 Comparison across populations . . . 111
4.4.1 Eye movements: Prediction Time window . . . 111
4.4.2 Eye movements: Word recognition . . . 111
4.5 Identification task . . . 113
4.6 General Discussion . . . 116
4.6.1 The disfluency bias . . . 118
4.7 Conclusion . . . 123

5 Meaning interpretation in speech comprehension **127**
5.1 Meaning interpretation in spoken language: The role of manner of delivery as (dis)fluency . . . 129
5.1.1 The case of deception . . . 132
5.2 Accounts of meaning interpretation in language comprehension . . . 134
5.2.1 Speaker's linguistic background: Non-native speakers . . . 139
5.2.2 Listener's linguistic background: Non-native comprehenders . . . 141
5.3 Conclusion and Chapter aims . . . 144

6 Disfluency as Deception **147**

6.1 The disfluency-as-deception bias . . . 149
6.1.1 Deception bias as an association . . . 150
6.1.2 Deception bias as an inference . . . 151
6.2 Experiment 3: The role of speaker identity . . . 153
6.2.1 Methods . . . 155
6.2.2 Results . . . 159
6.2.3 Discussion . . . 167
6.3 Experiment 4: The role of listener identity . . . 170
6.3.1 Methods . . . 172
6.3.2 Results . . . 176
6.3.3 Discussion . . . 182
6.4 Comparison across populations . . . 185
6.5 Non-linear time course of eye movements . . . 186
6.6 General Discussion . . . 190
6.6.1 The disfluency-as-deception bias . . . 192
6.6.2 Biases against non-native speakers . . . 196
6.7 Conclusion . . . 198

7 General Discussion **199**
7.1 Disfluency as difficulty . . . 201
7.2 Disfluency as deception . . . 204
7.3 Accounts for comprehending disfluent speech . . . 205
7.4 Limitations and future directions . . . 207
7.5 Conclusion . . . 209

References **211**

Appendices **253**

A Supplementary materials Experiments 1 and 2 **254**
A.1 Experiment 2 participants' linguistic profile . . . 254

A.2 Visualization of eye-movements 257
A.2.1 Experiment 1: Visualization of raw probabilities in the integration time window 257
A.2.2 Experiment 2: Visualization of raw probabilities in the prediction time window 259
A.2.3 Experiment 2: Visualization of raw probabilities in the integration time window 260
A.3 Comparison between populations 261
A.3.1 Eye movements: Prediction time window 261
A.3.2 Eye movements: Word recognition 263

B Supplementary materials Experiments 3 and 4 265

B.1 Experiment 3: Visualization of the referent disadvantage 265
B.2 Experiment 4 participants' linguistic profile 267
B.3 Experiment 4: Visualization of the referent disadvantage 268
B.4 Generalised Additive Mixed Models 270
B.4.1 Model results 272

List of Figures

2.1 Visual display of a trial in the Visual World Paradigm, from Cooper (1974). 24

4.1 Trial sequence of Experiments 1 and 2. 68
4.2 Pattern of fixations for fluent utterances in Experiment 1. 72
4.3 Pattern of fixations for disfluent utterances in Experiment 1. 72
4.4 Time course of the low-frequency advantage in fluent utterances in Experiment 1. 76
4.5 Time course of the low-frequency advantage in disfluent utterances in Experiment 1. 76
4.6 Time course of the target advantage for high- and low-frequency targets post-target onset in Experiment 1. 83
4.7 Predicted pattern of fixations of the target advantage post-target onset in Experiment 1. 84
4.8 Time course of the low-frequency advantage in fluent utterances in Experiment 2. 99
4.9 Time course of the low-frequency advantage in disfluent utterances in Experiment 2. 99
4.10 Pattern of the target advantage for high- and low-frequency targets post-target onset in Experiment 2. 103
4.11 Predicted pattern of fixations of the preference to fixate on the target post-target onset in Experiment 2. 104

6.1 Trial sequence of Experiments 3 and 4. 159

6.2 Pattern of fixations for fluent utterances in Experiment 3. 163
6.3 Pattern of fixations for disfluent utterances in Experiment 3. 164
6.4 Pattern of fixations for fluent and disfluent utterances in the first and third epochs of Experiment 3. 167
6.5 Pattern of fixations for fluent utterances in Experiment 4. 179
6.6 Pattern of fixations for disfluent utterances in Experiment 4. 180
6.7 Pattern of fixations for fluent and disfluent utterances in the first and third epochs of Experiment 4. 181
6.8 Predicted non-linear referent disadvantage in Experiment 3. 188
6.9 Predicted non-linear referent disadvantage in Experiment 4. 189
6.10 Difference of the differences between fluent and disfluent utterances between speakers for Experiment 3 and Experiment 4. 190

A.1 Pattern of fixations on high- and low-frequency targets post-target onset in Experiment 1. 257
A.2 Pattern of fixations for fluent utterances in Experiment 2. 259
A.3 Pattern of fixations for disfluent utterances in Experiment 2. 259
A.4 Pattern of fixations on high- and low-frequency targets post-target onset in Experiment 2. 260

B.1 Pattern of the referent disadvantage for fluent utterances in Experiment 3. 266
B.2 Pattern of the referent disadvantage for disfluent utterances in Experiment 3. 267
B.3 Pattern of the referent disadvantage for fluent utterances in Experiment 4. 269
B.4 Pattern of the referent disadvantage for disfluent utterances in Experiment 4. 270

List of Tables

1.1 Overview of the experimental work of this thesis. 6

2.1 Taxonomy of disfluencies, adapted from Shriberg (2001). 10

4.1 Duration (in ms) and pitch (in Hz) of the auditory stimuli in Experiments 1 and 2. 66

4.2 Mean reaction times (ms) in Experiment 1. 69

4.3 Estimated parameters of two generalised linear mixed models of looks to low-frequency objects in fluent and disfluent utterances in Experiment 1. 74

4.4 Estimated parameters of two linear mixed models of the low-frequency advantage in fluent and disfluent utterances in Experiment 1. 77

4.5 Estimated parameters of a linear mixed model of the target advantage post-target onset in Experiment 1. 82

4.6 Experiment 2 participants' self-reported English proficiency, LexTale scores and demographic information. 94

4.7 Mean reaction times (in ms) in Experiment 2. 97

4.8 Estimated parameters of two linear mixed models of the low-frequency advantage in fluent and disfluent utterances in Experiment 2. 100

4.9 Estimated parameters of a linear mixed model of the target advantage post-target onset in Experiment 2. 102

4.10 Mean accuracy of participants' recognition of fluent and disfluent utterances. 115

6.1 Breakdown of sentence types for filler trials for the native and non-native speaker condition. 157
6.2 Proportion of object clicked by speaker's linguistic background and manner of delivery in Experiment 3. 161
6.3 Experiments 4 participants' self-reported English proficiency, LexTale scores and demographic information. 174
6.4 Proportion of object clicked by speaker's linguistic background and manner of delivery in Experiment 3. 177

A.1 Experiment 2 participants' answers to the Language Experience and Proficiency Questionnaire (LEAP-Q) 255
A.2 Experiment 2 participants' reported use of English. 256
A.3 Estimated parameters of two mixed effects models of the low-frequency target advantage in fluent and disfluent utterances comparing Experiment 1 and Experiment 2. 262
A.4 GCA model fit for the target advantage comparing Experiment 1 and 2. . 264

B.1 Experiment 4 participants' answers to the Language Experience and Proficiency Questionnaire (LEAP-Q) 268
B.2 Experiment 4 participants' reported use of English. 268
B.3 GAMM results for Experiment 3. 273
B.4 GAMM results for Experiment 4. 274

Chapter 1

Introduction

A month before his death, Douglas Adams gave a talk on his book, *Last Chance to See*, an account of a series of journeys in the search of animals close to extinction. This is an excerpt of a transcription of this talk, where he describes the Aye-Aye:

> *The Aye-aye is a very very peculiar animal. It looks like the agglomeration of all sorts of other different animals. So, for instance, it has a sort of foxy ears, and it has a little sort of bitty rabbit's teeth, and it has a kind of ostrich feathered tail, and it has very weird eyes, actually it has Marty Feldman's eyes. The kind of sort of looking slightly beyond you into a sort of other dimension just over your left shoulder. But it also has one very very very peculiar characteristic, which is its middle finger on both hands is skeletally thin and very very long.*

Although most of us would agree that Douglas Adams was a great writer and speaker, this transcription is not an accurate representation of his talk. A more accurate representation of how this talk was delivered would be this:

> ***Um, uh, buuut*** *the Aye-aye is a very very peculiar animal. It looks like the agglomeration of all sorts of other different animals. So, for instance,* ***it has [silence] um, it has*** *a sort of foxy ears,* ***aaand*** *it has a little sort of bitty*

> *rabbit's teeth,* ***aaand*** *it has* ***um*** *a kind of ostrich* ***tail uh ostrich feathered*** *tail,* ***um*** *and it has,* ***it-it's-it has*** *very weird eyes, actually* ***er*** *it has Marty Feldman's eyes.* ***Uh****, the kind of sort of looking slightly beyond you into a sort of other dimension just over your left shoulder.* ***Um and um*** *but it also has one very very very peculiar characteristic, which is its middle finger on both hands is skeletally thin and very very long.*

The comparison between these two transcripts reveals one important characteristic of spoken language: It contains more than just words. Douglas Adams' speech, as everyone's, contained additional elements: He repeated (e.g., *it-it's*) and elongated words (e.g., *aaand*), corrected himself (e.g., *kind of ostrich tail, uh, ostrich feathered tail*), and filled pauses with *um* and *uh*. These phenomena, referred to as *disfluencies*, have been previously defined as "interruptions to the flow of speech that do not add propositional content to an utterance" (Fox Tree, 1995, p. 709). Data from corpus studies suggest that disfluencies are a standard feature of spoken language: Speakers average between seven to 15 disfluencies per every hundred words produced (Shriberg, 1996). Given their ubiquity in speech, the last decades of research in psycholinguistics have started to explore how the presence of these elements affects the comprehension of speech (Shriberg, 2001).

This thesis explores one type of disfluencies: Filled pauses, such as *uh* or *um* in English. In particular, the work described here focuses on how disfluent speech is processed and interpreted[1]. To date, research on listeners' comprehension of speech riddled with filled pauses has largely explored how their presence affects the processes underlying speech comprehension. This line of experimental work has shown that the presence of filled pauses can have consequences for semantic prediction (e.g., Arnold et al., 2004; Arnold et al., 2007; Corley et al., 2007), syntactic parsing (Bailey & Ferreira, 2003, 2005;, Ferreira et al., 2004), or even the representation of discourse (Cevasco & van den Broek, 2016). More recent studies have started to explore whether and how disfluent speech is interpreted differently than when it is produced fluently. This line of research

[1]Note that the work presented here revolves around non-pathological disfluent speech, as opposed to that produced by people who stutter (PWS).

has shown that filled pauses have consequences for message representation (Corley et al., 2007; Diachek & Brown-Schmidt, 2022; Fraundorf & Watson, 2011), and for message and speaker evaluation: For example, disfluent speech is more likely to be taken as deceptive (Loy et al., 2017; see also King et al., 2018; Li et al., 2022), as a speaker's attempt to save face (Loy et al., 2019), or as evidence of speaker's uncertainty (Brennan & Williams, 1995). Although these two streams of research arguably tackle the same phenomenon, there has not been a systematic effort to determine whether the same mechanisms that underlie the processing of disfluent speech explain the resulting interpretation (e.g., by using the same manipulation). Further, most research has been conducted with listeners comprehending their first (or native) language, usually produced by a native speaker of the language (with the exception of Bosker et al., 2014; Bosker et al., 2019; Morin-Lessard & Byers-Heinlein, 2019; Watanabe et al., 2008).

This thesis steps into this arena to bridge research exploring the comprehension processes and the interpretations of disfluent speech, with a focus on speakers' and listeners' linguistic backgrounds. This thesis aims to explore whether the comprehension of disfluent speech can be understood as the outcome of fast, cost-free mechanisms or as the outcome of relatively slow, cognitively demanding mechanisms. Throughout this thesis, we juxtapose two predictions of these mechanisms: Their automaticity (and thus the speed of their emergence) and their flexibility. We argue that given the wealth of research suggesting that speakers' and listeners' linguistic backgrounds can affect language comprehension (e.g., Foucart et al., 2015; Foucart & Hartsuiker, 2021; Grey & van Hell, 2017; Grey et al., 2019; Lev-Ari, 2015), these factors (i.e., differences that could arise in second-language (or non-native) comprehension, or due to non-native accents) should be considered to develop a better understanding of how disfluent speech is comprehended.

Regarding listeners' linguistic background, as we will review in further detail sections 3.2.2 and 5.2.2 of this thesis, the explanations put forward for previously reported differences between first- and second-language comprehension tackle the properties of the two mechanisms discussed in this thesis. Specifically, we exploit the differences in

cognitive load between first- and second-language comprehension. Our argument is that relatively automatic and cost-free mechanisms should have similar time courses, and if such a mechanism is to underlie the comprehension of disfluent speech, then we should find no differences between first- and second-language comprehenders. If, on the contrary, the comprehension of disfluent speech relies on more cognitively costly mechanisms, then second-language listeners should exhibit a delay in the emergence of previously reported biases upon encountering a filled pause. It is important to note that in this thesis we are thus focused on *late bilinguals*: Individuals who acquired an additional language at or after the age of six, and usually in circumstances different from those of people who acquired a language from birth (e.g., via formal instruction). Throughout this thesis, we will use the native/non-native labels to emphasize that participants were attending to speech produced by an individual who is a late bilingual, or that they were comprehending a language that is not their first, or mother tongue. We, however, acknowledge that the use of these labels within psycholinguistic research is problematic due to the ambiguity of what is a (non)-native speaker (see Cheng et al., 2021). To both ensure that participants were rightly categorised as belonging to either and that we did not exclude minoritized populations, we gathered a set of measures such as age of acquisition, country of origin, mode of acquisition, and language use.

In sections 3.2.1 and 5.2.1 we will review how speaker identity is a factor that can likewise guide speech comprehension. In this case, we exploit the different stereotypes associated with first- and second-language speakers, whereby the latter are believed to be less reliable speakers compared to the former and, specifically, may be expected to produce speech that contains more disfluencies. We argue that an automatic and cost-free mechanism comes at the expense of inflexibility, so that disfluencies produced by either speaker should yield similar biases, and in a similar time course. Cognitively-demanding mechanisms, however, offer flexibility: In this case, listeners should exhibit different biases, and different time courses, depending on who produces the disfluency. It is important to note that by social effects we are concerned with the inferences made about a speaker based on group membership, as opposed to effects that are speaker-specific (e.g.,

shared common ground due to discourse history). In the former case, individuals rely on the information they have (e.g., beliefs) about a speaker's membership to build an internal model of the speaker, which would be applied to any speaker perceived to belong to the same category. In the latter case, individuals rely on information they have about a specific speaker, which would not be transferred to another speaker from the same category.

Given growing concerns in the field of psychology about the replicability of findings, the experiments herein described are replications and extensions of two studies that tackled each of these streams (i.e., Bosker et al., 2014; Loy et al., 2017, respectively). The experiments described in this thesis focus on the effects of filled pauses on the (1) prediction of upcoming items in speech and (2) interpretation of deceit. These experiments tested comprehenders attending to speech in their first- or second-language, produced by either a first- or a second-language speaker. Due to the characterisation of the aforementioned mechanisms, these experiments explore the time course of disfluent speech comprehension by using the eye-tracking technique. It is important to note that our choice of exploring predictive processing and meaning interpretation in relation to disfluent processing is not meant to reflect the divide between the automatic, cost-free, and slow, cognitively demanding mechanisms that this thesis explores. Rather, either process can be characterised by either mechanism (e.g., predictive processing could comprise a cognitively costly component, and interpreting meaning could occur rather fast in a context; see, for example Ito & Pickering, 2021, for a discussion of automaticity in prediction).

1.1 Thesis overview

Chapter 2 provides an introduction to research in filled pauses. It includes an overview of the filled pauses literature by discussing their production: It reviews the production of filled pauses in terms of the language production architecture and the conditions in which they typically occur. This discussion highlights two properties of filled pauses that

can subsequently guide comprehension of disfluent speech: Their distributional patterns, and the causes for a speaker to be disfluent. We will then discuss what methodologies can help us explore the comprehension of disfluent speech. Given the two streams that this thesis explores, this thesis is divided into two parts as depicted in Table 1.1.[2]

Table 1.1

Overview of the experimental work of this thesis.

		Part I: Processes Lexical prediction, Bosker et al. (2014)				Part II: Interpretation Interpretation of deceit, Loy et al. (2017)			
		Speaker					Speaker		
		Native	Non-native				Native	Non-native	
Listener	Native	Experiment 1			Listener	Native	Experiment 3		
	Non-native	Experiment 2				Non-native	Experiment 4		

Part I is devoted to exploring how filled pauses affect one mechanism put forward to explain efficient speech comprehension: prediction. In Chapter 3, we briefly review the evidence suggesting that filled pauses can modulate predictive processes underlying speech comprehension in terms of the properties highlighted in Chapter 2. We then contrast this evidence with the mechanisms put forward to account for prediction in speech comprehension. This comparison will highlight that differences in speakers' and listeners' linguistic backgrounds can affect predictive processing. We will then propose that they can also interact with the effects of disfluency in lexical prediction. We will then evaluate the evidence available for these two factors, a discussion that will set the scene for Chapter 4.

Chapter 4 comprises the first set of experimental work of this thesis. A body of literature suggests that upon encountering a filled pause, comprehenders' expectations about upcoming elements in the signal are biased, what we refer to as the *disfluency bias*. We begin the chapter with a description of two mechanisms that may explain the effects of disfluency in linguistic prediction. In this chapter, the focus is placed on listeners' preference to fixate on objects whose labels are low in frequency over objects whose label is high-frequency (i.e., low- and high-frequency words). Bosker et al. (2014)

[2]Experiments are reported in this order for ease of reading. In fact, they were run in the following order: Experiment 3, Experiment 1, Experiment 2, and Experiment 4, with the last three being conducted simultaneously.

showed that native Dutch listeners displayed anticipatory eye movements towards low-frequency words when encountering a filled pause produced by a native speaker, but this anticipatory behaviour was attenuated when the speaker was producing speech in their second language (i.e., a non-native speaker). Experiment 1 replicates Bosker et al. (2014) in a sample of native English listeners. Experiment 2 extends Experiment 1 to a sample of non-native listeners, to explore the effects of language experience and cognitive load on the disfluency bias. These two experiments showed eased recognition of the low-frequency item when it was preceded by a disfluency. In the case of native listeners, this was regardless of the speaker's linguistic background. However, while non-native listeners exhibit a similar benefit for filled pauses produced by a native speaker, they benefited from the presence of a disfluency when it was produced by a non-native speaker, regardless of the frequency of the object the speaker referred to.

Part II of this thesis explores how filled pauses affect the interpretation of disfluent speech by exploring the interpretation of deceit. Chapter 5 parallels the structure of Chapter 3: It begins by presenting the evidence suggesting that disfluent speech is interpreted differently from when it is delivered fluently, with a particular focus on deceit. We then explore the accounts put forward to explain how listeners interpret meaning. By doing so, we will discuss how interpretations can be subject to different factors, amongst which are speakers' and listeners' linguistic backgrounds. This will be followed by an evaluation of the evidence exploring how these two factors have been shown to modulate what individuals interpret, a discussion that will lead us to Chapter 6.

Chapter 6 comprises the second set of experimental work. This body of work explores how the presence of a filled pause can lead listeners to interpret deceit. The chapter follows a structure akin to Chapter 3: We firstly propose two mechanisms whereby disfluent utterances may bias listeners to interpret deceit, what we refer to as *disfluency-as-deception bias*. Loy et al. (2017) demonstrated that native listeners of English were more likely to interpret disfluent utterances as deceitful, with an early emergence of this interpretation as reflected in participants' eye movements. Experiment 3 replicates and

extends Loy et al. (2017) by presenting native English listeners with disfluent speech produced by either a native or a non-native speaker. Experiment 4 extends this paradigm to a sample of non-native listeners to explore the effects of cognitive load on the disfluency-as-deception bias. The results demonstrate that in scenarios where the costs of misinterpreting a filled pause are high, listeners are insensitive to who produces the disfluency, and this interpretation does not seem to be cognitively costly.

Chapter 7 draws together the findings from both parts. This chapter includes a brief summary of the findings of the experiments described in the thesis, and presents a possible mechanism to account for how disfluent speech is comprehended.

Chapter 2

What we know about disfluencies

This chapter offers an introduction to filled pauses. It first covers what disfluencies are, with a focus on their characteristics in terms of production. From there, it goes into a review of the two camps accounting for the production of disfluencies: the 'fillers-as-symptoms' and the 'fillers-as-signals' accounts. Learning about the characteristics of filled pauses will show what features they have that can subsequently impact speech comprehension. Afterwards, we will discuss what methods can help us investigate whether and how filled pauses impact speech comprehension, by discussing online and offline measures.

2.1 What are disfluencies?

Disfluencies have been defined as "interruptions to the flow of speech that do not add propositional content to an utterance" (Fox Tree, 1995, p.709). Under this umbrella fall several different devices, such as unusual periods of silence (i.e., a silent pause) and those filled in with a sound (i.e., a filled pause). Table 2.1 depicts a taxonomy of disfluencies.

Table 2.1

Taxonomy of disfluencies, adapted from Shriberg (2001). The disfluency is written in italics in each example.

Disfluency type	Definition	Example
Filled pause	Hesitation devices such as *uh*, *er*, *um* in English. Also known as fillers.	*Uh* we live in Dallas.
Silent pause	Unsual periods of silence in speech.	It is *[silence]* a very peculiar idea.
Repetition	Repeated phonemes, words, phrases.	All *the-the* tools.
Deletion	A subtype of repair, where the speaker commits an error, stops, and resumes speaking without going back to the moment before the error was committed.	*It's - I could* get it where I work.
Substitution	A subtype of repair, where a speaker makes a speech error and wishes to change a phoneme, word, or phrase, and so they interrupt speech and resume speaking from the moment before the error occurred.	Any health *cover - Any health insurance.*
Insertion	A subtype of repair, where the speaker makes an error by omitting a phoneme, word, or phrase. The speaker interrupts speaking and resumes it to the moment before it occurred.	And *I felt - I also felt.*
Articulation error	A subtype of repairs, where the speaker mispronounces a phoneme. The speaker interrupts speaking and resumes speech before the error occurred.	And *pin* - pistachio nuts.
Lexical fillers	Words that do not add propositional meaning to the utterance. Sometimes categorised as discourse markers.	I was *like* so shocked.
Prolongation	Syllabic lengthening.	*Aaaaand* it also has a kind of ostrich tail.

This thesis is concerned with *filled pauses*, also referred to as *fillers*: Hesitation phenomena characterised by an interruption and a delay of speech, without the addition of propositional content, followed by a continuation of the utterance, without backtracking to the pre-disfluency utterance, such as in (1):

(1) **Uh**, walk to the corner and then turn to thee **uh** left[1].

The first component of the definition refers to the interruption and subsequent delay of speech. Corpus studies have shown that this delay can occur at the beginning of an utterance (i.e., utterance-initial; '**Uh**, walk to the corner') and in the middle of an utterance (i.e., utterance-medial; 'Turn to thee **uh** left'), although compound languages like Swedish allow for mid-word filled pauses (Eklund & Shriberg, 1998). This interruption has been said to span from 100 ms to 750 ms (Shriberg, 2001), with utterance-initial filled pauses lasting longer than utterance-medial ones (Swerts, 1998).

The second component of the definition refers to the sound that fills the pause. In most languages, a filled pause represents a period of articulation of non-propositional content that fits a language-specific convention (Rose, 2017). For example, French speakers fill their pauses with *eu* or *eh*, while German speakers produce *äh* or *ähm*, although some languages allow for more complex forms and employ lexicalised fillers (e.g., *este*, 'this', in Spanish) (Clark & Fox Tree, 2002). Across languages, filled pauses tend to be built around central vowels and are usually followed by a nasal consonant (Crible et al., 2017). In British English, filled pauses take the form of *uh* and *um* (de Leeuw, 2007), where the quality of the vowel is close to a flat schwa (Shriberg, 2001), and its duration differs between these two forms (*uh* can go up to 1.5 s, longer than the vowel duration of *um*; Hughes et al., 2016).

Further features of filled pauses include their prosodic contours and the elements with which they co-occur. Regarding the former, the prosodic contour of a disfluent utterance has been shown to be a crucial factor to identify filled pauses (Shriberg et al.,

[1]Cases where there is backtracking (e.g., "Turn to the, uh, turn to the left" are outwith the scope of this thesis.

1997). Generally speaking, a filled pause is characterised by a low fundamental frequency, and a falling tone (O'Shaughnessy, 1992). When a filled pause is produced mid-utterance, its fundamental frequency is relative to the prior prosodic context (Shriberg & Lickley, 1993). Finally, filled pauses often co-occur with other disfluencies and hesitation markers: For example, the word prior to the filled pause may be prolonged, and a silent pause may follow them (Grosjean & Deschamps, 1975), and some authors argue that discourse markers occur in their periphery (e.g., "*I mean* I think *uh* space is *you know*", Degand & Gilquin, 2013).

Due to these characteristics, filled pauses separate themselves from other types of disfluencies. In contrast to filled pauses, repair disfluencies include lexical material that the speaker will replace, while silent pauses do not include a sound. This potentially suggests that each disfluency may have different effects on comprehension (see MacGregor, 2008): For example, the inclusion of material that needs to be replaced in repair disfluencies means that these disfluencies are recognised as such because of their context (Corley, 2010). The properties described in this section thus far propose that filled pauses may affect speech comprehension due to the delay they represent in the speech signal. In this following section, we discuss what additional properties they may have that can have consequences for comprehension. We do so by describing the two main accounts put forward to explain why speakers produce disfluencies.

2.2 When are disfluencies produced?

This section describes the two broad camps that have suggested why filled pauses are produced: the 'fillers-as-symptoms' and 'fillers-as-signals' accounts. The former has its origins in the most widely-accepted architecture for language production, and therefore this review also includes a brief description of this model. The 'fillers-as-signals' account takes a radically different view and conceives filled pauses as intentional messages from the speaker, and thus includes a social component. Each section reviews the empirical evidence supporting the claims for each account.

2.2.1 The fillers-as-symptoms account

The 'fillers-as-symptoms' account considers fillers as a by-product of trouble in language production. The different causes for a speaker to experience difficulty when producing speech can be understood under Levelt's (Levelt, 1983, 1989) proposed architecture for language production. For explanatory purposes, let us first describe the basic structure of Levelt's (1983, 1989) model[2].

This model comprises three principal components: the *conceptualiser*, the *formulator*, and the *articulator*. Speech production starts in the conceptualiser: A speaker engages in the conceptual preparation of a given message. This involves conceiving the intention they want to engage in, selecting the relevant information to express it, and ordering how information will be delivered. In this stage, the speaker needs to keep track of the discourse context (e.g., what was said before, with whom they are talking). The outcome of the conceptualisation stage is a *pre-verbal message* that needs to be translated into linguistic units in the formulator. This translation involves accessing and retrieving *lemmas* from the mental lexicon, elements that include the declarative knowledge of the label associated with the concept required, and the syntactical properties of the label. The last stage involves the specification of the phonetic and articulatory commands that are fed to the articulator to produce speech.

This division of labour highlights the different steps in producing speech where speakers may struggle. Levelt's (1983, 1989) model includes a monitoring component whereby speakers can self-monitor speech production. In the early stages, speakers can monitor their internal speech: For example, they can keep track of whether the system is producing the concept they want to express. In the later stages, speakers can hear themselves and monitor their overt speech: For example, they can track the volume with which they speak. In the *perceptual loop theory*, Levelt (1983, 1993) states that this monitor component oversees both covert and overt speech and takes action when it detects an error. Specifically, errors are handled following the *main interruption rule*:

[2]These basic steps are part of other models of speech production such as Dell et al.'s (1997).

'Stop the flow of speech immediately upon detecting trouble' (Nooteboom, 1980). In the case of covert speech, because the error has not been overtly produced, the main interruption rule has consequences for the fluency of speech: A filled pause is likely to occur when an error is encountered in covert speech but a repair is not available at the time of the interruption.

It is worth noting that the perceptual loop theory does not fully account for all the reasons that can lead to disfluent speech. A filled pause (and other hesitation phenomena taken as 'covert repairs') can also occur due to factors other than repairing an error (e.g., processing load) and therefore it is difficult to assess whether an error is actually being repaired at the time a filled pause is produced (Hartsuiker & Kolk, 2001). Further, Levelt (1989) conceptualises disfluency production as a process involving attentional control. Consequently, the fewer attentional resources available to monitor speech, the fewer disfluencies speakers should produce. However, in situations of divided attention, normally fluent speakers produce more filled pauses (Oomen & Postma, 2001). Indeed, several cognitive taxing situations have been shown to increase the rate of filled pauses, such as anxiety (Christenfeld & Creager, 1996). This suggests that filled pauses are automatic responses to problems in speech planning (and not necessarily errors), and that monitoring might not be a crucial component in their production. Filled pauses are better understood as reflecting difficulties in language production, but not necessarily as evidence of errors in covert speech.

Evidence for the presence of filled pauses as a by-product of trouble in the conceptualisation stage is mixed. On the one hand, speakers are more likely to be disfluent when they are describing a topic they are less familiar with (Merlo & Mansur, 2004), arguably because less knowledge hinders the creation of the pre-verbal stage. Similarly, speakers' uncertainty about their knowledge on a given topic increases the rate of filled pauses (Brennan & Williams, 1995; Smith & Clark, 1993; Swerts & Krahmer, 2005). On the other hand, in carefully controlled experiments, this association is harder to find. Experiments exploring the production of filled pauses commonly employ the Network Task,

where participants are asked to describe to an interlocutor a route through a network of images. Researchers manipulate the characteristics of these networks (e.g., the number of names that can be used to refer to one image, i.e., name agreement) to challenge speech production at its different stages. Schnadt and Corley (2006) explored whether difficulties at the conceptualisation stage lead to an increase in the rate of disfluencies. The authors manipulated the ease with which speakers could provide instructions by presenting elements for which the image was blurred or not, under the assumption that blurring would hamper the identification of items and subsequently may pose a problem for the conceptualiser. Schnadt and Corley (2006) did not find an increase in the rate of filled pauses when speakers referred to blurred images as opposed to clear ones, suggesting that difficulties in conceptualising a message do not necessarily lead to an increase in disfluency (see also Pistono & Hartsuiker, 2023, for similar findings).

Problems at the formulation stage seem to increase the rate of filled pauses. Retrieving the appropriate lemma can be difficult when the information from the pre-verbal message activates a cohort of semantically-related lemmas, increasing the competition for selection (Hartsuiker & Kolk, 2001). For example, individuals who speak more than one language experience competition between lemmas, and are consequently more disfluent (Bergmann et al., 2015). This idea would also explain why, for example, lecturers in the arts average more hesitations than those in the sciences (Schachter et al., 1991), as arguably the former allows for more choices than the latter (Schachter et al., 1994): Filled pauses are, thus, a reflection of the act of choice (Beattie & Butterworth, 1979). Indeed, in carefully designed experiments, the rate of disfluency appears to be a function of the number of options available to refer to an entity and the competition between those (see also Pistono et al., 2023; Rapoeye et al., 2022).

Additionally, lemma retrieval is impacted by its accessibility. Goldman-Eisler (1958) reported that words with low transitional probabilities (i.e., less predictable given their preceding context) were associated with a higher rate of disfluencies. Experiments employing the Network Task methodology have shown that elements with low name

agreement and low frequency (which entail lower lemma accessibility) lead to an increase in the production rate of filled pauses (Hartsuiker & Notebaert, 2010; Jaeger et al., 2012; Pistono & Hartsuiker, 2021; Schnadt & Corley, 2006). A lemma's accessibility is said to be lower in second-language production (e.g., weaker link hypothesis, Gollan et al., 2008) and it could partially explain the higher rates of filled pauses in bilinguals (Davies, 2003; De Jong, 2016). Likewise, lemma accessibility lowers with age (e.g., transmission deficit hypothesis, Burke et al., 1991) and indeed, older adults are reportedly more disfluent (Mortensen et al., 2006, but cf. Arslan & Göskun, 2022; Gósy et al., 2014), a pattern similar to that reported in patients with Alzheimer's disease (Rohanian et al., 2021) and in the early stages of Huntington's disease (Tovar et al., 2020).

To summarise, under the 'fillers-as-symptoms' account, the presence of filled pauses can be explained following a three-stage architecture of language production. Different problems at different stages during this process can lead to what is seen as a disfluency. In this sense, there is nothing 'special' about filled pauses: They are likely to be a by-product of difficulties at different levels in speech production. Understanding the production of filled pauses under this light shows one of their properties: They occur in arguably predictable patterns (e.g., before low-frequency words). This, in turn, offers a way for filled pauses to influence speech comprehension: via statistical learning, as we will discuss in section 4.1.

2.2.2 The fillers-as-signals account

A contrasting view to the 'fillers-as-symptoms' account is that filled pauses are signals instead. While a symptom is a 'natural sign without an intervening intention' (Fox Tree & Clark, 1997, p.164), signals are meaningful acts produced by speakers to convey information to their interlocutors (Fox Tree & Clark, 1997). The fillers-as-signals account takes the fillers-as-symptoms proposal one step further by adding a social component to the production of disfluencies: They are intentional signals from the speaker.

This account is framed under the *backchannel hypothesis*. Briefly put, this hypothesis argues that individuals have at their disposal two streams to communicate: (1) the main channel, where they deliver the *primary signal* (i.e., the content), and (2) the secondary channel, where they deliver *collateral signals* (i.e., the displays). While the former channel allows speakers to communicate a message via linguistic devices, the latter represents a set of tools (e.g., verbal responses such as *mhm* or facial expressions, Tolins & Fox Tree, 2016) that allow speakers to coordinate the message delivered via primary signals and to coordinate their communication (Clark, 2006). Via collateral signals, speakers may comment on their performance to attract their interlocutor's attention.

Under the backchannel hypothesis, filled pauses are collateral signals. Specifically, they are *asides*: Expressions speakers use to comment on their performance (Clark, 2004) and their metacognitive state (Smith & Clark, 1993). In particular, fillers serve to communicate that 'At time t(filler) there will be a minor or major delay in speech' (Clark & Fox Tree, 2002) e.g., they can signal 'I am searching for a word' or 'My conversational turn is over' (Clark, 2004), while providing a sense of continuity in the signal (instead of ceasing speech) (Clark, 2002). Clark (2002) opposes the fillers-as-symptoms account by appealing to the properties of filled pauses that cannot be merely understood as a reflection of trouble in speech production alone. The idea that filled pauses have a social component is based on their variance (i.e., the different forms filled pauses can take; Clark & Fox Tree, 2002), speakers' control over their production, and data from corpus studies. Clark (1994) argues that while filled pauses do occur at stages where speakers can encounter problems, disfluency is produced *intentionally* by the speaker to warn their interlocutor that they are experiencing trouble in speech production, i.e., a filled pause is the visible display of a problem the interlocutor does not have access to (Clark, 1997, 2002).

Both primary and collateral signals undergo the same Leveltonian architecture for speech production: They need to be conceptualised, formulated, and articulated. As a consequence, filled pauses are conceptualised as words. For Clark and Fox Tree (2002),

filled pauses are interjections, a type of linguistic item that can constitute utterances on their own and does not form constructions with other word classes (Wilkins, 1992, as cited in Clark & Fox Tree, 2002). Filled pauses thus have the same characteristics as other lexical items (i.e., semantics, syntax, phonology, and prosody). Their basic meaning refers to an upcoming delay at the time they are produced but can adopt an implicated meaning (Clark, 2004). Several authors have endorsed the proposal that filled pauses belong to lexical categories, although different groups have been proposed: For example, Kirjavainen et al. (2022) argue that filled pauses are grammatical clitics, while Tottie (2011) conceptualises them as 'planners', which belong to the same category as discourse markers such as 'well' (cf. Kosmala & Crible, 2022).

There are several predictions from the 'fillers-as-signals' account regarding the production of filled pauses. Firstly, if hesitations are solutions to particular types of problems (Clark, 1994), then there should be variability in the forms filled pauses take depending on the struggle encountered. Clark and Fox Tree (2022) explored the production of filled pauses in the London-Lund corpus and found that the differences in duration between *uh* and *um* correlate with the amount of time it takes for speech to be resumed. The authors argue that speakers thus choose a particular form depending on the time it will take them to solve the trouble they have found, with *um* signalling a longer delay (see Fox Tree & Clark, 1997, for a similar argument regarding the production and elongation of *the*). Although some authors have reported similar findings (e.g., Kendall, 2013), this idea that filled pauses vary in how they are produced has been contested due to how the pause following a filled pause was measured, as it relies on subjective measures of how long the pause was perceived to be (O'Connell & Kowal, 2005).

Secondly, speakers do not produce filled pauses every time they struggle with language production (Clark, 1994). Corpus studies have shown that the production rate of filled pauses depends on the nature of the conversation. Filled pauses are more likely in dialogue than in monologue (Shriberg, 2001), which supports the idea that speakers produce filled pauses to communicate with their interlocutors. However, these results

have not been replicated in carefully controlled experiments (Finlayson & Corley, 2012). Further, Oviatt (1995) reported that while the per-word rate of disfluencies in human-human interactions was 0.06, it dropped to 0.008 for human-computer interactions - a rate than can increase when there is another person in the room (Walker et al., 2014). The relationship between the rate of production of filled pauses and the presence of another person supports the notion that filled pauses are listener-oriented.

Conceptualising fillers as intentional signals predicts that different communicative needs in a conversation can impact their rate of production. Maclay and Osgood (1959) found that filled pauses were more likely to be produced at syntactic junctures, i.e., between two separate utterances. Maclay and Osgood (1959) argue that speakers fill their pauses between utterances because they need to signal that they wish to hold the floor, contrary to mid-utterance silences, where the interlocutor is aware that the speaker has not finished talking. This parallels the finding that filled pauses characterise the beginning of a turn (i.e., a stalling act, Stenstrom, 2014): Speakers communicate that they are deciding what to say, but that the floor is theirs (Shriberg, 2001). In this sense, filled pauses can help interlocutors manage turn-taking (Kjellmer, 2003). Indeed, speakers produce more filled pauses when they cannot see their interlocutor (Belz & Reichel, 2015; Kasl & Mahl, 1965): As visual presence smooths conversation, speakers may compensate for the lack of this channel by increasing their use of verbal collateral signs.

Collateral signals can also help the speaker present themselves to their interlocutor in a particular manner. Smith and Clark (1993) argue that commenting on one's meta-cognitive state can aid speakers' self-presentation in situations when they are uncertain about their knowledge. When answering a question for which they do not know the answer, or they feel unsure about whether their answer is correct, speakers produce more filled pauses than when they are sure they do not know the answer, or they are uncertain about their knowledge (*Feeling of One's Knowing*, FOK, Brennan & Williams, 1995). Saving face can also explain why filled pauses may accompany dispreferred responses. For example, declining an invitation may have potential underiable social effects. To

compensate, speakers accompany the linguistic expression of the refusal with filled pauses and other markers, such as 'well' (Kirjavainen et al., 2022; Rose, 1998). Conversely, in scenarios where speakers do not want to display their meta-cognitive state, they may reduce the amount of collateral signals they produce. For example, speakers are less disfluent when they are being deceptive (Loy et al., 2018).

Finally, if filled pauses are socially-oriented, then they should be subject to individual differences in aspects pertaining to social cognition. For example, there are individual differences that make speakers more prone to be disfluent (Branigan et al., 1999) such as gender, social class, or personality traits such as extraversion (Dewaele & Furnham, 1999; Tottie, 2011, 2019) - traits that have been shown to impact individuals' social orientation. Likewise, individuals on the Autistic Spectrum tend to produce fewer filled pauses compared to individuals who are not on it (Bellinghausen et al., 2020; Gorman et al., 2016; Lake, 2010; MacFarlane et al., 2017), arguably because individuals on the spectrum may differ from neurotypical individuals in how they use social cues.

To sum up, the 'fillers-as-signals' account adds a social dimension to the production of filled pauses. Under this view, filled pauses are intentionally produced signals from the speaker to comment on their linguistic performance and their meta-cognitive state. Specifically, filled pauses take the basic meaning 'there will be a delay in the signal', which can then be further refined via implicatures. The addition of this social component highlights another aspect of filled pauses: They offer a window to the speaker's mental state, to which listeners may be tuned and thus guide their comprehension following the perceived meta-cognitive state of their interlocutor.

2.2.3 Interim discussion

Filled pauses have characteristics that distinguish them from other kinds of hesitation phenomena. They are interruptions that delay speech, which speakers fill with sounds that commonly involve a vowel but that do not carry propositional content. Their production can be understood as a symptom of struggle in the production of speech. In

an extreme position, they are intentional signals from the speaker to comment on their linguistic performance.

Across experiments and corpus analyses, filled pauses have been shown to occur in predictable patterns. This is in line with a three-stage architecture with a monitoring component. Specifically, data suggests that filled pauses are likely to emerge due to difficulties in formulating the linguistic message. Further, the production of filled pauses also seems to fluctuate as a function of the conversational context: Filled pauses can take different forms depending on the length of the delay in speech, and they are more likely to arise when speakers interact with another human. Therefore, the production of filled pauses can be partially explained by conversational needs: For example, they are more likely to occur when speakers wish to hold the floor and when they want to express uncertainty about the knowledge conveyed in the utterance. This suggests that there is evidence for both the 'fillers-as-symptoms' and the 'fillers-as-signals' accounts.

As Corley and Stewart (2008) propose, while there is a strong correlation between the production of filled pauses and a speaker's cognitive state, filled pauses themselves do not need to be intentional signals. For example, the distributional patterns of filled pauses do not necessarily suggest that these are points in which a speaker may want to signal they are struggling; that is, they can signal trouble, but without speakers intentionally producing them. Crible et al. (2017) argued that the division between these two accounts is a consequence of regarding disfluency in general under sequential (i.e., patterns of combination), situational (i.e., social and contextual norms), and ambivalent (i.e., situation-specific) factors. A better account for filled pauses should conceptualise them as being on a continuum between these two categories, i.e., sometimes produced unintentionally, and sometimes intentionally. In some scenarios, filled pauses can be better explained as a symptom, where they may arise due issues with to lexical retrieval (and therefore likely to co-occur with other hesitation phenomena). In other circumstances, filled pauses can be signals, likely to arise at syntactic boundaries (and to co-occur with discourse markers) (Kosmala & Crible, 2022).

2.3 How can filled pauses affect speech comprehension?

Thus far, we have reviewed the accounts proposed for the production of filled pauses and the evidence in support of each. Regardless of whether they are symptoms or signals, their ubiquity in speech and the different factors affecting their production rate (e.g., problems in production, communicative contexts) may affect how their presence affects disfluent speech comprehension.

Originally, filled pauses were considered to be treated as 'noise' by the comprehension system (Martin & Strange, 1968). Under this account, the comprehension system filters disfluencies out. This viewpoint was further justified by evidence showing that listeners do not exhibit good recall for the position of filled pauses in utterances just heard (Duez, 1985; Lickley & Bard, 1998; Lindsay & O'Connell, 1995). However, these findings are hard to reconcile with evidence showing that speech peppered with filled pauses affects comprehension. For example, words preceded by a filled pause decrease listeners' reaction times when selecting the referent (Fox Tree, 2002), and listeners' evaluations of disfluent speech differ from those of fluent speech (e.g., perception of dishonesty, Fox Tree, 2002). These differences suggest that, instead of being filtered out, the comprehension system perceives and integrates them. The advancement of new methodologies such as eye-tracking have further shown that the presence of filled pauses affects the online processing of speech: For example, listeners display anticipatory eye movements towards certain elements (over others) if the speaker is disfluent (Arnold et al., 2004, 2007; Bosker et al., 2014; Bosker et al., 2019; Heller et al., 2015).

In this thesis, we are interested in the effects disfluencies may have on comprehension from two perspectives: The *online processes* that take place while speech is being comprehended, and the *interpretations derived* from speech, i.e., what happens while listeners attend to disfluent speech and its outcome. These two components are explored in this thesis by using the prediction of upcoming elements in speech and the interpretation of deceit following disfluent utterances. In what follows, we briefly review how the methodology employed in this thesis can help us explore these two perspectives.

2.3.1 Methodology

Online measures: The Visual World Paradigm

One method to explore the online processing of speech comprehension involves measuring listeners' eye movements while they inspect a visual scene and attend to speech. This paradigm is known as the *Visual World Paradigm* (henceforth, VWP; Cooper, 1974; Tanenhaus et al., 1995). Participants are presented with a set of visual stimuli while they are listening to speech under the assumption that speech comprehension shifts participants' attention around the visual input (Magnuson, 2019).

Cooper (1974) first demonstrated how eye movements over a visual scene are guided by accompanying speech. Figure 2.1 depicts the visual display of an example trial in Cooper (1974). Listeners were presented that scene while attending to a narrative (e.g., 'While on a safari in Africa [...] I noticed a hungry lion slowly moving through the tall grass toward a herd of a grazing zebra'). Upon encountering the word *Africa*, participants' fixations were biased: They were more likely to fixate on elements such as the lion, the snake, and the zebra. This showed that listeners tend to look at visual elements related to the linguistic information they are processing in real-time. As the narration continued, participants tended to fixate on the elements mentioned (e.g., the lion) shortly after they were produced. Cooper's (1974) methods were then popularised by Tanenhaus et al. (1995) and have been used since to investigate different levels of language comprehension (Huettig et al., 2011) and with different populations (e.g., children, Trueswell, 2008). For example, the VWP has been used to investigate phonological processing (Allopenna et al., 1998), semantic prediction (Altmann & Kamide, 2007), pre-activation of shape features of predicted upcoming linguistic elements (Rommers et al., 2013) and pragmatic comprehension (Grodner et al., 2010; Kurumada et al., 2014). Research suggests that eye movements are oriented towards named objects as soon as 100 ms (Altmann, 2011) or 200 ms (Matin et al., 1993) after the spoken name becomes unambiguously identifiable.

A typical VWP display contains at least two objects. This display is usually presented on its own for a brief period of time (the *preview* window), and then the auditory

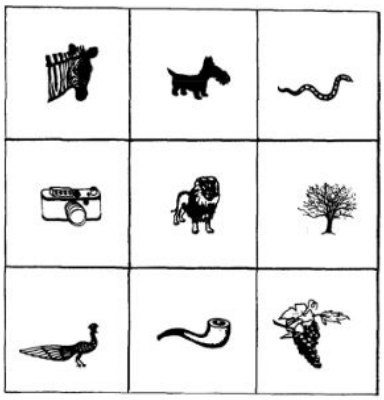

Figure 2.1. Visual display of a trial in the Visual World Paradigm, from Cooper (1974).

stimulus is presented. The auditory stimulus commonly refers to one of them (i.e., the *target* or *referent*), and the unmentioned element serves as a *distractor*. Researchers can be interested, for example, in knowing when fixations are oriented towards the item being mentioned, or how the differences (or similarities) between the referent and the distractor can delay participants' fixations towards the referent. The VWP is therefore a time-locked measurement suitable for investigating speech comprehension as utterances unfold (Ferreira et al., 2013; Huettig et al., 2011; Tanenhaus, 2007), with language-mediated eye movements being largely automatic and unintentional (Mishra et al., 2013).

Importantly to this thesis's aims, the VWP can also aid in understanding how disfluencies can guide comprehension. Arnold et al. (2007) capitalised on the fact that filled pauses commonly precede unfamiliar, and thus harder-to-name objects. In their study, participants were presented with a four-image array, where two items were familiar objects (e.g., ice cones) and two were unfamiliar objects (e.g., squiggly shapes), with each familiar-unfamiliar pair matched in colour (red and black). By having this manipulation, the authors could explore whether the disfluency biased participants towards an element before it is disambiguated. Participants were asked to follow auditory instructions, which could be delivered fluently or disfluently (*Click on the red* [item] versus *Click on* ***thee uh*** *red* [item]). Upon encountering filled pauses, participants fixated on unfamiliar objects before their names were uttered, suggesting that they predicted the upcoming referent based on the manner of delivery.

The VWP has also been previously employed to investigate the time course of interpretations triggered by disfluent speech. Loy et al. (2017) exploited the stereotype individuals hold about the sound of deceit, whereby liars are thought to produce more filled pauses (Zuckerman et al., 1981). Participants saw a two-image visual array while listening to a potentially deceitful speaker refer to either object as the potential location of a treasure. Participants were asked to select the location where they thought the treasure actually was, with the speaker's speech as the only information upon which they could make their decision. Importantly, the speaker would sometimes produce fluent or disfluent instructions (*The treasure is behind the* [item] versus *The treasure is behind* ***thee uh*** [item]). Participants' fixations towards the unmentioned elements increased shortly after encountering the disfluency, and they were more likely to ultimately choose them as the location of the treasure i.e., disfluent utterances were interpreted as deceptive. The time course of fixations demonstrated that the ultimate interpretation of a disfluent utterance in a given context emerges early while comprehending speech.

These two experiments suggest that the VWP is an appropriate technique to investigate how disfluencies affect speech comprehension. However, it is important to note that the VWP paradigm is not without caveats. As discussed in Huettig et al. (2011), the presentation of visual stimuli alongside speech can further impact predictive processing by activating the linguistic items corresponding to the images presented, which in turn can confound pre-activation due to auditory stimuli with that of visual stimuli (although cf. Rommers et al., 2013, for the effects of preview time). How eye movements are analysed can also be the subject of critiques, starting from how to represent eye movements (e.g., as binary or proportional data; Barr, 2008; Donelly & Verkuilen, 2017; Jaeger, 2008) to the statistical approach used (see Ito & Knoeferle, 2022; Stone et al., 2021). Importantly, different research questions would require different analyses (Ito & Knoeferle, 2022).

In this thesis, we are interested in the biases exerted by disfluent speech. We understand a bias as a preference towards an element which comes at the expense of a

dispreference towards another. Therefore, throughout this thesis, we will measure eye movements via empirical logits (Barr, 2008). Further, eye movements rarely follow a linear pattern (Porretta et al., 2018). The questions of interest in this thesis in fact revolve around non-linearity, by asking the rate of growth of a preference to fixate on a particular element (Experiments 1 and 2), and when the preference to fixate on a particular object emerges (Experiments 3 and 4). In this thesis, we accounted for the non-linearity of eye movements over time presenting additional analysis using Growth Curve Analysis (Mirman et al., 2008; Mirman, 2017) in Chapter 4 and Generalised Additive Mixed Models (Wood, 2006, 2011; see also Wieling, 2018) in Chapter 6.

Offline measures: Participants' choices

To investigate how speech is represented, researchers have commonly relied on offline measures, such as comprehension questions. These measures offer a window on speech comprehension and commonly include asking participants how grammatical a sentence was (Ferreira & Yang, 2019), what event was conveyed by a sentence (e.g., Cai et al., 2022; Gibson et al., 2013), or what words they can recall from the experiment (e.g., Corley et al., 2007). Likewise, these measures can show under what conditions listeners believe the speaker to be sarcastic (Bazzi et al., 2022; Caffarra et al., 2018; Woodland & Voyer, 2011), or how listeners evaluate the speaker (Hendriks et al., 2017; Lindemann, 2005; Schüppert et al., 2015).

Offline measures have been particularly useful to deepen our understanding of how disfluent speech is represented. Bailey and Ferreira (2003) found that the presence of a filled pause in an otherwise garden path sentence affected participants' grammaticality ratings (Exp. 1, although this could be explained by the increased distance until the disambiguation point, Exp. 2), and importantly, the presence of a filled pause in unusual positions decreased grammaticality ratings (Exp. 3). Likewise, listeners remember better the details of a story if they are preceded by disfluencies (Fraundorf & Watson, 2011), and the presence of a filled pause boosts the memory of the words said (Corley et al., 2007; Diachek & Brown-Schmidt, 2022).

Crucial for Part II of this thesis is the fact that disfluent speech can yield a different interpretation from that triggered by fluent speech. Listeners ascribe less certainty about knowledge of a given topic to a speaker if they are disfluent (Brennan & Williams, 1995), rate disfluent speakers as less intelligent (Christenfeld, 1995), and, crucially, disfluent utterances are more likely to be interpreted as deceitful (Loy et al., 2017; see also King et al., 2018; Li et al., 2022). In this thesis, we will rely combine the online measures described above with offline measures to investigate how disfluency impacts the interpretation of speech. Specifically, we will evaluate participants' interpretations of disfluent speech as deceiving. As Ferreira and Yang (2019) point out, linking offline and online measures can provide a better understanding of how speech is processed and represented. Therefore, these measurements will be particularly relevant in Chapter 6, when we will discuss them in relation to participants' online processing of disfluent speech.

2.4 Conclusion

This chapter was a brief introduction to filled pauses. We have discussed what characterises them, and the two broad theoretical proposals that account for their production. From this discussion, we have learnt that filled pauses occur in predictable patterns that have correspondence with trouble in speech production at the level of formulating the message (and with mixed evidence for trouble at the level of conceptualising the message). Further, filled pauses' production also has a social dimension, whereby speakers appear to be more likely to produce them when interacting with social agents.

The ubiquity of filled pauses in speech, alongside their characteristics, suggests that they may impact speech comprehension. Further, these properties can have different effects depending on the communicative contexts, i.e., the processes involved in language comprehension, and the interpretation derived from speech. The first area will be explored in Part I of this thesis, and we will return to the second in Part II. Across both parts, we will use online measures to investigate how filled pauses affect the time course of language

comprehension. In Part II, we will also look into participants' interpretations to explore how the presence of a disfluency can impact the interpretation of an utterance.

Part I:

Processes in speech comprehension

Chapter 3

Prediction in speech comprehension

Speech comprehension is a remarkable ability. As listeners, we have to deal with three to five words per second (Picheny et al., 1986) and the errors that speakers may produce as they talk to us (cf. Chapter 2). Sometimes we do not pay full attention to our interlocutor, or there is background noise. Technically, all these factors should make comprehension of spontaneous speech a difficult task. However, most of us would agree that comprehension in real life feels rather effortless, but why is that so?

Traditionally, language comprehension has been thought to reflect 'bottom-up' processing (Forster, 1979). Comprehension was regarded as the outcome of linking upcoming information with previously encountered input and the comprehender's background knowledge (Long & Lea, 2005). Brown and Hagoort (1993) explored the comprehension of words preceded by semantically related and unrelated primes (e.g., *peace-war* versus *star-hour*), where primes could be masked or unmasked. This manipulation entailed that in the former condition, the prime word could not be perceptually identified, but related primes could still activate semantically related words by unconscious processing. Brown and Hagoort (1993) found that words preceded by an unmasked unrelated prime elicited a brain response associated with semantic processing (i.e., the N400 effect) in contrast to unmasked related primes, which was absent for masked, unrelated primes. The authors interpreted their findings as reflecting the easier integration of upcoming

words due to an already-built representation resulting from the prime. This increased ease in processing words thanks to the preceding context also explains why predictable words (relative to unpredictable words) are easier to comprehend (Rayner et al., 2004; see also Kutas et al., 2011). Under this view, speech comprehension reflects the ease with which one can build a coherent representation and link information: Listeners passively receive information and link it with what they have already encountered.

This view is in contrast with both anecdotal (e.g., the feeling of knowing what our interlocutor will say next) and experimental evidence (e.g., shadowing experiments, Marslen-Wilson, 1973). In their seminal eye-tracking experiment, Altmann and Kamide (1999) demonstrated that descriptions of a scene with different objects (e.g., a cake, a balloon) that included constraining verbs (i.e., verbs that can only refer to one object, e.g., only a cake can be eaten) led listeners to fixate on objects that were semantically related and plausible (e.g., *eat-cake*) before these objects were encountered in the signal. Indeed, there is now a wealth of research suggesting that comprehenders can anticipate speech continuations based on semantic (e.g., van Berkum et al., 2005; Wicha et al., 2004) as well as morphosyntactic information (e.g., Dahan et al., 2000; Kamide et al., 2003; Knoeferle et al., 2005; Lew-Williams & Fernald, 2007; Sussman & Sedivy, 2003).

In line with this evidence, current proposals of language comprehension have started to include a predictive mechanism, with authors ascribing different degrees of importance to it (e.g., Dell & Chang, 2013; Fitz & Chang, 2019; Huettig, 2015; Huettig & Mani, 2016; Nieuwland, 2019; Pickering & Gambi, 2018; Pickering & Garrod, 2004; 2014). According to this view, comprehension not only involves the previously described integrative mechanisms but is also guided by anticipatory processes where the system tries to get ahead of what it will encounter next. One way of characterising prediction is by assuming that relevant linguistic information can be activated before it is encountered in the (bottom-up) signal (Pickering & Gambi, 2018). Several questions regarding prediction in language comprehension remain unanswered (Ryskin et al., 2020), such as the representational level at which prediction occurs (e.g., syntax, Levy, 2008; semantics,

Federmeier, 2007), or the level of detail of predictions (see for example, the debate on whether phonological details of upcoming elements are activated, DeLong et al., 2005; Ito et al., 2017; Kochari & Flecken, 2019; Nicenboim et al., 2020; Nieuwland et al., 2017; Yan et al., 2017). Nonetheless, it is clear that prediction does occur, to varying degrees, in language comprehension.

Predictive processing can be guided not only by linguistic information (e.g., semantic or morphosyntactic information): *How* an utterance is delivered can also affect its comprehension. For example, prosody can increase the perceptual salience of a word (Cutler, 2005), which in turn eases its recognition (Donselaar et al., 2005), as well as mark an utterance's syntactic boundaries (Cole et al., 2015). In Part I of this thesis, we are interested in whether and how manner of delivery, in the sense of fluency, can modulate listeners' expectations with a focus on what factors can interact with manner of delivery to guide comprehension.

The structure of this chapter is as follows: We first review the literature pertaining to the effects of manner of delivery on predictive processing, by discussing the evidence on whether filled pauses can bias comprehenders' anticipations. This review will show that the previously described properties of filled pauses can explain these findings, i.e. their distribution and the causes for a speaker to be disfluent. In order to understand how these two properties can account for the effects reviewed, we will contrast them with two mechanisms put forward to account for predictive processing in speech comprehension: an *associative* and an *inference* mechanism. From this discussion, we will learn the implications of each account for the comprehension of disfluent speech: In particular, we will discuss what these accounts entail for the speaker's and the listener's characteristics. Our argument is as follows: The mechanisms proposed to aid prediction in speech comprehension predicate on specific features (e.g., exposure to a language, cognitive resources, speaker identity) that differ between native and non-native listeners and speakers. By highlighting how predictive processing can vary as a function of the linguistic background of the interlocutors, we will argue that exploring the comprehension of disfluent speech

under the light of this feature can deepen our understanding of the potential constraining effects of filled pauses on prediction.

For simplicity of argument in this section, we will couch our discussion in terms of prediction (and pre-activation), but much of what we say could also be construed in terms of ease of integration. In fact, some authors propose that prediction and integration are better understood as two sides of the same coin (e.g., Ferreira & Chantavarin, 2018; Kuperberg & Jaeger, 2016), such that speech comprehension is aided by both processes (Luke & Christianson, 2016). For example, Ferreira and Chantavarin (2018) advocate for a "forward-looking process that occurs before the linguistic input is received (...) in this view, precise lexical prediction is not common, but preparedness for relevant semantic categories and gestures is part of normal language processing" (Ferreira & Chantavarin, 2018, p. 446).

3.1 Predictive processing in spoken language: The role of manner of delivery

A speaker's manner of producing an utterance can vary in many ways, with each exerting potentially distinctive effects on how speech is comprehended. Weber et al. (2006) exploited German's free word order and use of case and gender marking to assign thematic role to explore how prosody can mark syntactic structure. In German, articles marked as masculine take different forms to mark nominative and accusative cases (i.e., *der* and *den*, respectively), while in contrast, the feminine case takes the same form: *die*. Consider sentences (1a) and (1b):

(1a) *Die Katze*$_{\text{nominative}}$ *jagt womöglich den Vogel*$_{\text{accusative}}$.

The cat$_{\text{nominative}}$ is possibly chasing the bird$_{\text{accusative}}$.

(1b) *Die Katze*$_{\text{accusative}}$ *jagt womöglich der Hund*$_{\text{nominative}}$.

The cat$_{\text{accusative}}$ is possibly chased by the dog$_{\text{nominative}}$

Due to the ambiguity of the feminine case, the thematic role of the cat is ambiguous until the second noun is encountered both in (1a), with the structure Subject-Verb-Object (SVO) and (1b), with the structure Object-Verb-Subject (OVS). In an eye-tracking experiment where participants were presented with scenes that included all characters of sentences (1a) and (1b) (i.e., a cat, a bird, a dog), the authors manipulated the utterance's prosody by either placing the stress on *cat* or on *chase*. In the former case, the utterance is said to have an SVO-type intonation, while the former aligns with an OVS-type intonation. At the moment when *womöglich* was encountered, listeners started to fixate on the dog if *chase* had been stressed (i.e., listeners understood the cat was being chased, and expected an agent for the action). Importantly, this pattern of fixations emerged before the disambiguating noun (i.e., the dog) was encountered in the bottom-up signal, suggesting that manner of delivery in the form of prosody constrained listeners' expectations about the upcoming syntactic structure (for similar findings for the role of prosody, see Hirose & Mazuka, 2015; Nakamura et al., 2012; Nakamura et al., 2022; Snedeker & Trueswell, 2003). Taken together, these findings speak to the fact that how an utterance is produced can guide listeners' expectations. In the following section, we review the effects of fluency as another dimension of manner of delivery in predictive processing.

3.1.1 Disfluency in predictive processing

Speakers can vary in how fluent they are in uttering a sentence by, for example, producing filled pauses (cf. Section 2.2). While the specifics of the mechanisms involved in disfluent speech comprehension will be reviewed in Section 4.1, in the next section we will review the effects found for disfluent speech on predictive processing of speech by dividing these effects into those that align with filled pauses' distributional properties, and those that consider the speaker's mental state.

The distribution of filled pauses in speech

One way in which disfluencies can guide listeners' anticipations is via their distribution. In Section 2.2, we discussed how one key property of filled pauses is that they occur in

predictable patterns: They are more likely to precede certain lexical items over others e.g., those that are harder to retrieve such as low-frequency words (Hartsuiker & Notebaert, 2010). In line with this hypothesis, eye-tracking experiments employing the Visual World Paradigm (VWP, see Section 2.3.1) have shown that comprehenders display anticipatory eye movements towards discourse-new referents (Arnold et al., 2004), unfamiliar elements (Arnold et al., 2007), or items with low-frequency labels (Bosker et al., 2014, Exp. 1) following a filled pause.

A parallel stream of research has measured event-related potentials (ERPs) to investigate the comprehension of disfluent speech without the confounding effect of visual presentation. Corley et al. (2007) investigated the modulation of the N400 component, an ERP component related to semantic processing and previously found to be elicited by unpredictable words given their preceding contexts. In their study, participants were presented with sentences such as *Everyone's got a bad habit and mine is biting my...* that could have a predictable (*nails*) or an unpredictable (*tongue*) continuation. Given that filled pauses commonly precede unpredictable words, the authors hypothesized that the presence of a filled pause should at least ease the integration of unpredictable words. Indeed, they found that while fluent utterances with unpredictable endings did evoke an N400 effect, the processing of disfluent utterances with unpredictable endings did not elicit this response. Their results align with those findings from experiments employing the VWP in that disfluencies affect the processing of otherwise harder-to-process words.

Additional evidence for the comprehension of filled pauses as a function of their distributional patterns comes from developmental studies. Previous research has shown that children are sensitive to the distributional patterns of the linguistic input (Romberg & Saffran, 2010; Saffran et al., 1996), suggesting that they may be equally likely to predict following filled pauses. Kidd et al. (2011) demonstrated that children as young as 20 months old display anticipatory eye movements towards unfamiliar and discourse-new entities following a disfluent utterance. However, the fact that the predicted elements were both unfamiliar and discourse-new (due to their experimental design) may have

inflated the predictive utility of the disfluency: The salience of the referent could be heightened because it meets two criteria (instead of one) to follow a disfluency. When these two factors are disentangled, children appear to display anticipatory eye movements towards discourse-new entities (Owens & Graham, 2016), but not for unfamiliar objects (Owens et al., 2018). This latter finding aligns with research demonstrating that children learn new labels equally well regardless of how they are produced in terms of manner of delivery (White et al., 2020, Exp. 2; see also Libersky et al., 2022, Exp. 1 and 2 for bilingual adults).

Overall, these findings indicate that disfluencies can affect predictive processes underlying speech comprehension by virtue of their distributional patterns. Comprehenders have learnt an association between disfluency and the elements that contextually co-occur with them via exposure to speech. At face value, this statistical learning can be taken as a 'blind' association. However, as we will see in the next section, the distributional pattern of a filled pause cannot fully account for all the evidence showing that filled pauses can constrain listeners' predictions.

The reasons for a speaker to produce a filled pause

The presence of a filled pause can be traced back to the speaker's cognitive state. For example, speakers are more likely to produce a disfluency when they experience cognitive load (Bortfeld et al., 2001). When speakers produce a filled pause, listeners can predict what the speaker will say next by reasoning about the causes for them to produce speech in such a manner, i.e., they can learn due to a causal inference between disfluency and what can cause it. Indeed, there is evidence supporting that there is a social aspect to the effects of filled pauses on speech comprehension: Disfluencies in synthetic speech do not show the same benefits in comprehension as those produced in vocoded speech, arguably because the latter sounds more natural (Dall et al., 2014). Likewise, disfluencies seem to orient comprehenders' attention to the speaker (Huizeling et al., 2022; Yeo & Alibali, 2017), suggesting that upon perceiving that the speaker struggles with speech

production, listeners focus on them to improve communication by, for example, preparing to take action (Eklund & Ingvar, 2016, Mårback et al., 2009).

One way whereby the perceived speaker's mental state can impact the predictive value of a filled pause is via comprehenders' expectations of that speaker's language use. Arnold et al. (2007, Exp. 2) presented listeners with a visual array containing familiar (e.g., ice cones) and unfamiliar (e.g., squiggly shapes) drawings. Half of the participants were explicitly told that the speaker had a condition that made speech production difficult (i.e., agnosia) and that therefore the speaker might be disfluent. While those listening to an 'unimpaired' speaker displayed anticipatory eye movements towards unfamiliar objects following a filled pause, because producing a new label is harder than producing an already-used one, this pattern was attenuated for those listening to the speaker who was introduced as having agnosia. This suggests that listeners' a priori expectations about a speaker's production abilities were taken into consideration during comprehension and, in turn, constrained the predictive value of disfluencies. A similar finding has been reported when children are presented with a speaker who is introduced as 'forgetful' (and thus produces more filled pauses) and a 'knowledgeable' speaker. In contrast to a knowledgeable speaker, the forgetful speaker's disfluencies did not lead children to display anticipatory eye movements towards unfamiliar items (Orena & White, 2015). Even when listeners are not explicitly told about the potential struggles a speaker may experience when producing speech, there is an attenuation in the predictive value of filled pauses when the speaker's identity entails a different pattern of disfluencies (e.g., if the speaker's accent indicates that they are producing speech in their second language, Bosker et al., 2014, Exp. 2).

Yet not all a priori expectations about a speaker's language use seem to modulate predictive processing. Filled pauses produced by people who stutter (and thus should be expected to produce more disfluencies) lead listeners' eye movements towards unfamiliar objects in a similar way to typically-produced filled pauses even when listeners are forewarned about the speaker's condition (Leonard et al., 2016). This contrasts with

Arnold et al.'s (2007, Exp. 2) results, where instructing participants that the speaker's speech will likely include filled pauses (i.e., the speaker had agnosia) led to a decrease in anticipatory eye movements. Leonard et al. (2016) argue this lack of effect might be attributable to listeners' quick adaptation to the speaker's distribution. Leonard et al.'s (2016) findings align with other studies exploring how certain identities stereotyped as producing disfluent speech still modulate listeners' predictions. For example, disfluencies produced by older adults bias eye movements towards discourse-new elements similarly to disfluencies produced by younger adults, even when the former group is said to be more disfluent and listeners report holding that belief (Saryazdi et al., 2021). Taken these results together, it may be possible that in some scenarios, the stereotype about the speaker's ability to produce fluent speech may be not relevant - or, at least when it comes to predicting what they will say next.

The studies described so far have used a static identity of the speaker (i.e., it is a constant attribute). A stronger test of the idea that listeners consider the speaker when comprehending disfluent speech involves exploring scenarios where the production of a disfluency is context-specific (instead of speaker-specific). Barr and Seyfeddinipur (2010) explored whether the preference for discourse-new entities following a disfluency reflected listeners' or speakers' perspectives in an eye-tracking experiment. In practice trials, participants listened to one speaker refer to elements in a visual scene, while in test trials, participants could hear either the same speaker or be introduced to a new speaker. By doing so, the authors manipulated the discourse status of the elements the speaker could refer to: In the old-speaker condition, the discourse status of the elements on the screen was the same for both parties. In the new-speaker condition, however, there was a mismatch: All elements were discourse-new for the speaker, while only some were for the listener. Barr and Seyfeddinipur (2010) found that disfluencies in the old-speaker condition led to anticipatory eye movements towards discourse-new referents, similarly to previous studies (Arnold et al., 2004). In contrast, those who were introduced to a new speaker did not display anticipatory eye movements following a disfluency. This suggests that listeners dynamically adapt to their interlocutor and predict according to

their perspective, with children as young as four years old showing this ability (Yoon et al., 2020, Exp. 2, but cf. Thacker et al., 2018, for a lack of effect in a similar task).

Heller et al. (2015) took Barr and Seyfeddinipur's (2010) design one step further by testing participants with an artificial language. Across two experiments, the authors tested the degree of flexibility of comprehension of disfluent speech. By employing an artificial lexicon, the authors aimed to disentangle whether anticipations following a disfluency are modulated due to listeners' prior experience, or by recently acquired knowledge. In these experiments, participants saw scenes depicting objects for which they had not learnt a label (and thus disfluency could be a cue for these referents, as disfluency is associated with objects with unconventional names, that require longer descriptions, and that have not been referred to previously), alongside objects with recently learnt labels (and thus disfluency could be associated with these labels, because they might be perceived as difficult to retrieve). Further, Heller et al. (2015) additionally manipulated the speaker's and the listener's perspectives: Participants were informed about whether the speaker knew all the labels or not. This manipulation explored whether comprehenders anticipated based on their own perspective, or if they could consider their interlocutor's knowledge.

In two eye-tracking experiments, Heller et al. (2015) reported that disfluent utterances modulated participants' expectations. Specifically, the presence of a disfluency biased participants' fixations towards objects with newly acquired labels (instead of objects that require long descriptions or that lack a conventional name, Exp. 1). However, when participants received a longer training session with the artificial lexicon (Exp. 2), this preference for objects with newly acquired labels following a disfluency disappeared, suggesting that the findings in Experiment 1 could be attributed to listeners' inferences of what is difficult to produce. However, in both experiments, the bias elicited by the presence of a disfluency did not consider the speaker's perspective: Regardless of whether the speaker was said to know or not know the label, participants preferred to fixate on elements that could pose problems in production *given their own knowledge*. This is at

odds with Barr and Seyfeddinipur's (2010) evidence of perspective-taking while comprehending disfluent speech, but aligns with a larger body of research pointing out that there may be limits to the sources that can modulate the effects of disfluency on comprehension. For example, filled pauses that can be attributed to an external cause (e.g., environmental noises that can distract the speaker) still bias listeners' eye movements towards elements that contextually co-occur with disfluency such as unfamiliar objects (Arnold et al., 2007, Exp. 3).

In line with Heller et al. (2015), further studies have shown that the properties of the experiment itself can override listeners' previous preferences. Bosker et al. (2014, Exp. 2) reported that a non-native speaker's disfluencies do not cue low-frequency items for native listeners, which the authors attributed to listeners' experience with non-native-accented speech, which is riddled with disfluencies and is perceived as more disfluent than its native counterpart (Pinget et al., 2014). Bosker et al. (2019) explored whether listeners can update these prior expectations when there is evidence (or lack thereof) in the signal that a filled pause is (or is not) a reliable cue of the target's frequency. Across three experiments, the authors manipulated the overall distribution of disfluencies in the experiment: Half of the participants were presented with a speaker whose disfluency distribution was typical (i.e., the speaker was fluent when producing high-frequency targets, and disfluent when producing low-frequency targets), and the other half encountered an atypical distribution.

In Experiments 1 and 2, participants listened to instructions produced by a native speaker. In these experiments, those in the typical-distribution condition displayed anticipatory eye movements towards low-frequency items following a disfluency, replicating previous findings (Bosker et al., 2014, Exp. 1). Those in the atypical-distribution condition showed an adaptation to the signal's properties: Towards the end of the experiment, filled pauses led to anticipatory fixations towards high-frequency items, suggesting that listeners' updated their expectations following the specific distribution in the signal. Interestingly, this updating of expectations was not seen to the same extent in Experiment

3, where participants listened to a non-native speaker. In this case, those exposed to a typical distribution adapted to the speaker, and displayed anticipatory eye movements towards low-frequency items (in contrast to Bosker et al., 2014, Exp. 2). However, those in the atypical distribution condition did not adapt to the signal's properties and did not display anticipatory eye-movements towards high-frequency items, in contrast to Experiments 1 and 2. This suggests while that listeners' expectations about the properties of speech can be updated, it is a function of comprehenders' a priori expectations about speakers (see also Thacker et al., 2018, for a similar finding with children).

This stream of research suggests that the distribution of filled pauses itself cannot fully account for how filled pauses guide prediction. Instead, there seems to be a social factor, be it in the sense of listeners' models about speakers, or in situ social reasoning, that underlies the comprehension of disfluent speech. The results of Heller et al. (2015) and Bosker et al. (2019) point out the fact that there is a close link between an association (i.e., the distributional patterns of disfluencies) and an inference (i.e., the causes for the speaker to be disfluent) account: While listeners can adapt to their interlocutor (Bosker et al., 2019), this is still limited by listeners' previous experience with language (Heller et al., 2015).

To complicate matters further, it is possible that the benefits of disfluency arise simply due to the interruption of the speech signal, in line with the properties discussed in Section 2.1. A filled pause can affect speech comprehension by highlighting word boundaries. Brennan and Schober (2001) found that listeners are faster at recognising the intended word when speakers produce a repair disfluency if they include a filled or a silent pause between the error and the repair(e.g., *yel-uh-orange* versus *yel-orange*). The *temporal delay hypothesis* (Corley & Hartsuiker, 2011) suggests that the benefits of filled pauses might be primarily explained by the temporal characteristics of filled pauses: The interruption and delay they cause in the speech signal can subsequently aid word segmentation (as sounds following the pause likely belong to another word), and potentially, provide more time for any ongoing processes (e.g., lexical activation)

(Corley & Hartsuiker, 2011). Indeed, Corley and Hartsuiker (2011) found that *any* break in the speech signal speeded up word recognition: Listeners were faster at selecting the appropriate target if it was preceded by a filled pause (Exp. 1), a silent pause (Exp. 2) or a pause filled with a non-sound (i.e., a sine wave, Exp. 3). Follow-up studies have shown that the time benefits of both filled and silent pauses enhance the processing of nouns that surround them by influencing lexical activation (Baker & Love, 2022), and silent pauses equally ease the integration of unpredictable words (MacGregor et al., 2009). However, a different stream of research has suggested that delays themselves cannot fully account for the differences associated with comprehending disfluent speech: For example, pauses of the same length as a filled pause filled with sounds such as coughs do not impact participants' predictive processing (Barr, 2001; Barr & Seyfeddinipur, 2010), and temporal delays introduced by discourse markers such as *oh* lead to faster recognition of words than when the *oh* is excised and there is only a pause (Fox Tree & Schrock, 1999).

The interruption to the speech signal can also impact processes involved in speech comprehension that are not prediction. It has been shown that changes in the speech signal (e.g., stress) can lead listeners to notice changes that otherwise go unnoticed, arguably because these changes attract listeners' attention (Sanford et al., 2006). Fox Tree (2001) found that participants were faster at responding to target words when they were preceded by filled pauses compared to utterances where filled pauses had been excised and thus were fluent, with a difference between *uh* and *um*: Participants were faster responding to the former than to the latter, which showed no differences from its fluent counterpart. Fox Tree (2001) argued that this difference was attributable to the different lengths of *uh* and *um*: Short delays, signalled by *uh*, 'heighten listeners' attention for upcoming speech' (Fox Tree, 2001, p. 325). This heightened attention can also increase listeners' ability to detect changes in the transcription of previously encountered text that contained disfluencies (Bosker et al., 2015).

Online evidence for the attention-orienting effects of filled pauses comes from Collard et al. (2008). In an ERP study, participants were presented with a speech stream

where target words could be acoustically manipulated. Crucially, half of the time, these manipulated words were preceded by a filled pause. The authors argued that, if filled pauses act as 'attention attractors', then ERP components associated with increased attention (due to novel stimuli, e.g., P300) should be diminished for manipulated words if they are delivered disfluently because the disfluency would have already re-oriented listeners' attention to the speech signal. Collard et al. (2008) found that, while manipulated words embedded in fluently delivered sentences elicited components associated with the re-direction of attention to the speech signal, their disfluent counterparts showed a reduction. In fact, Mårback et al. (2009) reported that filled pauses themselves elicited a P300 response (yet cf. Agmon et al., 2022, for a lack of replication). Additional evidence suggests that filled pauses increase activation in brain areas associated with attention heightening (e.g., Primary Auditory Cortex, PAC; Eklund & Ingvar, 2006, yet cf. Smirnova et al., 2020, for a lack of replication). Taking these results together, filled pauses seem to increase listeners' attention to the speech signal, and subsequently, the comprehension of the following linguistic items could be facilitated by this heightened attention.

This review of the evidence highlights how the presence of a filled pause can impact comprehension. This discussion suggests that the ways in which filled pauses can affect comprehension can be traced back to their properties; namely, their distribution, the reasons for the speaker to be disfluent, and the interruption to the speech signal. We have specifically argued that the presence of disfluencies can guide comprehenders' anticipations of what the speaker will say next, which is influenced by the patterns of a language (what we refer to as an *association*) as well as social reasoning (what we refer to as an *inference*). These two mechanisms are akin to those proposed by several authors as guiding predictive processing in speech comprehension (e.g., Pickering & Gambi, 2018; Pickering & Garrod, 2004, 2014; Huettig, 2015; Dell & Chang, 2014; Federmeier, 2007)[1].

[1]Our discussion concerns accounts with *at least* two mechanisms for prediction in language comprehension. However, some authors have proposed one-stage models (e.g., MacDonald et al., 1994; Altmann & Mirkovic, 2009), where all sources of information (i.e., linguistic – rule-based – and non-linguistic – world-knowledge) constrain predictions simultaneously (Metusalem et al., 2012). However, one-stage

Let us thus explore how these mechanisms have been described in the literature to understand their implications and subsequently the factors that should be considered when investigating the effects of disfluencies in language prediction.

3.2 Accounts of prediction in language comprehension

Most accounts of prediction in speech comprehension involve *at least* two stages. For example, Pickering and Gambi (2018) propose two mechanisms to account for predictive processes in language comprehension: *prediction-by-association* and *prediction-by-production*, while Huettig (2015) argues for predictions based on associations and those that rely on combinatorial mechanisms (alongside predictions that originate from the comprehender's own production system, and simulations). We will refer to these two mechanisms as an *associative* and as an *inference* mechanism.

On the one hand, the associative mechanism accounts for predictions originating from individuals' experience with linguistic input, as a form of Hebbian learning (Huettig, 2015). For example, anticipations may originate via priming due to spreading activation from encountered linguistic input. This spreading of activation can be due to learnt associations in semantic networks (Anderson, 1983). Consequently, upcoming linguistic input activates elements associated with it (e.g., *chair* activates *to sit*, Huettig et al., 2022). This form of learning can also involve acquiring the statistical properties of the input which can subsequently drive comprehension (MacDonald, 2013). For example, probabilistic models of speech comprehension account for individuals' experience with linguistic input as a factor that can modulate anticipations about upcoming elements in the speech signal (Christiansen & Chater, 2016; Kleinschmidt & Jaeger, 2015; Kuperberg & Jaeger, 2015). This form of prediction is said to be automatic and thus is cost-free and effortless, with its effects emerging rather soon after encountering the bottom-up input that triggered the prediction. This mechanism can thus account for the evidence discussed at the beginning of the chapter: For example, encountering *eat* in the speech

accounts are difficult to falsify (cf. Huettig, 2015) and thus the discussion here will not consider these models.

signal activates its representation in a semantic network, which propagates to related concepts, such as *cake* (Altmann & Kamide, 1999).

On the other hand, the inference mechanism accounts for predictions that are informed by non-linguistic information (Pickering & Garrod, 2004, 2014). For example, anticipations might be further informed by individuals' knowledge about the world. Van Berkum et al. (2008) demonstrated that speaker identity can impact the integration of upcoming linguistic input. In their EEG study, listeners were presented with sentences where the event conveyed could match the stereotypes associated with the speaker (i.e., pragmatically consistent, e.g. (where italics signal the manipulation), 'I speak to my *son* a lot on the weekends', said by an adult) or there could be a mismatch (i.e., pragmatically inconsistent, e.g., 'I speak to my *son* a lot on the weekends', said by a child). Listeners' responses were contrasted against semantically inconsistent sentences (e.g., 'I speak to my *valley* a lot on the weekends'). The authors found that pragmatically incongruent sentences elicited similar responses to semantically incongruent sentences, but additionally, pragmatically incongruent sentences elicited responses as early as 200-300 ms post-target onset, which the authors took as evidence of listeners' difficulty in integrating the linguistic input with the model they had built of the speaker. Since this study, there has been an increasing emphasis on the role of the speaker in guiding comprehension (Kleinschmidt & Jaeger, 2016; Kuperberg & Jaeger, 2015). Predictions emerging from the inference mechanism are described to be costly, effortful, and non-automatic due to the interaction of different sources of information. Consequently, they emerge later compared to those produced by the associative mechanism.

There is evidence supporting this proposed time course for the effects of inferences emerging from associative and inference mechanisms. Corps et al. (2022) had participants observe a scene depicting two (out of four) wearable objects, where each was stereotypically associated with males or females (e.g., a *tie* and a *dress*), while listening to male and female speakers. The authors found that listeners first displayed anticipatory eye movements towards both wearable objects upon encountering the verb *to wear*, to

shortly later fixate on the object that was associated with the speaker's gender. This occurred regardless of the participant's gender, and whether the agent of the action was the speaker (Exp. 1), the listener (Exp. 2), or a third character (Exp. 3). These pattern of fixation supports the idea that predictions triggered by associative mechanisms (i.e., the spreading activation of the verb semantics towards elements that are associated with this information) precedes predictions triggered by inference mechanisms (i.e., the activation of the object that is stereotypically associated with the speaker).

The properties of these proposed mechanisms present interesting consequences for how prediction occurs during speech comprehension; specifically, with regard to whether and how one can expect some degree of variability. Understanding prediction as an (at least) dual process that can operate with stable representations (i.e., associations) while adapting to variability in the signal (i.e., inferences) suggests that a cue that can be associated with certain elements, and thus bias anticipations towards those, can see its effect attenuated when an inference entails that such association is not representative in the current context of comprehension.

In fact, this is similar to the evidence we discussed in Section 2.2 for the effects of filled pauses in comprehension: Listeners' anticipations seem to align with the distributional patterns of disfluencies, but can be overridden (in some contexts) by information about the speaker that renders this association less reliable. Therefore, if the same mechanisms that support prediction in speech comprehension are the ones underlying the effects of disfluency, then a potential way to examine how disfluencies affect predictive processing is by exploring it under factors previously shown to modulate linguistic prediction in general. In this thesis, the focus is on linguistic background (i.e., 'nativeness') as an attribute of speakers and listeners that can have an effect on what comprehenders anticipate. In what follows, we review the evidence supporting the idea that linguistic background does affect prediction.

3.2.1 Speaker's linguistic background: Non-native speakers

The inference mechanism can be influenced by factors such as speaker identity. One pertinent feature of a speaker's identity is their linguistic background, which in speech can be indexed by their accent. Non-native speakers of a language commonly produce speech with an accent (Abrahamsson & Hyltenstam, 2009), which includes non-canonical word pronunciations (Hanulíková & Weber, 2011). These speakers usually differ from native speakers in terms of their lexical choices (Dewaele & Pavlenko, 2003) and their likelihood to commit grammatical mistakes (Clahsen & Felser, 2006). These differences in their speech patterns allow listeners to identify speakers' nativeness as early as 100 ms after being exposed to their speech (Jiang et al., 2019).

It is important to note that a speaker's linguistic background may affect speech comprehension solely due to their accent. Non-native-accented speech is harder to comprehend (Grey & van Hell, 2017). Following an *intelligibility account*, the effects reported for non-native-accented speech comprehension are due to the decreased ease of processing it (Lev-Ari, 2015). This decreased intelligibility can explain why accented speech elicits negative attitudes (Dragojevic et al., 2017), as well as the fact that detecting grammatical mistakes in non-native-accented speech is usually delayed (Gosselin et al., 2021; Sanders & de Bruin, 2022).

Alternatively, but not exclusively, comprehenders' a priori expectations about non-native speakers' speech can exert an effect by modulating different mechanisms underlying comprehension. Non-native speakers' accents may invoke stereotypes whereby these speakers are believed to be less competent than native speakers (Lev-Ari, 2015). These stereotypes cascade onto how their speech is comprehended: For example, non-native-accented utterances containing grammatical mistakes are processed differently than their native counterparts, arguably because listeners expect them to commit them: The particular model a comprehender has of these speakers affects how their speech is comprehended (Hanulíková et al., 2012; Grey & van Hell, 2017; Caffarra & Martin, 2018; Grey et al.,

2020). It is thus possible that nativeness affects other aspects of comprehension, including prediction.

Lev-Ari (2015) proposes an *expectation-based account* for the comprehension of non-native-accented speech. In her proposal, stereotypes of non-native speakers' low linguistic competence affect comprehension via an inference mechanism, which impacts both the processing and the interpretation of their speech. A stereotype of non-native speakers as less capable of conveying their intentions via language leads the comprehension system to be guided by other sources of information to compensate (i.e., they do not consider the speaker). According to Lev-Ari (2015), predictions about upcoming elements in non-native-accented speech are more permissive and less detailed: For example, a native listener is more likely to represent *pie* in a broader manner, including elements like *brownie*, when they are listening to a non-native speaker.

Lev-Ari (2015) conducted a VWP experiment to explore whether native listeners' comprehension of speech differed depending on the speaker's linguistic background. In her study, participants were presented with a series of scenes depicting several characters and were warned that there was an underlying common theme to the characters the speaker will refer to. For example, one set of instructions would refer to imaginary creatures (e.g., *the witch*, *the man on the magic carpet*). This manipulation ensured that the context itself could constrain listeners' anticipations. At critical trials, the visual scene included elements that could match an ambiguous speech signal e.g., *fɛri* can be taken as referring to a *ferry* and a *fairy*. Eye movements towards the ferry would suggest listeners' reliance on upcoming linguistic information, whilst the opposite would be evidence of listeners' reliance on discourse knowledge. When listening to a non-native speaker, participants were more likely to fixate on the fairy (as opposed to the ferry), suggesting that comprehension relied more heavily on discourse expectations than on the bottom-up signal to resolve an ambiguity. Lev-Ari (2015) took this as evidence of a compensatory mechanism due to listeners' assumption that their interlocutor had less competency in reliably producing speech.

Alternatively, the perceived reduced reliability of non-native-accented speech can halt all predictive mechanisms, or reduce the set of elements expected. Romero-Rivas et al. (2016) conducted an ERP experiment to measure listeners' responses to sentence completions with varying degrees of semantic relatedness. For example, in the sentence *In the pirates' map there was an X showing the location of the...*, the word *treasure* is the best completion, but *chest* is also acceptable, in contrast to semantically unrelated words such as *enemy*. When attending to native-accented speech, participants showed a gradient N400 response: *Chest* elicited an N400 higher than *treasure*, but smaller than *enemy* (in line with previous findings, Federmeier & Kutas, 1999). However, for non-native-accented speech, both semantically related and unrelated endings elicited a similar N400 response. This difficulty in integrating semantically-related words in the non-native speaker condition was taken by the authors as evidence of listeners' reduced set of expectations for upcoming elements in speech.

Romero-Rivas et al. (2016) measured responses after the linguistic input had been encountered, and thus cannot explore whether fewer predictions were triggered. To explore the pre-activation (or lack thereof) of linguistic elements, Schiller et al. (2020) measured ERPs for articles preceding nouns that varied in the degree to which they were the best sentence completion. The authors exploited the gender marking of Dutch, where articles agree with the noun, by presenting participants with sentences where the expected and unexpected sentence completions differed in gender. Consider the differences between (2a) and (2b):

(2a) *Mijn gelezen boeken staan op de onderste plank in* ***de*** $_{\text{neuter}}$ *kast*$_{\text{neuter}}$.

My read books stand on the lowest shelf in **the**$_{\text{neuter}}$ (book)shelf.

(2b) *Mijn gelezen boeken staan op de onderste plank in* ***het***$_{\text{non-neuter}}$ *bureau*$_{\text{non-neuter}}$.

My read books stand on the lowest shelf in **the**$_{\text{non-neuter}}$ desk$_{\text{non-neuter}}$

Bookshelf is a predictable ending following the sentence context, while *desk* is not. Since articles in Dutch agree with nouns in gender, encountering *het* when the listener

anticipates *bookshelf* (and thus, *de*) should elicit an N400 response. While encountering *het* when *de* was expected elicited an N400 response both for the native and non-native accents, mismatching articles in the non-native speaker condition did not elicit any neural response in an early time window. The authors interpret the lack of early modulation for non-native-accented unexpected articles as aligning with Romero-Rivas et al. (2016), in that non-native-accented speech reduced predictive processing. However, it is worth querying whether Schiller et al.'s (2020) lack of effects for non-native-accented speech is due to a halting of predictive processing or if, instead, they reflect less detailed predictions (i.e., listeners do not activate morphosyntactical details of anticipated lexical items when attending to a non-native speaker). When comprehending non-native-accented speech, predictions about upcoming elements might not include its morphosyntactic details in order to avoid "deal[ing] with conflicting information only at the levels where the input might be most misaligned with one's predicted features" (Schiller et al., 2020, p. 9).

Although the evidence presented here is mixed with respect to how a speaker's linguistic background affects predictive processing, it shows that it is a potential constraint, although the direction in which it can constrain prediction is unclear. These findings support the notion that listeners' beliefs about the communicative abilities of non-native speakers, whereby these speakers are expected to be less reliable, guide their comprehension. The fact that there is some evidence supporting the existence of prediction suggests that difficulties in comprehending non-native-accented speech cannot fully account for the differences in comprehending native and non-native speakers (Lev-Ari, 2015). An interesting consequence of these findings is the question of what areas non-native speakers are expected to be less competent in. For example, a stereotype of low linguistic competence may entail an expectation that non-native speakers' speech is less fluent. In what follows, we move on to what happens when it is the listener who is a non-native.

3.2.2 Listener's linguistic background: Non-native comprehenders

Conceptualising prediction as the outcome of mechanisms that depend on language exposure and varying degrees of effort entails differences across individuals. One source of variability between individuals is whether they are comprehending speech in their first or second language (or native and non-native, respectively). Kaan (2014) posits that second-language comprehension does not differ from first-language comprehension in terms of the mechanisms that underlie it; rather, they are impacted by the same factors. She identifies five factors that affect predictive processing: (1) the frequency information comprehenders have, which is heavily associated with exposure to a language, (2) the presence of competing information to create a prediction, (3) the reliability of the information comprehenders hold to guide their comprehension, (4) the processes and strategies induced by the task at hand, and (5) the cognitive resources involved in comprehension, including motivation and control; factors that are interrelated (Schlenter, 2022)[2]. Consequently, the differences found between first- and second-language comprehension can be explained by differences between comprehenders in these five factors.

Non-native comprehenders differ from their native counterparts in terms of their exposure to the target language. Increasing exposure to a language can yield more stable and stronger associative networks, which in turn increases both spreading activation and lexical access, subsequently easing predictions based on associations. As non-native comprehenders are less exposed to their non-native language, they have slower lexical access due to weaker semantic networks (Gollan et al., 2008; Ivanova & Costa, 2008). For example, the word frequency effect, whereby higher-frequency words are recognised faster than low-frequency ones, is exacerbated in second-language comprehension arguably due to a more dramatic division between these two categories in their lexicon (Duyck et al., 2008). As lexical representations do not only include semantic information but also

[2]A complete review of non-native language comprehension is outside of the scope of this thesis (and for that, we refer the reader to Kaan, 2014; Ito & Pickering, 2021; Kaan & Grueter, 2021; Schlenter, 2022), in this section we review how predictive processes in non-native language comprehension can be understood in the light of the accounts previously described.

morphosyntactic details (cf. Section 2.2.1), the increased difficulties in lexical retrieval cascade into difficulties in morphosyntactic processing (Hopp, 2018).

Predictive processing may be thus attenuated in second-language comprehension by virtue of weaker networks. Experiments using simple stimuli (akin to Altmann & Kamide, 1999) have reported a similar time course of eye movements for predictions triggered by semantic information (e.g., Corps, Liao, et al., 2022; Dijkgraaf et al., 2017; Ito et al., 2018). However, when more complex stimuli are used (e.g., *I know the friend of the dance that will* ***open/get*** *the present*, where the bold signals the manipulation, and the underlined word is the target), the emergence of predictions triggered by verb semantics were delayed for non-native listeners in comparison to native listeners (Chun & Kaan, 2019; see also Dijkgraaf et al., 2019, for similar findings)

Weaker networks can also impact predictions following morphosyntactic information. Lew-Williams and Fernald (2010) found that non-native Spanish (L1: English) listeners did not employ gender marking to anticipate upcoming referents, contrary to native Spanish listeners. Similarly, non-native Japanese speakers (L1: English) do not show anticipatory eye movements towards elements cued by morphological case (Mitsugi & MacWhinney, 2016; Mitsugi, 2017). However, when second-language listeners have experience with these cues, be it because comparable cues exist in their first language (Foucart et al., 2014, 2016), because of an increase in proficiency (Dussias et al., 2013; Foucart & Frenck-Mestre, 2012) or because of their specific knowledge of the rule (Hopp, 2013, 2015, 2016), these listeners exhibit prediction towards referents marked by such cues. Taking these findings together, the evidence suggests that predictions via an associative mechanism can arise in second-language comprehension, although it is a function of individuals' exposure to the target language and the similarities with their own language.

The mechanisms discussed in Section 3.2 posit that the amount of resources available can impact predictive processing. Anticipatory eye movements have been shown to be impacted by individuals' available working memory resources (Huettig & Janse,

2016; Ito et al., 2018). This is particularly relevant for second-language comprehension, which is assumed to be more cognitively taxing (Segalowitz & Hulstijn, 2005; Weber & Broersma, 2012; although cf. Dijkgraaf et al., 2019). This may be because this population seems to activate both their language systems and lexicons in parallel and thus experience more lexical competition (Blumenfeld & Marian, 2013; Costa et al, 2006; Duyck, 2005; Duyck et al., 2007; Weber & Cutler, 2004), with their second language experiencing more cross-linguistic influence (Karaca et al., 2021). This, in turn, requires them to exercise cognitive control to only employ the relevant knowledge for the task at hand and inhibit incorrect predictions (Cunnings, 2016). The resource demands associated with second-language comprehension render it less automatic (Ito & Pickering, 2021), subsequently shaping how predictive mechanisms operate.

Additionally, predictions in second-language comprehension can be affected by the paucity of resources available - which may be particularly salient for predictions triggered by an inference mechanism as they are resource-intensive. In accordance with this prediction, integrating semantic information that runs against world knowledge is more costly when comprehending one's second language, arguably because this entails an extra cost (Foucart et al., 2016; Romero-Rivas et al., 2017). Corps et al. (2022) extended Corps et al. (2022) to a sample of non-native listeners. In this study, participants' predictive fixations towards elements that were associated with verb semantics (e.g., *to wear* → a *tie*, a *dress*) had a time course similar to native listeners. However, anticipatory fixations towards elements that were associated with the speaker's gender stereotypes (e.g., a *tie* for a male voice) were delayed in comparison to native listeners. This suggests that, when it comes to predictions informed by inferences, there may be a delay in second-language comprehension due to its increased cognitive load.

The evidence presented here suggests that second-language comprehension can also be guided by predictive processing. The findings can be understood in the light of the mechanisms proposed to underlie prediction in language comprehension, given the

factors that can affect them. Specifically, predictions triggered by an associative mechanism arise in second-language comprehension, but they seem to be weaker as non-native comprehenders tend to have less exposure to the target language. Predictions via an inference mechanism are effortful; since second-language comprehension is said to be cognitively costly, these form of predictions are delayed in this population. Importantly, the differences in how an associative and an inference mechanism operate in second-language comprehension suggest that predictions guided by manner of delivery, and specifically, by fluency, can differ in this population.

3.3 Conclusion and Chapter aims

In this chapter, we presented the evidence suggesting that comprehension of disfluent speech affects predictive processes underlying speech comprehension. Specifically, we reviewed how listeners seem to anticipate certain elements over others as likely continuations of what the speaker would say. We evaluated this evidence against two properties of filled pauses: their distribution, and the reasons for a speaker to be disfluent, what we referred to as an associative and an inference mechanism. These two mechanisms were contrasted with those put forward to account for predictive processing in speech comprehension in general. Elements can be pre-activated due to spreading activation from linguistic input to elements that they are associated with or semantically related to (i.e., an associative mechanism), or due to a refinement of predictions thanks to the integration of non-linguistic information such as speaker identity (i.e., an inference mechanism). These mechanisms differ in the costs they require and the flexibility of predictions. An associative mechanism arises from priming effects, which are effortless yet require listeners' exposure to a language; and an inference mechanism arises from a refinement of those initial predictions, which in turn require more cognitive effort.

These two mechanisms not only align with two properties of filled pauses, but can also account for the effects found throughout the literature suggesting that the presence of a filled pause in an utterance can guide listeners' predictions about upcoming elements

in the signal. We explored the circumstances under which these two mechanisms can give rise to different predictions, by taking speakers' and listeners' backgrounds as proving grounds. This discussion also highlighted the fact that comprehension of non-native-accented speech differs from that of native-accented, due to an inference mechanism: Listeners seem to hold a stereotype of non-native speakers as less competent, which impacts comprehension in several ways, with most of them aligning with the idea that they are less reliable speakers. Finally, predictions in second-language comprehension seem to be attenuated, particularly when predictions involve an inference mechanism.

This literature review opens up several questions with regards to the processing of disfluent speech. Would disfluent speech comprehension differ when it is produced by a native speaker than when it is produced by a non-native speaker? And would there be differences between first- and second-language comprehension of disfluent speech? In the following chapter, we will argue that exploring these two questions can deepen our understanding of how disfluent speech informs predictive processing. We will formalise how these associative and inference accounts operate and their implications. Specifically, we will discuss how they differ in terms of costs and flexibility. Experiments 1 and 2 were designed with this conceptualisation in mind to explore under what circumstances either mechanism might be more salient.

Chapter 4

Disfluency as Difficulty

In spontaneous speech, speakers produce more than just words: Excluding pauses, speakers average around six disfluencies per every 100 words (Shriberg, 1996). Filled pauses (e.g., *uh* or *um* in British English) are the disfluencies most widely attested to impact language comprehension: For example, they can affect the comprehension of lexical units (Beattie & Butterworth, 1979; Corley & Hartsuiker, 2011; Arnold et al., 2004, Arnold et al., 2007), how a sentence is parsed (Lau & Ferreira, 2005), how a discourse is represented (Cossavella & Cevasco, 2021), and can increase the recall of items that accompany them (Corley et al., 2007, Diachek & Brown-Schmidt, 2022, Fraundorf & Watson, 2011). At the lexical level, filled pauses have been shown to affect the comprehension of the linguistic units that follow them, so that in some circumstances, the comprehension of otherwise hard-to-process elements is eased - what we will refer to as the *disfluency bias*. Several explanations have been put forward to account for this bias, implicating time, statistical learning, and speaker modelling (Arnold et al., 2007, Bosker et al., 2014, Bosker et al., 2019, Corley & Hartsuiker, 2011).

To date, the vast majority of research on the disfluency bias has focused on first language comprehension, in which listeners are responding to utterances spoken in their native language, commonly produced by a native speaker. What remains less clear is what happens when a speaker is hearing their second, or non-native, language, in which

they are presumably less proficient than in their native language. As we argue below, investigating the disfluency-bias cross-linguistically allows us to form some conclusions about the mechanisms which underlie it. The two experiments reported in this chapter explore a previously reported bias whereby filled pauses lead comprehenders to anticipate low-frequency words (versus high-frequency ones) if they are attending to a native speaker, but not when the speaker is non-native (Bosker et al., 2014). Although due to the nature of our auditory stimuli we cannot draw firm conclusions about the mechanisms that underlie the disfluency bias, the evidence presented here is a first step towards comprehending how this bias operates.

In what follows, we briefly review the accounts explaining the disfluency bias. We will review the evidence supporting each account, and highlight their different predictions and the degree of specificity of what listeners can anticipate. We then present Experiment 1, in which we replicate Bosker et al. (2014) in a sample of native English listeners. Contrary to Bosker et al. (2014), we failed to find evidence for anticipatory eye movements towards objects with low-frequency labels following a native's filled pause. Instead, we found that both native and non-native disfluencies aided in the recognition of low-frequency words. After discussing the differences between Bosker et al. (2014) and Experiment 1, we will present Experiment 2, where we extended Experiment 1 to a sample of non-native English listeners. We will review how non-native listeners deal with disfluencies in speech, and explain how the characteristics of second language comprehension can shed light on the mechanisms underlying the disfluency bias overall. In Experiment 2, we found similar effects to those found in Experiment 1 for a native listener: Filled pauses aided in the recognition of low-frequency words. In contrast, non-native disfluencies aided word recognition regardless of its frequency. We will then conclude this chapter by comparing the findings of these two experiments against the mechanisms put forward to account for the disfluency bias.

4.1 The disfluency bias

Filled pauses introduce an interruption to the speech signal without adding propositional content (Fox Tree, 1995). These interruptions follow a non-arbitrary distribution (Bailey & Ferreira, 2007): Filled pauses are more likely to precede items without a conventional label (Arnold et al., 2007) as well as items whose label is less accessible, such as low-frequency words (Hartsuiker & Notebaert, 2009). This distribution can be partially explained because speakers are more likely to be disfluent under circumstances of cognitive load (Bortfeld et al., 2001): Deciding how to refer to an object or retrieving less accessible lexical items is cognitively costly, and thus disfluencies commonly precede such items (see Section 2.2). These three characteristics (i.e., an increase of time, distributional patterns, and origins) of filled pauses' production have been used to explain their effects on comprehension, which fall into two broad camps: the *fillers-as-time* and the *fillers-as-tokens*.

The fillers-as-time has its basis in the fact that filled pauses introduce extra time without propositional content. This interruption and delay may benefit the comprehension of the linguistic units that follow a filled pause simply by virtue of highlighting word boundaries and thus aid word recognition (Corley & Hartsuiker, 2011). Therefore, any pause in speech should exert similar effects. In line with this prediction, the interruption introduced by filled pauses, silent pauses and sine waves of the same length has been shown to lead to faster word recognition (Corley & Hartsuiker, 2011) and both filled and silent pauses ease the integration of words with low contextual probabilities (Corley et al., 2007, MacGregor et al., 2010).

Additionally, these interruptions may affect attentional processes that in turn impact word recognition (Fox Tree, 2002). Indeed, filled pauses have been shown to re-orient attention to the speech signal (Collard et al., 2008) and to increase the activation of brain areas associated with attention (Eklund & Ingvar, 2016). As increased attention to an object can ease its processing (Posner, 1988), the processing of elements following a filled pause can be enhanced simply because listeners are paying attention to it. Crucial

for the fillers-as-time account, no new processes are triggered by filled pauses; rather, the delay and increase in time may just provide a window for any processes that were already underway, e.g., prediction about upcoming items.

One assumption of the fillers-as-time view is that listeners do not need to recognise a filled pause as such. An alternative possibility is that for the disfluency bias to arise, listeners need to first recognise that the speaker is being disfluent. For example, the discourse marker *like* influences the activation of phonologically overlapping words (e.g., *lightbulb*, Bosker et al., 2021). At some point, this discourse marker must be distinguished from these content words (and from its homonyms) for the listener to comprehend what the speaker is referring to.

In opposition to the fillers-as-time view, according to the fillers-as-tokens account, filled pauses are recognised as such and are processed by the language system. An influential version of this account was proposed by Clark and Fox Tree (2002), who argue that fillers are collateral signals that allow speakers to signal their mental states (e..g, difficulty finding a word), alongside the primary, contentful words they are uttering (cf. Section 2.2.2). Consequently, listeners are tuned to speakers' displays, which can guide their comprehension of speech. For the fillers-as-tokens account, the recognition of filled pauses can trigger additional processes underlying language comprehension: These signals allow listeners to refine their predictions of what speakers might be about to say. Indeed, predictive processing has been put forward as a mechanism that accounts for the efficient and rapid comprehension of speech (Pickering & Gambi, 2018; Pickering & Garrod, 2004, 2014) and thus it would be reasonable to expect listeners to form predictions from any signal in the message, including filled pauses.

If fillers are tokens, a simple way in which a listener's predictions might be updated is via learnt associations between these tokens and the items which usually co-occur with them. We will refer to this account as the *tokens-associative* account, which can be seen as learning patterns. According to this view, there is nothing 'special' about fillers: They are simply elements that follow distributional patterns that listeners can learn

and this knowledge drives their comprehension. Arnold et al. (2007) argue that a simple mechanism to account for the disfluency bias involves learning patterns between disfluency and specific stimuli: Given filled pauses' distribution (e.g., individuals are more likely to have encountered a disfluency accompanying a discourse-new entity than a discourse-old one; Arnold & Tanenhaus, 2007), they can be a probabilistic distributional cue that can be learnt and consequently guide language comprehension in a relatively automatic manner. In this sense, disfluency might be linked to different sets of objects. Therefore, upon encountering a filled pause, listeners should reliably predict that the speaker will produce language which is (contextually) associated with filled pauses. At this level, it is important to note that this pre-learned association can either involve disfluencies and the types of elements that co-occur with it, or an understanding of why certain sets of elements follow a disfluency. In line with this view, Visual World eye-tracking studies have shown that upon encountering a disfluency, listeners display anticipatory eye movements towards harder-to-name objects (Arnold et al., 2007), low-frequency items (Bosker et al., 2014), or discourse-new entities (Arnold et al., 2004) - all elements that are usually preceded by disfluencies in spoken, spontaneous language.

A more complex way for a listener to predict via filled pauses is by considering the speaker. Indeed, integration of what is said with who says it is common in language comprehension (Berkum et al., 2008), and listeners' expectations appear to reflect stereotypes about speakers (Corps, Brooke, et al., 2022). Following the fact that the distributional pattern of filled pauses presumably has its origins in the speaker's production system, under the *tokens-inference* account, the disfluency-bias is explained as a paralinguistic, rather than as a linguistic (i.e., dependent on the knowledge of a language's distributional patterns), process - what some authors have referred to as perspective-taking. Predictions following a disfluency are based on listeners' knowledge of the causes for the speaker to be disfluent, such as difficulty in retrieving a low-frequency lemma. In this sense, what drives the disfluency bias is a causal inference between disfluency and the reasons for a speaker to be disfluent, which can be taken as a 'crystallised' form of learning, or as an inference that emerges when listeners encounter a filled pause. Supporting this account, the

disfluency-bias is attenuated when listeners believe the speaker has trouble with lexical production (Arnold et al., 2007), for non-native-accented speech (Bosker et al., 2014), or when the accessibility of candidate referents is low for the speaker (Barr & Seyfeddinipur, 2010).

Above we have outlined three main ways in which comprehension of spoken language could be impacted by filled pauses. Under the fillers-as-time view, the language comprehension system simply uses the time during which there is no propositional content to process to (continue to) predict upcoming items, as well as benefits from the re-orientation of attention to the speech signal. Under the fillers-as-tokens view, fillers are associated with contextually likely continuations; or, more radically, fillers act as signals to begin inferring what might have presented a cognitive load for the speaker. It is important to note that these views do not need to be mutually exclusive: The associative and the inference accounts can be seen as two ends of a continuum (Arnold et al., 2007).

As discussed in Chapter 3, prediction in speech comprehension can be understood as at least involving two mechanisms, one associative and one inference. The formalisation of these mechanisms is parallel to the token-associative and token-inference here described: The former involves associations arising from exposure to a language and subsequent learning of the statistical properties of the linguistic input, whilst the latter involves inferences following individuals' non-linguistic knowledge. Therefore, we can assume that the same properties of these two mechanisms apply to the ones here described. Namely, we can describe the token-associative account as automatic and thus effortless and cost-free, yet inflexible (inasmuch the system follows the learnt distribution without any specification), while the tokens-inference account can be seen as a non-automatic, effortful, and costly mechanism, but that in return allows the system to adapt to different scenarios (i.e., it is flexible).

To explore whether and how these mechanisms operate, this chapter presents two studies based on Bosker et al. (2014). To test the automaticity (in terms of processing costs) and the flexibility of the disfluency bias following the fillers-as-tokens account, we

use linguistic background as an attribute of speakers and listeners as a proving ground. In Experiment 1, we attempted to replicate Bosker et al.'s (2014) findings that support the idea that the disfluency bias can be flexible. In Experiment 2, we explored whether the disfluency bias is resistant to cognitive load. We selected Bosker et al.'s (2014) experiments because of the properties of the elements the authors found that a filled pause could anticipate. As we discussed in Section 2.2, a three-stage Leveltonian architecture of the speech production system presents three major points at which a speaker can struggle with production: They can struggle with what to say (conceptualisation), with what words to say it (formulation), or how to say it (articulation) (i.e., vulnerability points, Segalowitz, 2010). Most research has shown that listeners can anticipate objects that represent a struggle at the conceptualisation level (e.g., the difficulty in producing speech is due to referring to an object without a conventional label or due to referring to an object that has not been mentioned in the discourse, and the speaker has to decide how to refer to it; Arnold et al., 2004, 2007). It is important to note that biases towards discourse-new entities or objects without a conventional label do not require linguistic knowledge from the listener, as the cause for the speaker to be disfluent does not depend on language. In contrast, biases towards objects that can pose trouble at the formulation level do require linguistic knowledge from the listener due to the detail of the anticipation (cf. Corley & Hartsuiker, 2003). In this regard, Bosker et al. (2014) stands as the sole study exploring whether a disfluency can cue elements that can yield disfluencies due to problems at the formulation level (e.g., low-frequency words). Exploring Bosker et al. (2014) while manipulating listeners' linguistic background, as we will do in Experiment 2, allows us to understand whether the disfluency bias is also dependent on listeners' knowledge of the language, and thus aligning with the predictions of the fillers-as-tokens account, or if in these scenarios, the fillers-as-time account might be more accurate.

4.2 Experiment 1: The role of speaker identity

There is now recent evidence suggesting that what is expected due to the disfluency bias can also be reduced to difficulties in retrieving a lemma, i.e., at the formulation stage in Levelt's proposed architecture. Bosker et al. (2014) set out to explore the specificity and flexibility of the disfluency bias. For example, a speaker may struggle to find the label for an object because it has low accessibility (i.e., low-frequency words, Hartsuiker & Notebaert, 2010), which arguably is even harder to retrieve when a speaker is producing speech in their second language (Gollan et al., 2008). Bosker et al. (2014) explored whether the anticipations triggered by disfluencies can be as specific as trouble at the level of formulation by exploring the disfluency bias towards low-frequency words, and whether listeners can take into consideration the cause for the speaker to be disfluent, by presenting speech produced by a non-native speaker.

In their study, native Dutch listeners were visually presented high- and low-frequency items while following a speaker's instructions which could be fluent or disfluent (e.g. "*Klick op de* [target]"; Click on the [target] or "*Klik op uh de* [target]"; Click on uh the [target]). In their first experiment, instructions were produced by a native Dutch speaker; instructions in the second experiment were produced by a non-native Dutch speaker. Bosker et al. (2014) found that filled pauses produced by a native speaker elicited anticipatory eye movements towards low-frequency items (i.e., a disfluency bias), while this was absent for the non-native speaker counterpart. The authors attributed their findings to listeners' exposure to the distributional patterns of filled pauses in non-native speech, in which filled pauses are more arbitrary (De Jong, 2016; Davies, 2003) and thus reduces their value as predictive cues (an explanation that aligns with the tokens-associative view).

In Experiment 1, we set out to investigate the degree of flexibility of the disfluency-bias by replicating Bosker et al. (2014) in a different sample and language. In our experiment, a sample of native English listeners followed instructions provided by either a native or a non-native speaker (L1: Spanish), who would refer to one (out of two) items

on a computer screen fluently or disfluently (*Click on the* [target] or *Click on* ***thee uh*** [target]). Following Bosker et al. (2014), our participants should show anticipatory eye movements (i.e., prior to target onset) for disfluent instructions provided for the native speaker, but not for the non-native speaker.

4.2.1 Methods

All materials and data, including experimental and analysis scripts, can be found at https://osf.io/ct2ja/.

Participants

106 self-reported British English monolingual speakers took part in the study. Participants were aged between 18-30 years old, had normal or corrected-to-normal vision, and no hearing impairments. Participants were recruited from either the general population or the University of Edinburgh student pool. All participants gave their informed consent as approved by the University of Edinburgh PPLS Ethics Committee (reference number: 249-1819/3). Participants were compensated for their participation (either with an economic compensation - £5 - or with university credit) and were debriefed after the experiment.

We exceeded our desired sample size by four because some participants took part in exchange for university credit, and we did not want to prevent students from participating in experiments, as it is required for their degree. However, data from participants recruited after the desired sample size was reached is not included in the analysis (but data from all participants, including those not reported in the analysis, is reported online). Additionally, participants who reported 1) the aim of the study, 2) the manipulation, or 3) that the naturalness of the audio was below 4 were not included in the analysis. We identified six participants who were not included in the analysis. Our final sample size was 96 participants (48 in the native speaker condition, 48 in the non-native speaker condition, males = 19).

Materials

128 black and white line drawings from Weldon & Roediger, 1987 were used as visual stimuli. All pictures had high name agreement (H < 1; Weldon & Roediger, 1987). Drawings were grouped into 64 pairs comprised of one low-frequency and one high-frequency item. 32 pairs were selected as critical stimuli. Frequencies were obtained from the Johnston et al. (2010) database for British English. Low-frequency items had a log frequency of 0.41 (SD = 0.31) on average, and high-frequency items had a log frequency of 2.15 (SD = 0.31) on average ($t(62) = 22.56$, $p < .001$). Additionally, high-frequency words differ from low-frequency words in age of acquisition (HF = 2.09 (0.56), LF = 3.38 (0.94), $t(50.3) = -6.67$, $p < .001$), familiarity (HF = 6.2 (0.67), LF = 4.43 (1.15), $t(49.86) = 7.53$, $p < .001$), and length (HF = 4.6 (1.66), LF = 6.72 (1.8), $t(62) = -4.91$, $p < .0010$). The remaining 32 pairs were used for filler trials. The frequency of items mentioned (low- vs high-frequency items) and manner of delivery (fluent vs disfluent) were counterbalanced across four lists in a Latin Square design to ensure that all items were presented in all conditions.

A native British English speaker and a Spanish-English bilingual speaker were recorded for the auditory stimuli. Both speakers followed the script "Click on the [item]". As speakers were allowed to produce speech freely, they produced different types of hesitations. Following Bosker et al.'s (2014) procedure, we selected three fluent utterances and three disfluent utterances as templates for the critical trials, and then placed before the item. As a consequence, these templates differed in length by speaker and by template. Low- and high-frequency words differed in length in both the native (HF = 451.9 ms, LF = 576.1 ms, $t(62) = -5$, $p < .001$) and non-native speaker (HF = 452.4 ms, LF = 563.2 ms, $t(62) = -5.02$, $p < .001$). These differences in length between low- and high-frequency words were not statistically significant between speakers ($t(62) = 0.11$, $p = .91$ and $t(62) = 0.42$, $p = .68$ for high- and low-frequency words, respectively). Table 4.1 describes the length and maximum pitch of each segment of the critical carriers by speaker and manner of delivery. Each list included additional 32 filler utterances. Filler

utterances were recorded in their entirety. We aimed for a proportion of fluent:disfluent utterances of 3:1 for the low- and high-frequency items mentioned for filler trials similar to Bosker et al. (2014). This ensured that the experiment resembled natural speech (Fox Tree & Clark 1997).

Table 4.1

Duration (in ms) and pitch (in Hz) for each segment of the sentence templates employed as critical stimuli, by speaker's linguistic background and manner of delivery.

	Duration (ms)	Maximum pitch (Hz)
Native speaker		
Fluent		
Click	207, 204, 143	129, 113, 102
On	125, 107, 108	130, 108, 100
The	135, 99, 173	120, 100, 95
Disfluent		
Click	202, 174, 187	144, 149, 151
On	115, 156, 164	144, 158, 145
The	135, 172, 174	131, 131, 133
Uh	602, 795, 578	120, 113, 125
Non-native speaker		
Fluent		
Click	385, 319, 415	282, 467, 304
On	196, 204, 205	193, 206, 190
The	125, 135, 129	175, 188, 160
Disfluent		
Click	405, 308, 254	264, 349, 306
On	242, 216, 207	180, 210, 287
The	332, 379, 367	169, 183, 213
Uh	552, 575, 568	157, 157, 157

Procedure

Stimuli were presented on a 21" CRT monitor screen at a viewing distance of approximately 80 cm using OpenSesame v.3.2.7. (Mathôt et al., 2012). Participants' eye movements were recorded using a Tower-mounted SR Research EyeLink 1000 eye-tracker, sampling the right eye at 500 Hz. Mouse responses were collected with a standard computer mouse.

Participants were given a cover story at the beginning of the experiment. The experimenter explained that they would listen to a previously recorded participant who had given instructions to another participant regarding what item to click on a computer screen. They were told that the aim of the study was to investigate how visual and auditory information are processed when we have to follow instructions. This cover story was given to provide an explanation for the origin of the audio and why the speaker could stutter at times. Participants were instructed to click on the item mentioned as fast as possible and to move the mouse freely.

Eye-tracking began with a nine-point calibration and validation procedure, followed by an eight-trials practice session. Trials began with a drift correction, followed by a red fixation dot which lasted 500 ms at the centre of the screen. Items were then presented vertically centred on the left and right sides of the screen, the position of the item mentioned being counterbalanced across trials. After 1500 ms, audio was presented at a comfortable loudness level from two loudspeakers. Participants had 5000 ms once the audio was over to click on either object. If no response was recorded, the next trial began. Critical and filler trials were presented at random. Figure 4.1 depicts the trial sequence.

Afterwards, all participants filled in a post-experimental questionnaire to assess if they had noticed the manipulation. Questions included a 9-point scale rating on the audio's naturalness (i.e. if the voice sounded natural and/or not edited; 1: extremely unnatural, 9: extremely natural) and five open-ended questions about the experiment's

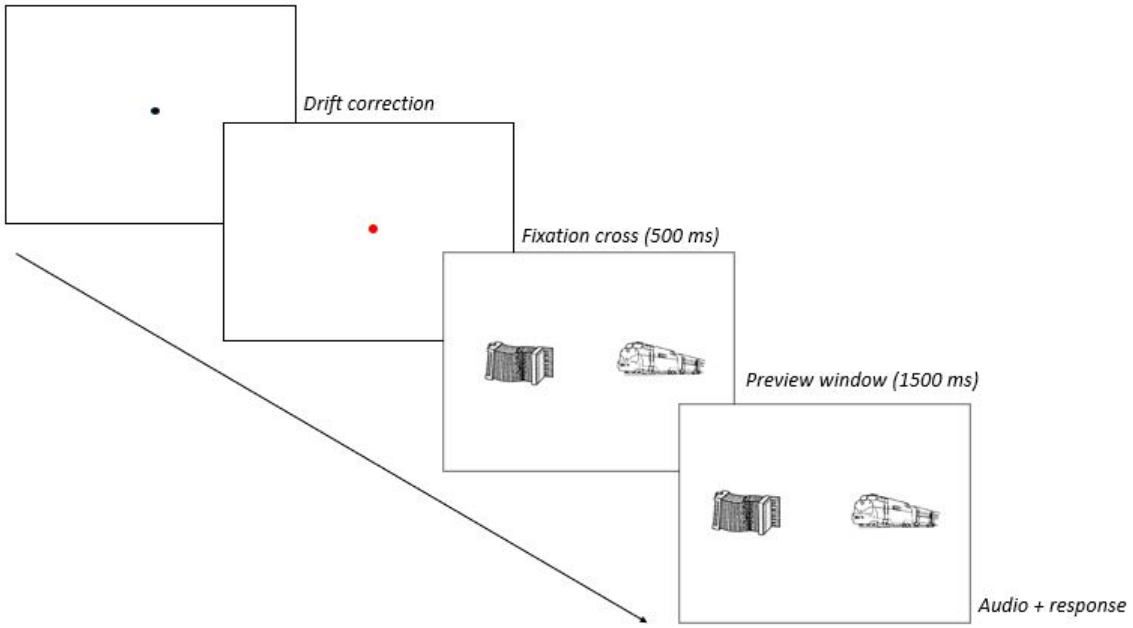

Figure 4.1. Trial sequence of Experiments 1 and 2.

stimuli and its aim. Participants who listened to the non-native speaker answered additional questions about the auditory stimuli (i.e. two 9-point scale ratings on perceived fluency and accentedness; 1: strongly disfluent, 9: strongly fluent; 1: strongly accented, 9: strongly unaccented). Additionally, these participants rated their exposure to foreign-accented English on two 9-point scales, measuring their daily interactions with second-language English speakers (1: never, 9: always) and their daily exposure to foreign-accented English (1: never, 9: always). Afterwards, participants were informed of the actual aim of the experiment and further asked if they had noticed the manipulation.

4.2.2 Results

All analyses were carried out in R version 4.2.0 (R Core Team, 2020), using the *lme4* (version 1.1.29., Bates et al., 2015) package. Data wrangling was done with the *tidyverse* (version 1.3.1., Wickham et al., 2019) package, and data visualization with *ggplot2* (version 3.3.6., Wickham, 2016) and *wesanderson* (version 0.3.6., Ram & Wickham, 2018) packages. Due to an error in presentation, data from three pairs had to be removed, as they overlapped phonologically (9.07% of trials). Data where no click was recorded or participants clicked on the wrong item were not included in the analysis (0.29% of trials).

Answers to questionnaire

On average, participants rated the naturalness of the audio as 8.04 (SD = 1.05) for the native speaker and 7.10 (1.26) for the non-native speaker ($t(94) = 3.36$, $p < 0.001$). Additionally, they rated the non-native speaker on average 6.73 (1.18) for fluency and 7.10 (1.31) for accentedness. Participants reported being exposed daily to non-native accented English at an average of 6.08 (2.34), and interacting daily with non-native speakers at 6.13 (2.36). These measures did not correlate with participants' behaviours and are not discussed any further.

Reaction time

On average, our participants' accuracy at selecting the right target was quite high (99.5% for the native, 99.67% for the non-native speaker). To investigate whether the presence of a filled pause eased recognition of the target (following the fillers-as-time account, Corley & Hartsuiker, 2011), we analysed participants' reaction times. Trials with reaction times that exceeded 2000 ms post-target offset were removed from analysis (0.07% of observations). Further, trials that exceeded 2 standard deviations by participant mean were not included in the analysis (4.74% of observations).Table 4.2 depicts participants' reaction times by manner of delivery, speaker's linguistic background, and referent frequency.

Table 4.2

Mean reaction times (ms) of participants' click on the referent for both the native and non-native speaker condition by manner of delivery and target's frequency (standard deviation in brackets).

	Native speaker RT (ms)	Non-native speaker RT (ms)
Fluent		
High-frequency	898 (201)	906 (215)
Low-frequency	946 (193)	962 (221)
Disfluent		
High-frequency	884 (182)	897 (200)
Low-frequency	940 (200)	936 (211)

We modelled participants' raw reaction times using a linear mixed model. The model included fixed effects for manner of delivery (fluent coded as -0.5, disfluent as +0.5), speaker's linguistic background (native coded as -0.5, non-native as +0.5), referent frequency (high-frequency coded as -0.5, low-frequency as +0.5) and their interactions. The maximal model (Barr et al., 2013) included by-participant and by-item random intercepts, with random slopes for manner of delivery, referent frequency, and their interaction by-participant, and random slopes for manner of delivery, speaker's linguistic background, and their interaction by-item. The maximal model failed to converge and thus was simplified by dropping the terms that explained the least variance. The final model included a random slope for manner of delivery by-participant, and random slopes for manner of delivery and speaker's linguistic background by-item. Results were deemed significant at $|t| > 2$ (Baayen, 2010).

The model showed that participants were faster at clicking on the referent when this was high-frequency ($\beta = 51.46$, SE = 12.52, t = 4.11), in line with the word-frequency effect (e.g., Howes & Solomon, 1951; Dahan et al., 2001). We found a not significant effect of manner of delivery in the expected direction ($\beta = 11.98$, SE = 6.83, t = 1.75) or speaker's linguistic background ($\beta = 8.61$, SE = 29.56, t = 0.29). Importantly, there was no interaction between frequency and manner of delivery ($\beta = 8.46$, SE = 13.51, t = 0.63). The interaction between frequency, manner of delivery, and speaker's linguistic background was close to significance ($\beta = 40.93$, SE = 22.04, t = 1.86). This lack of effect fails to replicate previous findings whereby the presence of a disfluency speeded up the selection only of low-frequency words (Bosker et al., 2014). We turn to an explanation as to why we did not replicate the effects of manner of delivery and speaker's linguistic background in the discussion.

Eye movements: Prediction Time window

Figures 4.2 and 4.3 depict participants' pattern of fixations for fluent and disfluent utterances respectively. The time course of fixations for fluent utterances does not seem to differ between the low- and the high-frequency item, regardless of the speaker condition.

In contrast, disfluent utterances present a contrasting pattern. Those listening to the native speaker started to fixate on the low-frequency item shortly after the onset of the filled pause, to then start fixating on the high-frequency item. In contrast, those in the non-native speaker condition increased their fixations towards either item shortly after encountering *the*. From *um* onset, fixations on the high-frequency item slightly decreased over time. These figures thus suggest that non-native disfluencies biased participants towards low-frequency items, while native disfluencies biased participants towards high-frequency items.

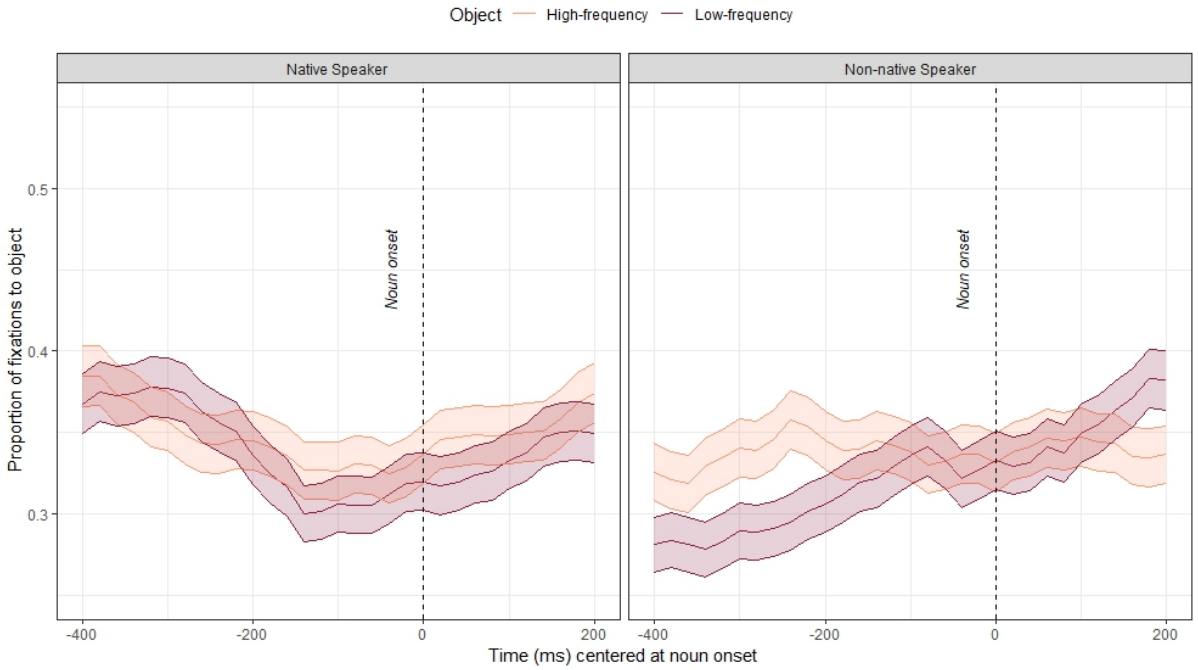

Figure 4.2. Mean proportion of fixations to high- (orange line) and low-frequency (red line) items in fluent utterances by speaker's linguistic background (native/non-native) over time. Proportions were calculated out of the sum of fixations towards either object for each 20-ms time bin. Shaded areas represent ± 1 standard error of the mean.

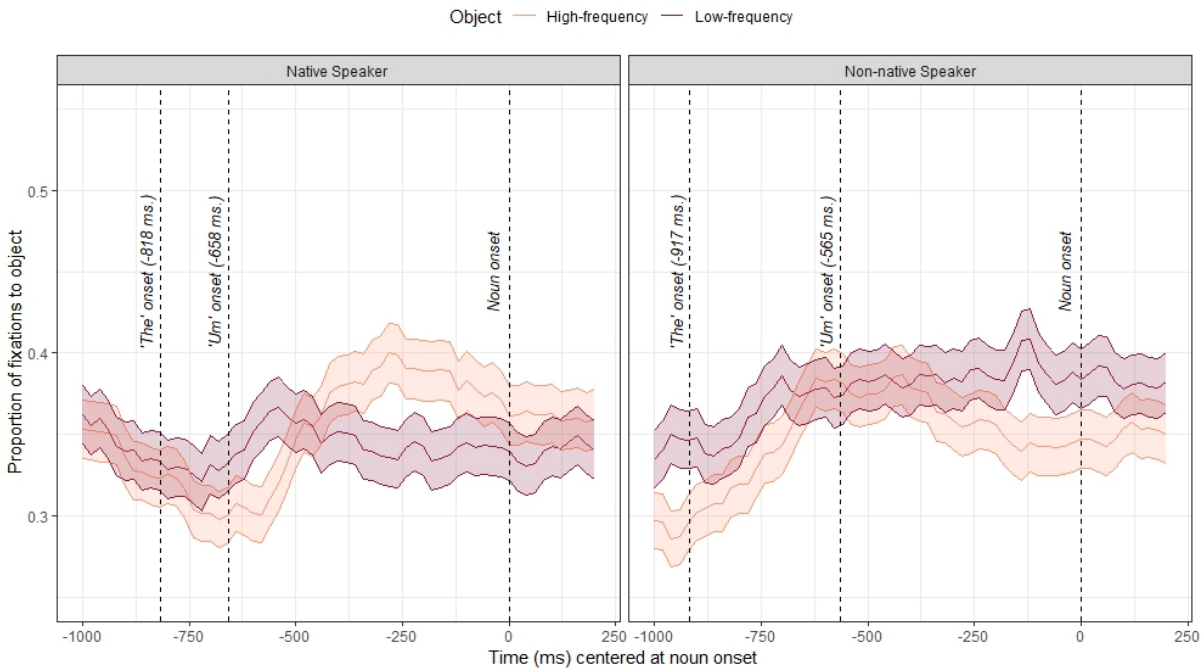

Figure 4.3. Mean proportion of fixations to high- (orange line) and low-frequency (red line) items in disfluent utterances by speaker's linguistic background (native/non-native) over time. Proportions were calculated out of the sum of fixations towards either object for each 20-ms time bin. Shaded areas represent ± 1 standard error of the mean.

We first analysed eye movements following Bosker et al.'s (2014) analysis. In two generalised linear mixed models (one per manner of delivery, due to the differences in length between fluent and disfluent utterances, cf. Bosker et al., 2014), we analysed participants' fixations on the low-frequency item. The time window of analysis began at *'the'* onset and ended at target onset. Both fluent and disfluent time windows were shifted by 200 ms post-target onset to account for the time it takes to launch a saccade (Matin et al., 1993).

We modelled fixations to the low-frequency item over time as a function of the speaker's linguistic background. To explore dynamic changes in participants' pattern of fixations, we included two polynomials as linear and quadratic time components (Mirman et al., 2008; Mirman, 2017). Both models included nativeness (native coded as 0, non-native as 1), linear time, and quadratic time, as well as the interaction between nativeness and each time component as fixed effects. For ease of convergence, we divided the linear and quadratic terms by 200. The models included random intercepts by-participant, by-item and by-sentence template.

Table 4.3 depicts the model estimates of each model. The first six parameters refer to the fixations towards the low-frequency item in fluent utterances. The model showed that participants were less likely to fixate on the low-frequency item at the *'the'* onset when listening to a native speaker (β = -0.93, SE = 0.19, z = -4.81, p <.001). This pattern did not change linearly (B = 0.08, SE = 0.14, z = 0.58, p = .56) or quadratically (B = 0.0003, SE = 0.0004, z = 0.63, p = .53). This pattern did not differ in the non-native speaker condition. Overall, this suggests that following fluent utterances, participants were not likely to fixate on the low-frequency item.

Table 4.3

Estimated parameters of two mixed effects logistic regression models (one per manner of delivery) of the looks to low-frequency objects, with fixed effects of speaker's linguistic background (native coded as 0, non-native as 1), a linear and a quadratic time component, and their interactions. Time window of analysis spanned from 'the' onset until 200 ms post-target onset.

	Estimate	Std. Error	z-value	p-value
Fluent utterances				
Intercept	-0.93	0.19	-4.81	<.001
Linear Time	0.08	0.14	0.58	.56
Quadratic Time	0.0002	0.0004	0.63	.53
Non-native Speaker	0.10	0.26	0.37	.71
Linear Time * Non-native Speaker	-0.10	0.18	-0.56	.58
Quadratic Time * Non-native Speaker	0.0001	0.0005	0.23	.82
Disfluent utterances				
Intercept	-0.83	0.16	-5.32	<.001
Linear Time	0.06	0.02	2.70	.007
Quadratic Time	-0.00007	0.00002	-3.15	.002
Non-native Speaker	0.03	0.21	0.15	.88
Linear Time * Non-native Speaker	0.07	0.03	2.15	.03
Quadratic Time * Non-native Speaker	-0.00002	0.00003	-0.58	.56

The next six parameters refer to fixations towards the low-frequency item in disfluent utterances. In this case, at the onset of a lengthened *'the'*, participants in the native speaker condition started off by not fixating on the low-frequency item (β = -0.83, SE = 0.16, z = -5.32, p <.001), followed by a linear increase (β = 0.06, SE = 0.02, z = 2.70, p = .007), followed by a quadratic decrease (β = -0.00007, SE = 0.00002, z = -3.15, p = .002), i.e., it followed a ∩ shape. This linear increase was slightly bigger in the non-native speaker condition (β = 0.02, SE = 0.21, z = 0.15, p = .03). This model suggests that there was a weak increase in fixations on the low-frequency item following a disfluency.

Analysing the time course of fixations binomially partially supports Bosker et al.'s (2014) findings. These results suggest that the presence of a disfluency, produced by either a native or a non-native speaker, increased fixations on the low-frequency item. It is important to note, however, that a binomial analysis only measures participants' preference for the low-frequency item, without measuring if it comes at the cost of a dispreference for the high-frequency item. While Bosker et al. (2014) took this measure as reflective of participants' update of predictions following a disfluency, modelling only fixations to the low-frequency item might only depict half of the picture, as suggested by Figure 4.3, given that their analysis does not compare participants' preference for the low-frequency item between fluent and disfluent items (because of the differences in time between the two conditions). Further, coding fixations binomially may yield different results when compared to analyses that consider instead the proportion of time fixating on each item (Ito & Knoerfele, 2022).

A better conceptualisation of the disfluency bias, in a situation where fluent and disfluent utterances cannot be compared, would be measuring the preference for the low-frequency item over the high-frequency one, a pattern not captured by a binomial analysis. Therefore, we conducted a separate analysis where we calculated the fixation proportions towards either object per bin and transformed them into empirical logits (Barr, 2008) to measure this preference (i.e., a low-frequency advantage): Positive numbers index a preference for the low-frequency item, while negative numbers index a preference for the high-frequency item. Figures 4.4 and 4.5 represent the low-frequency advantage modelled to help mapping the model estimates. We analysed this preference to fixate on the low-frequency item over the high-frequency one in two linear mixed models, one per manner of delivery, in the same time windows as our previous analysis. Each model included fixed effects for nativeness (native coded as 0, non-native as 1), linear and quadratic time components, and the interaction between nativeness and each time component. Each model included random intercepts by-participant, by-item and by-sentence template. We considered effects significant at $|t| > 2$ (Baayen, 2008).

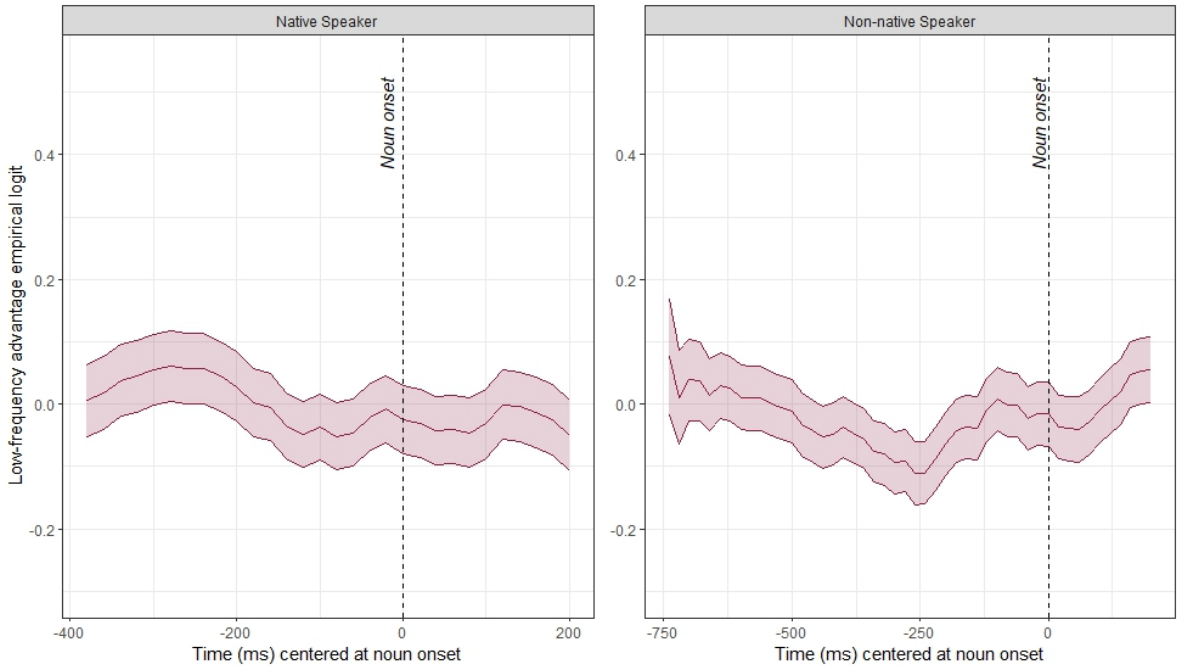

Figure 4.4. Mean low-frequency advantage in empirical logits in fluent utterances by speaker's linguistic background from audio onset to 200 ms post-target onset. Positive values index a preference to fixate on the low-frequency object, and negative values index a preference to fixate on the high-frequency object. Shaded areas represent ± 1 standard error of the mean.

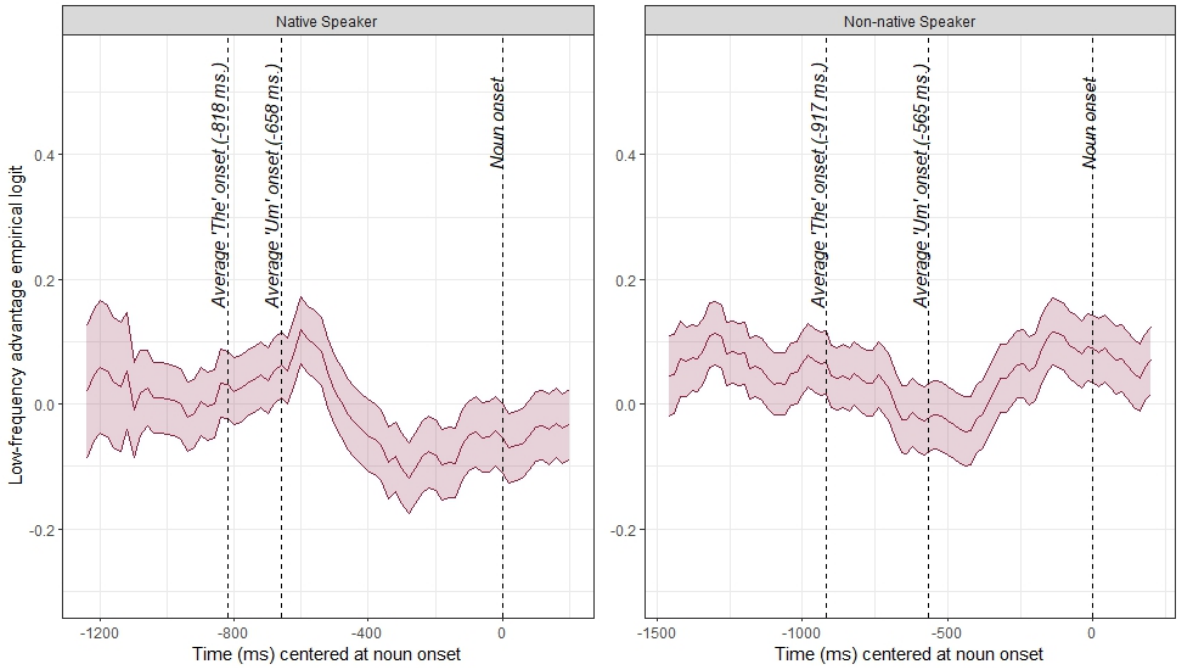

Figure 4.5. Mean low-frequency advantage in empirical logits in disfluent utterances by speaker's linguistic background from audio onset to 200 ms post-target onset. Positive values index a preference to fixate on the low-frequency object, and negative values index a preference to fixate on the high-frequency object. Shaded areas represent ± 1 standard error of the mean.

Table 4.4 depicts the result of these two models. For fluent utterances, there was no preference towards either item over time when listening to the native speaker, either linearly (β = 0.02, SE = 0.09, t = 0.22) or quadratically (β = -0.00005, SE = 0.0003, t = -0.20). This pattern did not differ in the non-native speaker condition (β = -0.02, SE = 0.11, t = -0.15; β = 0.00006, SE = 0.0003, t = 0.19 respectively).

Table 4.4

Estimated parameters of two linear mixed effects regression models (one per manner of delivery) on the preference to fixate on the low-frequency item via empirical logits, with fixed effects of speaker's linguistic background (native coded as 0, non-native as 1), a linear and a quadratic time component, and their interactions. Time window of analysis spanned from 'the' onset until 200 post-target onset.

	Estimate	Std. Error	t-value
Fluent utterances			
Intercept	-0.04	0.09	-0.46
Linear Time	0.02	0.09	0.22
Quadratic Time	-0.00005	0.0003	-0.20
Non-native Speaker	0.04	0.11	0.38
Linear Time * Non-native Speaker	-0.02	0.11	-0.15
Quadratic Time * Non-native Speaker	0.00006	0.0003	0.19
Disfluent utterances			
Intercept	0.11	0.06	1.95
Linear Time	-1.22	0.02	-6.62
Quadratic Time	0.0001	0.00002	5.60
Non-native Speaker	-0.08	0.07	-1.08
Linear Time * Non-native Speaker	0.11	0.02	4.44
Quadratic Time * Non-native Speaker	-0.00007	0.00002	-3.34

Disfluent utterances produced by the native speaker were characterised by a linear decrease in the low-frequency advantage (β = -1.22, SE = 0.02, t = -6.62), followed by a quadratic increase (β = 0.0001, SE = 0.00002, t = 5.60). In contrast, in the non-native speaker condition, the preference to fixate on the low-frequency item increased

linearly (β = 0.11, SE = 0.02, t = 4.44) and decreased quadratically (β = -0.00007, SE = 0.00002, t = 3.34). This suggests that the low-frequency preference following a disfluency increased quadratically for those attending to native-accented speech, and linearly for those attending to non-native-accented speech.

The present analyses highlight how different conceptualisations of eye movements yield different interpretations. Bosker et al. (2014) and the binomial analysis presented here explored whether participants fixate on a low-frequency item depending of the utterance's manner of delivery. When coding fixations binomially, similarly to Bosker et al. (2014), the results of our analysis replicate their results: The model suggests that there was a linear increase, albeit weak, in fixations on the low-frequency item over time. As we discussed, a more appropriate manner of measuring the potential bias exerted by a filled pause is by modelling a preference for the low-frequency item over the high-frequency one (via empirical logits). In this analysis, we failed to replicate Bosker et al.'s (2014) pattern of fixations following a native disfluency: Our results suggest that fixations towards the low-frequency item increased quadratically, instead of linearly.

Eye movements: Word recognition

Analysis of eye movements in our prediction window revealed two unexpected findings: (1) that non-natives' disfluencies (as opposed to natives') increased the preference to fixate on the low-frequency linearly, and (2) increased the preference to fixate on the high-frequency linearly, with the preference for the low-frequency item emerging quadratically. We decided to conduct a follow-up analysis to whether the presence of a filled pause impacted word recognition, that is, in a time window when listeners had encountered phonological information about the target. It could be possible that the presence of a filled pause eased the recognition of the upcoming word (in line with the fillers-as-time account), and that this benefit is larger for a low-frequency item (following a fillers-as-tokens account).

We thus conducted an analysis on the time window after target onset. Following Bosker et al.'s (2014) supplementary analyses, we defined our window of analysis based on visual inspection of participants' fixations: We selected a time window elapsing from target onset until participants' fixations suggested that they had fixated on the target (i.e., when fixations on the target reached an asymptote). In our case, this occurred within 760 ms post-target onset (see Figure A.1 for the target advantage, as defined below, and Appendix A, section A.2.1 for a visualization of the raw fixation probabilities). This time window covers the length of the longest word to produce: Reducing the interest period prior to word offset may mask any effects of interest, and it may be preferable to analyze longer time windows to capture delayed effects (Ito & Knoerfele, 2022).

The utterance's target phonological information was available for participants in this time window. As we are interested in the ease with which this information is recognised, we modelled the preference to fixate on the target (as opposed to the preference to fixate on the low-frequency item) using empirical logits (Barr, 2008), i.e., the target advantage: Positive values indicate fixations on the target, and negative values indicate fixations on the distractor. The model included factors of nativeness (native coded as 0, non-native as 1), manner of delivery (fluent coded as 0, disfluent as 1), frequency (high-frequency coded as 0, low-frequency as 1), linear and quadratic time components, and their interactions (bar interactions between time components themselves). For ease of convergence, we divided the linear and quadratic terms by 200. We included random intercepts by-participant, by-item and by-sentence template.

Due to the model's complexity, we will describe the parameters of Table 4.5 in relation to Figure 4.7. Starting with the native speaker condition (top panel of Figure 4.7), parameters A to C capture the preference to fixate on high-frequency targets produced fluently (orange, dashed line): In this case, the target advantage increased linearly (parameter B), with a weak quadratic increase (parameter C). When high-frequency targets were produced disfluently, the target advantage increased less linearly and more quadratically than its fluent counterparts (parameters H and I; orange, solid line in

Figure 4.7). This suggests a slight disadvantage in recognising high-frequency targets when they were preceded by a disfluency, as can be seen in the comparison between the solid (disfluent) and dashed (fluent) orange lines. The target advantage when it was a low-frequency word produced fluently was characterised by a weaker linear increase and a more quadratic increase than its high-frequency counterparts (parameters E and F, dashed, red line), what likely reflects the word-frequency effect, whereby word frequency affects the early stages of lexical access so that high-frequency words are recognised faster (Brysbaert et al., 2011; Dahan et al., 2001). The preference to fixate on low-frequency targets when produced disfluently, however, increased more linearly (parameter K, solid red line). Visual inspection of Figure 4.7 suggests that the presence of filled pauses facilitated the recognition of these targets in contrast to when they were produced fluently (comparison between the solid and the dashed red lines).

Moving on to the non-native speaker condition, the preference to fixate on high-frequency targets produced fluently increased less linearly and more quadratically than in the native condition (parameters N and O). Visual comparison between the top and bottom panels of Figure 4.7 (depicting the native and non-native speaker conditions, respectively) suggests that the overall recognition of targets produced by a non-native speaker was characterised by a more quadratic increase, possibly attributable to reduced intelligibility due to a non-native accent. The target advantage for high-frequency targets produced disfluently did not differ from that of the native speaker (parameters S and O, solid orange line): Visual exploration of Figure 4.7 suggests that the recognition of high-frequency words after a disfluency was impaired (comparison between the solid and the dashed orange lines). The preference to fixate on low-frequency targets produced fluently increased slightly more linear and less quadratically than its native counterpart (parameters Q and R). Importantly, participants were more likely to fixate at the offset on the target when it was a low-frequency word produced disfluently (parameter V), likely a spillover of the effects found in the prediction time window. The growth of the target advantage did not differ from that in the native speaker condition (parameters W and

Z), suggesting that the presence of a filled pause aided the recognition of low-frequency words.

Table 4.5

Estimated parameters of a mixed effects model for target preference measured as empirical logits in a time window from target onset to 760 ms post-target onset, with speaker's linguistic background (native coded as 0, non-native as 1), target frequency (high-frequency coded as 0, low-frequency as 1), manner of delivery (fluent coded as 0, disfluent as 1), linear and quadratic time components, and their interaction as fixed effects.

	Estimate	Std. Error	t value
A. Intercept	-0.14	0.07	-2.05
B. Linear Time	0.17	0.04	4.01
C. Quadratic Time	0.0005	0.0001	9.23
D. Low-frequency	0.03	0.07	0.38
E. Low-frequency * Linear Time	-0.20	0.06	-3.45
F. Low-frequency * Quadratic Time	0.0002	0.0001	3.22
G. Disfluent	-0.02	0.08	-0.28
H. Disfluent * Linear Time	-0.12	0.06	-2.03
I. Disfluent * Quadratic Time	0.0002	0.0001	2.72
J. Disfluent * Low-frequency	-0.13	0.07	-1.95
K. Disfluent * Low-frequency * Linear Time	0.29	0.08	3.49
L. Disfluent * Low-frequency * Quadratic Time	-0.0003	0.0001	-3.28
M. Non-native Speaker	0.16	0.09	1.80
N. Non-native Speaker * Linear Time	-0.37	0.06	-6.29
O. Non-native Speaker * Quadratic Time	0.0004	0.0001	5.98
P. Non-native Speaker * Low-frequency	0.01	0.07	0.15
Q. Non-native Speaker * Low-frequency * Linear Time	0.20	0.08	2.47
R. Non-native Speaker * Low-frequency * Quadratic Time	-0.0004	0.0001	-3.47
S. Non-native Speaker * Disfluent	-0.01	0.11	-0.11
T. Non-native Speaker * Disfluent * Linear Time	0.04	0.08	0.53
U. Non-native Speaker * Disfluent * Quadratic Time	-0.0001	0.0001	-1.00
V. Non-native Speaker * Disfluent * Low-frequency	0.21	0.09	2.24
W. Non-native Speaker * Disfluent * Low-frequency * Linear Time	-0.17	0.12	-1.46
Z. Non-native Speaker * Disfluent * Low-frequency * Quadratic Time	0.0002	0.0001	1.51

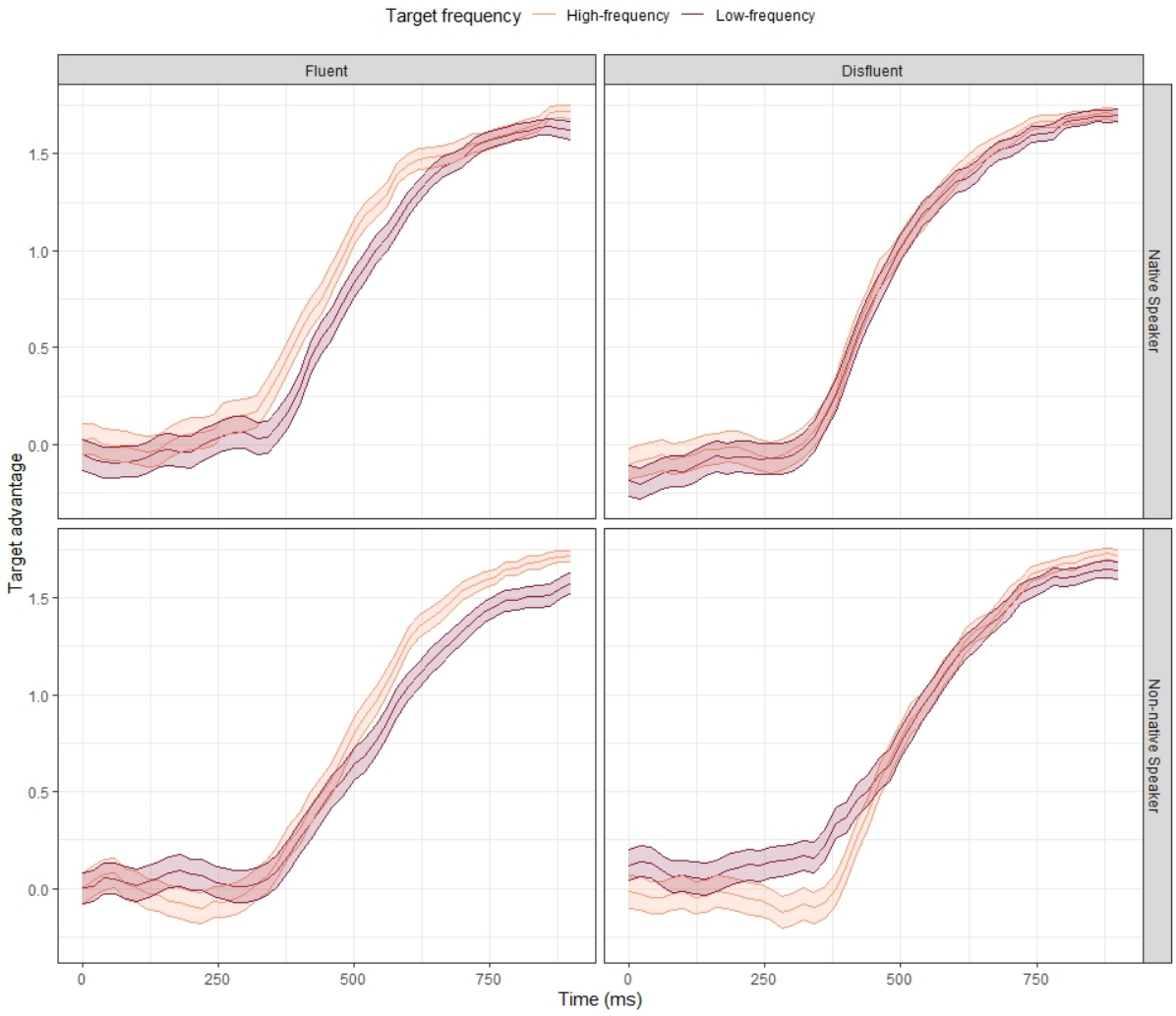

Figure 4.6. Mean target advantage by target frequency (red: low-frequency, orange: high-frequency) by manner of delivery and and speaker's linguistic background (native/non-native). Target advantage was calculated via empirical logits, where positive values indicate a preference to fixate on the target, and negative values a preference to fixate on the distractor. Time window of analysis spanned from target onset to 760 ms post-target onset. Shaded areas represented ± 1 standard error of the mean.

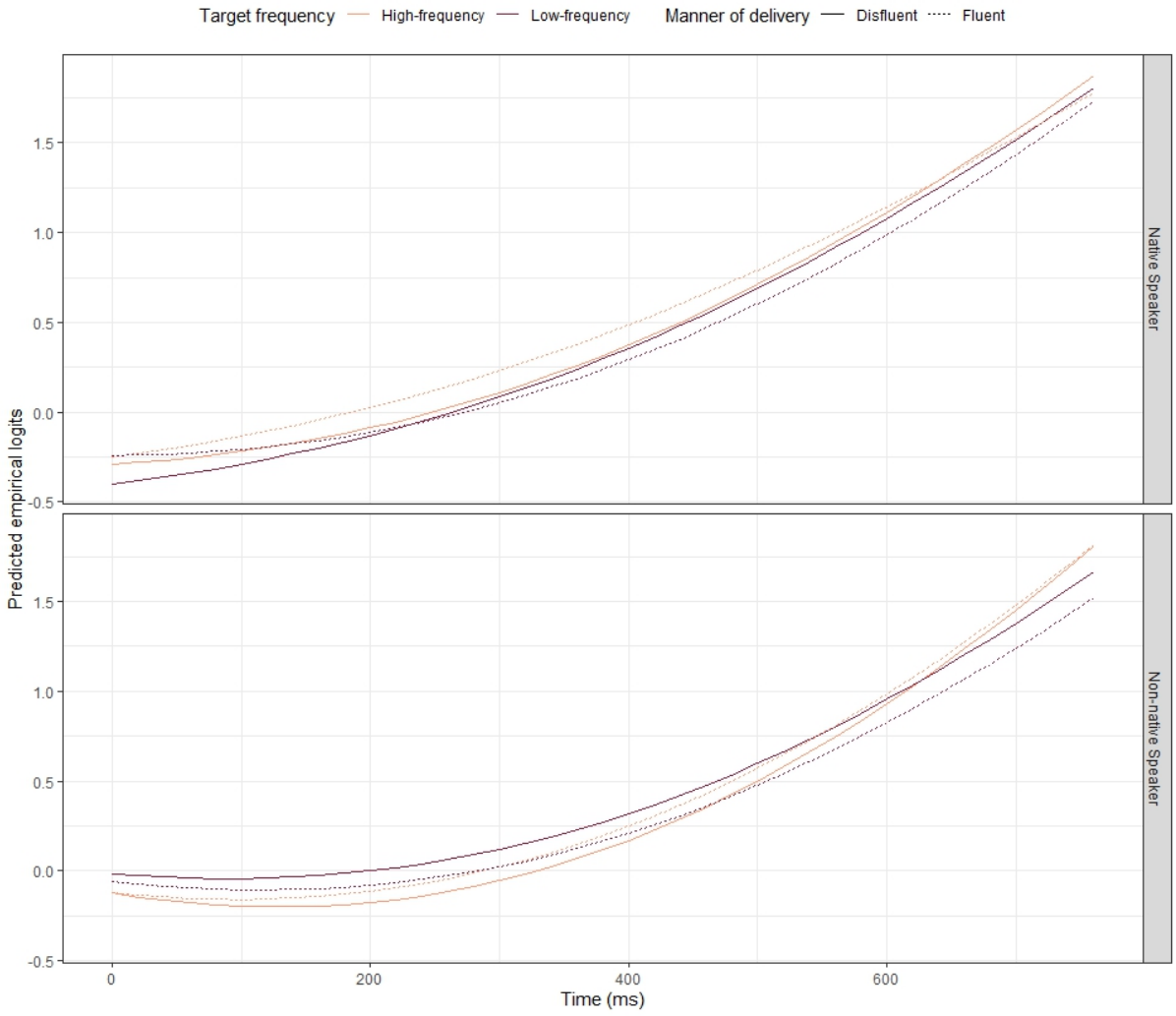

Figure 4.7. Growth Curve Analysis model fit of empirical logits to the low-frequency (red) and high-frequency (orange) items as a function of manner of delivery (dashed line: fluent, solid line: disfluent) and speaker's linguistic background (native/non-native).

4.2.3 Discussion

The findings of Experiment 1 are partially at odds with previous findings suggesting that native's, but not non-native's disfluencies, lead to anticipatory eye movements towards objects with low-frequency labels (Bosker et al., 2014). When listeners encountered a disfluency, we found evidence for anticipatory eye movements towards such objects if the speaker was non-native, but not for the native speaker. Subsequent analysis in a time window reflecting word recognition showed that the presence of a disfluency, regardless

of the speaker's linguistic background, facilitated the recognition of the low-frequency words and negatively affected the recognition of high-frequency words. The results of this secondary analysis partially align with those reported by Bosker et al. (2014), with the difference that in Experiment 1, non-native disfluencies affected word recognition similarly to native ones. However, we failed to find an effect of manner of delivery on reaction times, in contrast to Bosker et al. (2014). In this discussion, we will first focus on the differences between our results and Bosker et al.'s (2014) in the time window prior to target onset (reflecting prediction), to then move our discussion to the effects of manner of delivery once the target has been encountered.

The most striking finding of Experiment 1 is the lack of anticipatory eye movements towards low-frequency items following a native disfluency. In fact, our analysis suggests a dispreference to fixate on these items, which is not only at odds with Bosker et al. (2014), but also with a large body of research suggesting that disfluencies bias listeners' anticipations towards elements that are contextually associated with disfluency (e.g., harder to name objects, Arnold et al., 2004, Watanabe et al., 2008; discourse-new elements, Arnold et al., 2007, Heller et al., 2015). There are two potential explanations for this difference, implicating disparities between our analysis and that of Bosker et al. (2014), as well as the qualities of the auditory stimuli employed.

The most direct argument revolves around how the disfluency bias was conceptualised. In the original study, Bosker et al. (2014) explored fixations on low-frequency items prior to target onset binomially. As the authors themselves note 'we interpret a higher probability of fixations on low-frequency objects as a preference for low-frequency references, without excluding the possibility that the same result may be accounted for by a dispreference for high-frequency referents (Bosker et al., 2014, p. 108). In fact, when the disfluency bias is conceptualised as Bosker et al. (2014) did, we also found a preference to fixate on low-frequency targets following a native disfluency. Visual exploration of participants' pattern of fixations suggested that this approach might not be representative of the data: In fact, native disfluencies led to an increase in fixations towards the

high-frequency item. We argued that a better conceptualisation of the disfluency bias requires a dispreference towards high-frequency items because predicting a low-frequency word should come at the expense of not expecting a high-frequency word. Analysing participants' patterns of fixations with this conceptualisation (via empirical logits) showed that a native speaker's disfluencies led to a linear low-frequency disadvantage. At face value, this striking difference due to the disparity between the analyses casts some doubts on the original claims put forward by Bosker et al. (2014).

The second explanation concerns the qualities of our native speaker's filled pause. Bosker et al. (2014) reported filled pauses of around 900 ms of duration, which is well within the range of the duration of disfluencies employed in other studies (e.g., 1.3 s, Arnold et al., 2004). Additionally, previous studies reported that disfluencies were characterised by changes in length and prosody prior to the production of the filled pause (Arnold et al., 2003). The combination of these features in disfluent utterances might have made comprehenders notice that the speaker was about to be disfluent: Indeed, a filled pause is the obvious sign of a disfluency, but its surroundings are affected by its presence (cf. Section 2.1). In contrast, the length of the filled pause of our native speaker was 658 ms on average, with the length of the prolonged article not differing from that when the native speaker was fluent (160 ms and 135 ms on average, respectively). Therefore, in Experiment 1, those listening to the native speaker may have had less evidence of the speaker's hesitation than those listening to the native speaker. The overall preference to fixate on the high-frequency item might reflect listeners' anticipations given the lack of disfluency, with the quadratic increase reflecting the delayed effect of perceived disfluency. Since in fluent utterances the target name immediately follows the words 'Click on the', there is no obvious timeframe in which predictive looks to the high-frequency item might be detected.

The next difference between Experiment 1 and Bosker et al. (2014) concerns the effects of non-native disfluencies on prediction. Bosker et al. (2014) reported that listeners did not display anticipatory eye movements towards low-frequency words when

they encountered a filled pause produced by a non-native speaker. The authors took this attenuation as reflecting listeners' experience with non-native-accented speech, i.e., a more arbitrary pattern of disfluencies which reduced their predictive value, in line with the fillers-as-tokens account; and, specifically,the tokens-associative account. In our case, our participants did display anticipatory eye movements towards such elements. It is interesting to note that if Bosker et al.'s (2014) findings are attributable to individuals' learning of non-native-accented speech properties we should have found a similar effect unless there is a qualitative difference between the non-native speakers that participants in Bosker et al. (2014) are exposed to and the ones the participants in Experiment 1 are. Indeed, Bosker et al.'s (2014) participants reported being exposed to non-native-accented speech an average of 3.83 (SD = 2.13) on a 9-point scale; in contrast, our participants reported being exposed to non-native-accented speech and interacting with non-native speakers an average of 6.08 (SD = 2.34) and 6.13 (SD = 2.36). This suggests that a tokens-associative account might not fully explain Bosker et al.'s (2014) findings. As participants in Experiment 1 had more exposure than those in Bosker et al. (2014) to the non-arbitrary distribution of filled pauses in non-native-accented speech, yet displayed the disfluency bias, this entails that if the tokens-associative account can explain Bosker et al. (2014) we should have found similar findings.

An alternative possibility involves the similarity between our findings and previous studies exploring predictive processing upon encountering disfluent speech from individuals stereotyped to produce more filled pauses. For example, disfluencies produced by an old person still bias listeners' fixations towards discourse-new elements in a pattern similar to that arising following disfluencies produced by a young person, even when listeners hold the belief that the elderly are more disfluent (Saryazdi et al., 2021). The difference between our findings and Bosker et al.'s (2014) could lie in the amount of exposure listeners had to non-native-accented speech in their daily lives. Speculatively, as experience with a particular property of a signal increases (e.g., a foreign accent), the system habituates to this 'deviation' and might rely more on associations - what could also explain Saryazdi et al.'s (2021) findings. Indeed, research has shown that as listeners'

experience with non-native-accented speech increases, the system is more likely to engage in predictive processing (Porretta et al., 2017, Porretta et al., 2020). This explanation suggests that one factor that can increase the reliance on an associative mechanism is listeners' experience with the properties of the signal. This would suggest that Bosker et al. (2014) findings would actually align with a tokens-inference account: Listeners' stereotypes about non-native speakers' linguistic competence renders the disfluency a useless cue, because they are likely to experience more difficulties in speech production.

Analysis of a time window post-target onset demonstrated that both native and non-native disfluencies affected speech comprehension. Specifically, and partially in line with Bosker et al.'s (2014) supplementary analysis, we found that the presence of a filled pause facilitated the recognition of a low-frequency word and that, in fact, it impaired the recognition of high-frequency words. These patterns align with the fillers-as-tokens account, as they suggest that the benefit induced by a filled pause (in the form of facilitated recognition of subsequent linguistic units) depends on the congruency between manner of delivery and the properties of the element that follow. It is important to note, however, that this benefit did not cascade onto participants' reaction times as in Bosker et al. (2014). Arguably, this might be due to the lack of anticipatory processes in the native speaker condition, and the reduced intelligibility of our non-native speaker. It is possible that the reduced reaction time in selecting low-frequency items following a disfluency in Bosker et al. (2014) is due to the pre-activation of these elements.

The results of Experiment 1 suggest that disfluencies can affect speech comprehension, at least at the level of word recognition. The fact that disfluencies produced in non-native-accented speech led to similar effects to those in native-accented speech, contrary to Bosker et al. (2014), fails to support an association account for the disfluency bias in the form of learning probabilistic cues, as non-native-accented speech is more disfluent. This pattern of results opens the possibility of whether the association is not about the distributional properties of filled pauses themselves, but a crystallization of a causal inference: Listeners learn that disfluencies are commonly associated with difficulties in

speech production, and automatically attribute the origin of the problem when the set of potential referents is available to them. This possibility is supported by the speed with which the effects of disfluency constrained word recognition. A potential explanation for the differences between our results and those of Bosker et al. (2014) has to do with listeners' exposure to different signals: The system might follow this learnt association unless something in the speech signal triggers novelty (e.g., an accent) and thus considers the speaker's identity, and possibly resorts to an inference. However, these two possibilities cannot be answered by the present experiments. In what follows, we explore whether the bias towards low-frequency words can emerge in a scenario where listeners have less experience with the linguistic input, and arguably fewer cognitive resources to engage in inferences, but who may be more sensitive to speaker identity: non-native listeners.

4.3 Experiment 2: The role of listener identity

Non-native listeners also have to face disfluent speech, produced by both native and non-native speakers. Most research exploring how disfluent speech affects second-language comprehension has relied on offline measures, where non-native listeners are asked to transcribe disfluent speech and the number of errors (or lack thereof) is taken as an index of comprehension. Voss (1979) reported that speech including a range of hesitation phenomena, including filled pauses, led to perceptual problems in non-native listeners as reflected in the increase of errors in their transcriptions: A filled pause was taken as part of a word, or words were taken as hesitation phenomena (see Freedle & Kostin, 1999; Griffiths, 1991, for similar findings) - suggesting that, if anything, disfluencies impair second-language comprehension.

In contrast, other studies have found that disfluencies are beneficial in second-language comprehension, as long as non-native listeners recognise filled pauses as such. Blau (1991) found that speech peppered with hesitation phenomena, such as discourse markers and filled pauses, improved non-native listeners' comprehension of speech, as evidenced by their answers to questions about the content of the discourse (see also

Buck, 2001; Carney, 2022; for similar findings). Overall, this line of research suggests that filled pauses can aid comprehension following the predictions of the fillers-as-time account (Bloomfield et al., 2010): Non-native listeners may benefit from an interruption to the speech signal due to the increased difficulties associated with comprehending a non-native language. However, due to the methodologies employed, it is difficult to explain whether these effects are due to facilitated word segmentation or to the attention-orientating effects of filled pauses.

Alternatively, filled pauses could also guide comprehension by informing listeners' anticipations of what will follow, in line with the fillers-as-tokens account. There is, however, little evidence regarding the effects of disfluencies on the online processing of second-language comprehension. Watanabe et al.'s (2008) experiments stand as the sole study that partially tackles this question. In their study, the authors investigated whether filled pauses affected native and non-native listeners' anticipations about the complexity of upcoming elements, as measured by their reaction times in selecting the appropriate target. In a paradigm similar to that of Arnold et al. (2007), participants saw pairs of items consisting of a simple shape (e.g., a triangle) and a compound shape (e.g., a triangle with two arrows attached). Following Arnold et al.'s (2007) reasoning, a filled pause should lead listeners to anticipate a compound shape, because it is harder to refer to these objects in comparison to a simple shape.

Watanabe et al. (2008) found that native Japanese listeners were faster at selecting compound shapes when the speaker was disfluent. Non-native Japanese listeners (L1: Chinese) displayed different patterns depending on their proficiency. Proficient listeners' behaviours were similar to their native counterparts, while novice listeners' reaction times were not impacted in any direction by the presence of a disfluency. Intermediate listeners, however, showed a mixed pattern: They were faster at selecting the compound shape if the speaker was disfluent, but the presence of a disfluency before the speaker referred to a simple shape increased their reaction times.

These findings are consistent with the fillers-as-tokens account. If a delay itself facilitates object recognition, according to the fillers-as-time account, novice listeners would have shown faster reaction times following disfluent utterances. The differences between proficient and intermediate listeners can be understood in terms of prediction error: Filled pauses cued the wrong expectations when they preceded simple constituents, and recovery from this prediction error was harder for intermediate than for proficient listeners. Watanabe et al.'s (2008) explained their findings in terms of the tokens-associative account: The more exposed non-native listeners had been to the target language, the more likely it was for the disfluency bias to arise. This would be in line with theories where non-native comprehenders' use of cues to guide their comprehension is not as dependent on proficiency as it is on exposure (Dussias & Sagarra, 2007): In Watanabe et al. (2008), non-native participants' proficiency was measured by their length of residence in Japan. Watanabe et al.'s (2008) findings speak to the fact that non-natives' expectations following a disfluency are sensitive to language-extrinsic causes for the presence of a disfluency (i.e., describing an object that requires a longer description).

This leaves the question open of whether non-native listeners can anticipate objects whose labels are difficult to produce due to their accessibility (i.e., language-intrinsic). It is possible that this finer-detailed expectation is not available in second-language comprehension. Indeed, the word-frequency effect is exacerbated in non-native comprehension. Diependaele et al. (2013) showed that non-native comprehenders' reduced exposure to the target languages maximises the frequency differences between lemmas, so that the representation of low-frequency words is weaker and thus non-native comprehenders take longer to identify them (see also Gollan et al., 2008; Whitford & Titone, 2012). This opens the question of whether non-native comprehenders' associations (or inferences) about disfluencies can be this specific. Answering this question not only can inform our understanding of how disfluencies can impact the online processing of second-language comprehension, but more generally, it can shed light on the general mechanisms underlying the disfluency bias: Particularly, it can show whether the specificity reported in Bosker et al. (2014) depends on the amount of exposure to a language.

It is important to note that the characteristics of second-language comprehension can shed light on how effortless it can be to follow this association. In Section 3.2.2, we argued that predictions in second-language comprehension due to associations can emerge in patterns similar to those of first-language while inferences may be delayed due to reduced cognitive resources and automaticity (Ito & Pickering, 2021; Pickering & Gambi, 2018). This would translate into time courses for the disfluency bias differing between first- and second-language comprehension, depending on what mechanism underlies it. We further argued that non-native listeners tend to rely more on cues they have mastered well: While following an association can be automatic, non-native comprehenders may be also more sensitive to the speaker's identity (Futrell & Gibson, 2017). In turn, comprehending one's second language riddled with disfluencies may differ depending on who produces the filled pause.

In Experiment 2, we extended Experiment 1 to a sample of non-native English listeners. Participants heard instructions provided by either a native or a non-native English speaker who would refer to one (out of two) objects either fluently or disfluently. Contrary to previous studies on second-language comprehension, we did not test a homogeneous sample of non-native listeners. While homogeneity in participants' first language is desirable when the predictive cue in the non-native language is language-specific (e.g., gender marking), filled pauses are ubiquitous to all languages. Secondly, and most importantly, previous studies have shown that the different phonetic realisations of filled pauses across language do not impact their potential as a cue for upcoming items for bilinguals (Morin-Lessard & Byers-Heinlein, 2019). Finally, what is of relevance to us is our participants' proficiency in English (following Watanabe et al., 2008). Here, we tested a sample of bilinguals immersed in an English-speaking environment and gathered additional measures of their proficiency. We additionally measured daily interactions and exposure to native and non-native-accented speech to control for the role of experience.

4.3.1 Methods

All materials and data, including experimental and analysis scripts, can be found at https://osf.io/ct2ja/.

Participants

109 self-reported L2 English speakers took part in the study. Participants were aged between 18-30 years old, had normal or corrected-to-normal vision, and had no hearing impairments. Participants were recruited from either the general population or the University of Edinburgh student pool. All participants gave their informed consent as approved by the University of Edinburgh PPLS Ethics Committee (reference number: 249-1819/3). Participants were compensated for their participation (either with economic reimbursement - £5 - or with university credit) and were debriefed after the experiment. Similarly to Experiment 1, our recruitment method led to a larger sample size than desired (we tested one participant more than necessary): In the reported analysis, we only included participants who met our criteria up to our planned sample size.

To determine participants' eligibility, we included responses to the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian et al., 2007). Participants were excluded if they listed English as their first acquired language. Previous studies have used 6 years old as a cutoff for age of acquisition; in our case, we excluded participants if their usage of English was higher than 20% during childhood in comparison to their other languages. This was chosen to compensate for the fact that in many countries children are exposed to English in kindergarten. Based on these criteria, we identified and excluded five participants.

Additionally, we excluded participants if they reported 1) the aim of the study, 2) the manipulation, or 3) that the naturalness of the audio was rated lower than 4 in the post-experimental questionnaire. We further removed data from three participants due to their ratings. This resulted in a sample size of 96 participants (48 per speaker condition, males = 20) (see 4.6 and Appendix A, Section A.1).

Table 4.6

Participants' mean proficiency (and standard deviation) as measured by the LexTale as well as the number of participants in each CPF level of proficiency following their LexTale score, self-reported English proficiency (1 = not good at all, 10 = native-like) mean (and standard deviations), alongside participants' first language and country of origin, and length of residency in the United Kingdom.

	Native speaker conditon	Non-native speaker condition
LexTale score	Mean score = 74 (13.6) C1 - C2: 19 B2: 19 B1: 10	Mean score = 76.3 (11.7) C1 - C2: 23 B2: 21 B1: 4
Self-reported English proficiency in speaking	7.84 (1.44)	8.10 (1.51)
Self-reported English proficiency in reading	8.67 (1.15)	8.73 (1.23)
Self-reported English proficiency in listening	8.44 (1.29)	8.62 (1.23)
Self-reported English proficiency in writing	7.99 (1.47)	8 (1.52)
Length of stay in the UK	Range = 1 month - 19 years Mode = 2 months	Range = 1 month - 9 years Mode = 5 months
First language	Chinese (21), Polish (5), Spanish (3), Cantonese (2), Dutch (2), Greek (2), Mandarin (2), Czech (1), Danish (1), Farci (1), German (1), Latvian (1), Norwegian (1), Punjabi (1), Romanian (1), Slovak (1), Thai (1), Turkish (1)	Chinese (11), Dutch (3), French (3), Polish (3), Cantonese (2), Finnish (2), Hungarian (2), Greek (2), Malay (2), Mandarin (2), Slovene (2) Spanish (2), Arabic (1), Bulgarian (1), Croatian (1), Czech (1), Danish (1), Estonian (1), Indonesian (1), Italian (1), Lithuanian (1), Portuguese (1), Shona (1), Turkish (1).

Materials

Experiment 2 employed the same materials reported in Experiment 1, with the addition of post-experimental measures to control for participants' linguistic backgrounds.

The present study utilized a version of the English LexTale (Lemhöfer & Broersma, 2012) implemented using OpenSesame (Mathôt et al., 2012) to measure participants' English proficiency. In order to assess participants' eligibility, we employed a pen-and-paper adapted version of the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian et al., 2007). Participants were asked to list all the languages they spoke in the order of acquisition and rate their proficiency in speaking, writing, reading, and listening on a 10-point scale. For each language, participants reported the age of acquisition, age of fluency, age when the language was first used for communicative purposes (i.e. outside of a classroom setting), length of time spent learning the language, and mode of acquisition (i.e. classroom, interaction with other people, media, or mixed). Participants were also asked to rate their current exposure to each language on a 10-point scale in various contexts (e.g. interactions with relatives, classmates, or media) and indicate if they switched between languages. We included additional questions on language use (de Bruin, 2019) and was similar to those employed in prior research on second language (L2) speakers (Foucart et al., 2014; Ito et al., 2018): Participants reported the percentage of exposure to each language during childhood, adolescence, and currently. Finally, we collected demographic information such as country of origin, length of residency in the United Kingdom and any other English-speaking country. Completing these tasks took no longer than twenty minutes.

Procedure

The procedure was identical to that reported in Experiment 1 with the following exceptions. After the experiment, participants performed a version of the English LexTale (Lemhöfer & Broersma, 2012), implemented using OpenSesame (Mathôt et al., 2012). Afterwards, besides filling in the same post-experimental questionnaire as participants in

Experiment 1, participants filled in the adapted version of the LEAP-Q (Marian et al., 2007).

4.3.2 Results

As in Experiment 1, trials where no click was recorded or participants clicked on the distractor were discarded from analysis (1% of trials). Additionally, 9.38% of trials were not included in the analysis due to an error in presentation, as the experimental pairs overlapped phonologically.

Answers to questionnaire

Participants rated the naturalness of the audio an average of 8.04 (1.29) for the native speaker and 6.57 (1.62) for the non-native speaker ($t(94) = 5.24$, $p < .001$). They rated the non-native speaker's fluency as 5.84 (1.77) and the audio's accentedness as 7.04 (1.59). Non-native participants reported being exposed to non-native accented English at an average of 7.11 (1.74) daily and interacting daily with non-native speakers at 7.54 (1.73) on a 9-point scale. Additionally, participants reported interacting with British English speakers an average of 6.45 (2.14). These measures did not impact participants' behaviours and are not discussed any further in the results.

Reaction time

On average, participants were highly accurate in selecting the right target (99.32% for the native speaker, 98.66% for the non-native speaker). Table 4.7 depicts participants' average reaction times by manner of delivery, speaker's linguistic background and target frequency. Prior to analysis, we removed observations where reaction time exceeded 2000 ms from target offset (0.94 %) and reaction times above or below 2 standard deviations from the participant's mean (4.91 % data lost).

Table 4.7

Mean reaction times (ms) of participants' click on the referent for both the native and non-native speaker condition by manner of delivery and target's frequency (standard deviation in brackets).

	Native speaker RT (ms)	Non-native speaker RT (ms)
Fluent		
High-frequency	921 (214)	946 (231)
Low-frequency	1026 (241)	1028 (243)
Disfluent		
High-frequency	916 (197)	939 (225)
Low-frequency	1008 (236)	992 (230)

We modelled participants' raw reaction times via a linear mixed model with fixed effects for manner of delivery (fluent coded as -0.5, disfluent as +0.5), speaker's linguistic background (native coded as -0.5, non-native as +0.5), target frequency (high-frequency coded as -0.5, low-frequency as +0.5), and their interactions. The maximal model with by-participant and by-item random intercepts, and random slopes for fluency, frequency and their interaction by-participant, and random slopes for fluency, speaker's linguistic background and their interaction by-target failed to converge. The model was thus simplified by dropping hierarchically the parameters that explained the least variance. The final model's random effect structure included random slopes for frequency and fluency by-participant, and random slopes for speaker's linguistic background by-target.

The model showed a main effect of frequency, whereby low-frequency words lead to slower reaction times ($\beta = 90.94$, SE = 18.76, t = 4.85). We failed to find any effect for manner of delivery ($\beta = 13.64$, SE = 7.66, t = 1.78), speaker's linguistic background ($\beta = 9.52$, SE = 30.84, t = 0.31), or the interactions between these factors. Due to the variability in our participants' length of stay in the United Kingdom (see Table 4.6), we did not conduct any further analyses exploring the effects of proficiency.

Eye movements: Prediction Time window

Eye movements prior to target onset (i.e., a time window reflecting prediction) were analysed via two linear mixed models. As in Experiment 1, we conducted two analysis for fluent and disfluent utterances, covering the time window from 'the' onset to word onset. Following our discussion in Section 4.2.2, we analysed eye movements via empirical logits to measure the preference to fixate on the low-frequency item over the high-frequency one (i.e., the low-frequency advantage): Positive values reflect a preference to fixate on the low-frequency item, while negative values reflect a preference to fixate on the high-frequency item. Each model included speaker's linguistic background (native coded as 0, non-native as 1), two time components (a linear and a quadratic component), and the interaction of nativeness with each time component as fixed effects. For ease of convergence, we divided the linear and quadratic terms by 200. We included random slopes by-participant, by-item and by-sentence template. Values were considered significant at $|t| > 2$ (Baayen, 2008).

Figures 4.8 and 4.9 depict the low-frequency advantage for fluent and disfluent utterances respectively (participants' raw probabilities of fixations are depicted in figures A.2 and A.3 in Appendix A, section A.2.2). Figure 4.8 suggests that participants did not show any preference towards the low-frequency item (red line) over the high-frequency item (orange line), for both speaker conditions. In disfluent utterances (Fig. 4.9), the pattern of fixations towards either item did not differ over the time window where the phonological information about the disfluency was available. Figure 4.9 suggests that our non-native listeners' fixations were not influenced by the presence of a disfluency.

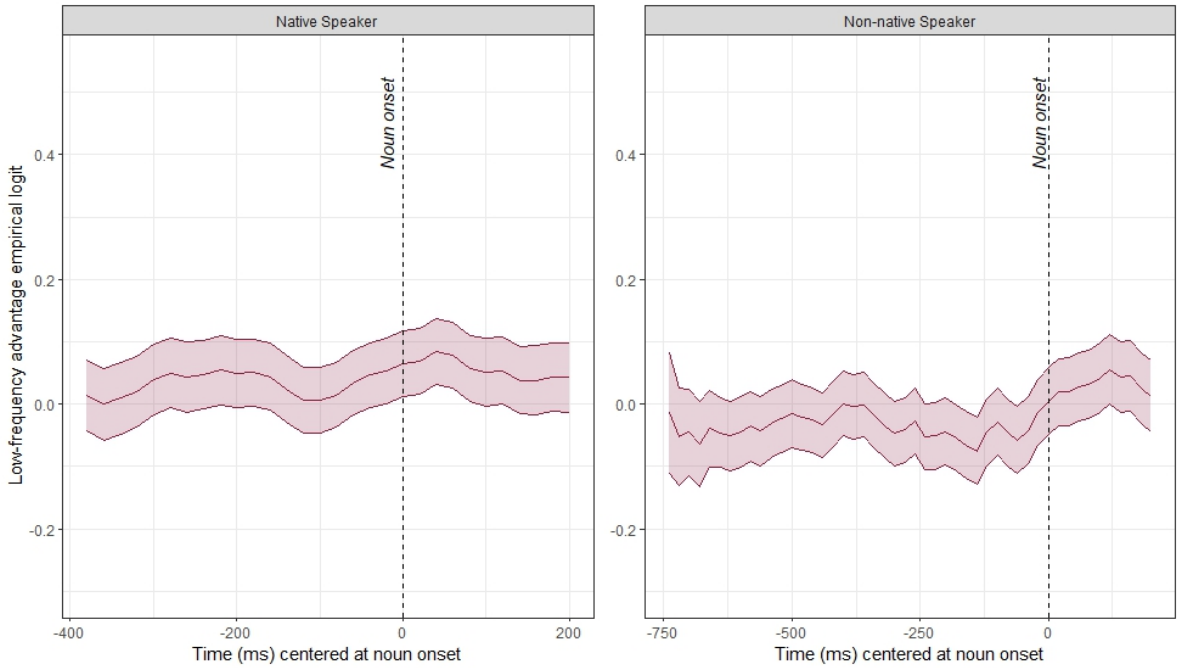

Figure 4.8. Mean low-frequency advantage in empirical logits in fluent utterances by speaker's linguistic background from audio onset to 200 ms post-target onset. Positive values index a preference to fixate on the low-frequency object, and negative values index a preference to fixate on the high-frequency object. Shaded areas represent ± 1 standard error of the mean.

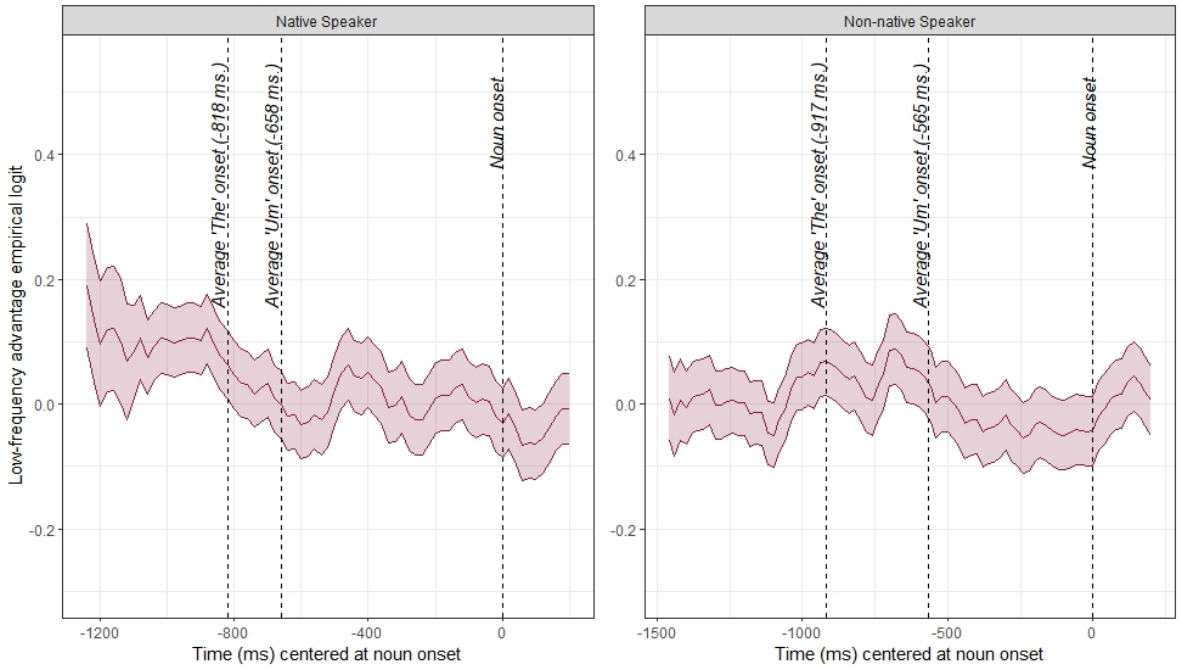

Figure 4.9. Mean low-frequency advantage in empirical logits in disfluent utterances by speaker's linguistic background from audio onset to 200 ms post-target onset. Positive values index a preference to fixate on the low-frequency object, and negative values index a preference to fixate on the high-frequency object. Shaded areas represent ± 1 standard error of the mean.

Table 4.8

Estimated parameters of two linear mixed effects regression models (one per manner of delivery) on the preference to fixate on the low-frequency item via empirical logits, with fixed effects of speaker's linguistic background (native coded as 0, non-native as 1), a linear and a quadratic time components, and their interactions. Time window of analysis spanned from 'the' onset until 200 post-target onset.

	Estimate	Std. Error	t-value
Fluent utterances			
Intercept	0.04	0.11	0.55
Linear Time	-0.006	0.09	-0.07
Quadratic Time	0.000008	0.0002	0.03
Non-native Speaker	-0.14	0.14	-1
Linear Time * Non-native Speaker	0.24	0.13	1.85
Quadratic Time * Non-native Speaker	-0.0006	0.0004	-1.50
Disfluent utterances			
Intercept	1.92	0.07	0.30
Linear Time	0.006	0.02	0.35
Quadratic Time	-0.00002	0.00002	-1.08
Non-native Speaker	0.07	0.08	0.82
Linear Time * Non-native Speaker	-0.08	0.03	-3.06
Quadratic Time * Non-native Speaker	0.00007	0.00002	3.15

Table 4.8 depicts the results of both models. Fluent utterances by the native speaker showed no preference towards either item over time (β = -0.006, SE = 0.09, t = -0.07; β = 0.000008, SE = 0.0002, t = 0.03). This lack of preference did not differ for fluent utterances produced by the non-native speaker. In line with Figure A.3, disfluent utterances produced by the native speaker did not lead to a preference towards either item (β = 0.006, SE = 0.02, t = 0.35; β = -0.00002, SE = 0.00002, t = -1.08). In contrast, disfluent utterances by the non-native speaker were characterised by a linear decrease in preference to look towards the low-frequency item (β = -0.08, SE = 0.03, t = -3.06), followed by a weak quadratic increase (β = 0.00007, SE = 0.00002, t = 3.15). This time

course contrasts with that of our participants in Experiment 1, in that filled pauses led to a low-frequency disadvantage, to what we will return in the discussion.

Eye movements: Word recognition

Following Experiment 1, we analysed whether manner of delivery affected the recognition of the target, especially given previous research suggesting that prediction in second-language comprehension may be delayed.

Following the approach taken in Experiment 1, we modelled the preference to fixate on the target (as opposed to the low-frequency item), because phonological information was available in this time window. We analysed in a linear mixed model the preference to fixate on the target via empirical logits (Barr, 2008) over a time window from target onset to 760 ms post-target onset, where positive values mean fixations on the target, and negative values reflect fixations on the distractor. The model included nativeness (native coded as 0, non-native as 1), frequency (high-frequency coded as 0, low-frequency coded as 1), manner of delivery (fluent coded as 0, disfluent coded as 1), their interactions, a linear and a quadratic time component, as well as the interactions between our variables with the time components. For ease of convergence, we divided the linear and quadratic terms by 200. We included random slopes by-participant, by-item and by-sentence template.

Table 4.9 reports the model's results, which are depicted in Figure 4.11 (but also see A.4 for the target advantage, and Appendix A, section A.2.3 for a visualization of the raw fixation probabilities). The discussion of this analysis is reported by referring to both the model's estimates and their visualisation for ease of reading.

Table 4.9

Estimated parameters of a mixed effects model for target preference measured as empirical logits in a time window from target onset to 760 ms post-target onset, with speaker's linguistic background (native coded as 0, non-native as 1), target frequency (high-frequency coded as 0, low-frequency as 1), manner of delivery (fluent coded as 0, disfluent as 1), linear and quadratic time components, and their interaction as fixed effects.

	Estimate	Std. Error	t value
A. Intercept	-0.22	0.07	-3.08
B. Linear Time	0.11	0.04	2.45
C. Quadratic Time	0.001	0.00001	10.67
D. Low-frequency	0.34	0.07	4.65
E. Low-frequency * Linear Time	-0.41	0.06	-6.78
F. Low-frequency * Quadratic Time	0.0003	0.0001	3.98
G. Disfluent	0.39	0.07	5.31
H. Disfluent * Linear Time	-0.31	0.06	-5.11
I. Disfluent * Quadratic Time	0.0003	0.0001	3.61
J. Disfluent * Low-frequency	-0.33	0.07	-4.66
K. Disfluent * Low-frequency * Linear Time	0.42	0.09	4.90
L. Disfluent * Low-frequency * Quadratic Time	-0.0005	0.0001	-4.11
M. Non-native Speaker	0.27	0.08	3.23
N. Non-native Speaker * Linear Time	-0.36	0.06	-5.85
O. Non-native Speaker * Quadratic Time	0.0003	0.0001	3.67
P. Non-native Speaker * Low-frequency	-0.34	0.07	-4.82
Q. Non-native Speaker * Low-frequency * Linear Time	0.43	0.09	4.94
R. Non-native Speaker * Low-frequency * Quadratic Time	-0.0004	0.0001	-3.50
S. Non-native Speaker * Disfluent	-0.46	0.10	-4.47
T. Non-native Speaker * Disfluent * Linear Time	0.41	0.09	4.70
U. Non-native Speaker * Disfluent * Quadratic Time	-0.0003	0.0001	-3.10
V. Non-native Speaker * Disfluent * Low-frequency	0.30	0.10	2.95
W. Non-native Speaker * Disfluent * Low-frequency * Linear Time	-0.33	0.12	-2.67
Z. Non-native Speaker * Disfluent * Low-frequency * Quadratic Time	0.0003	0.0002	2.23

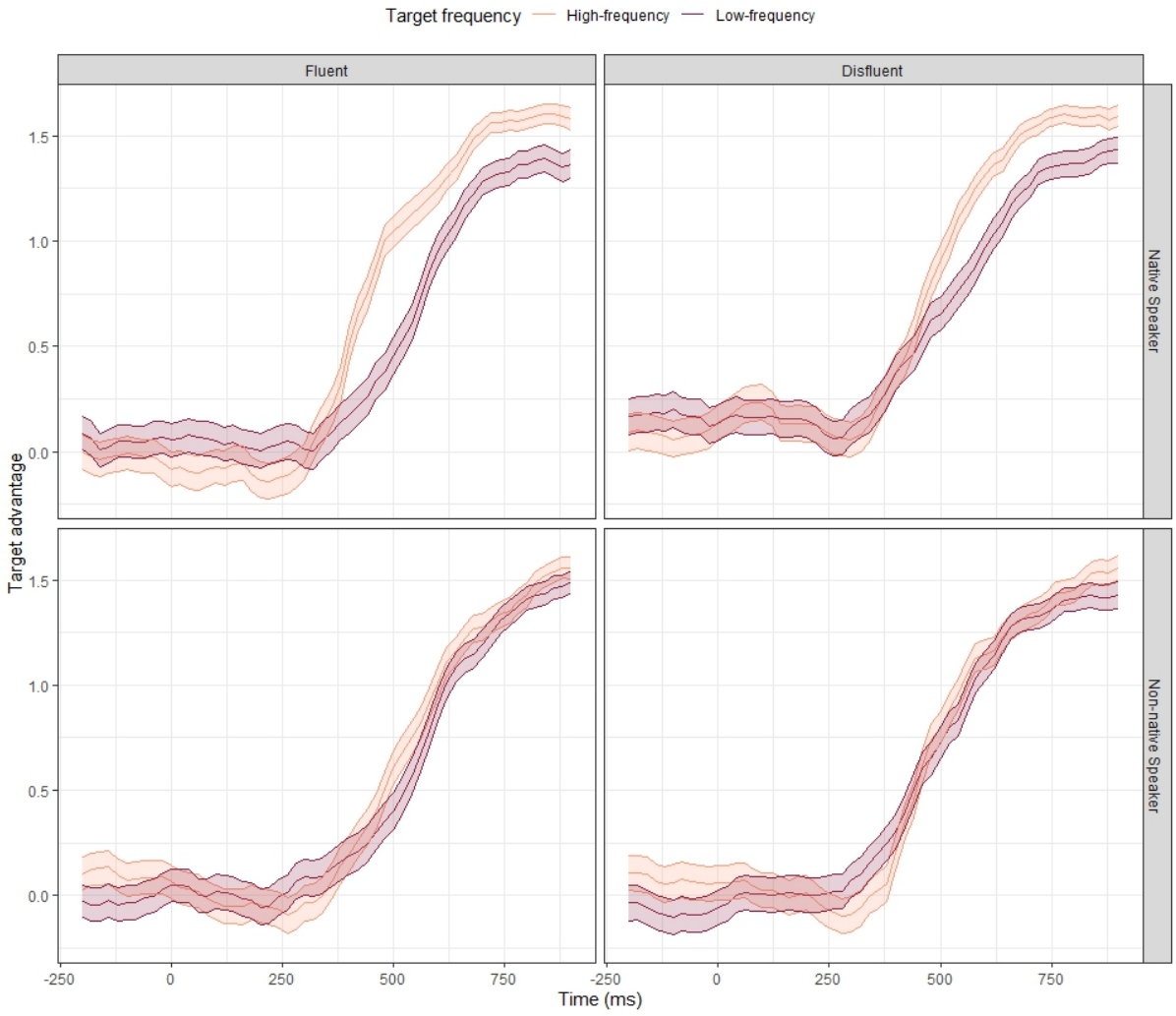

Figure 4.10. Mean target advantage by target frequency (red: low-frequency, orange: high-frequency) by manner of delivery and and speaker's linguistic background (native/non-native). Target advantage was calculated via empirical logits, where positive values indicate a preference to fixate on the target, and negative values a preference to fixate on the distractor. Time window of analysis spanned from target onset to 760 ms post-target onset. Shaded areas represented ± 1 standard error of the mean.

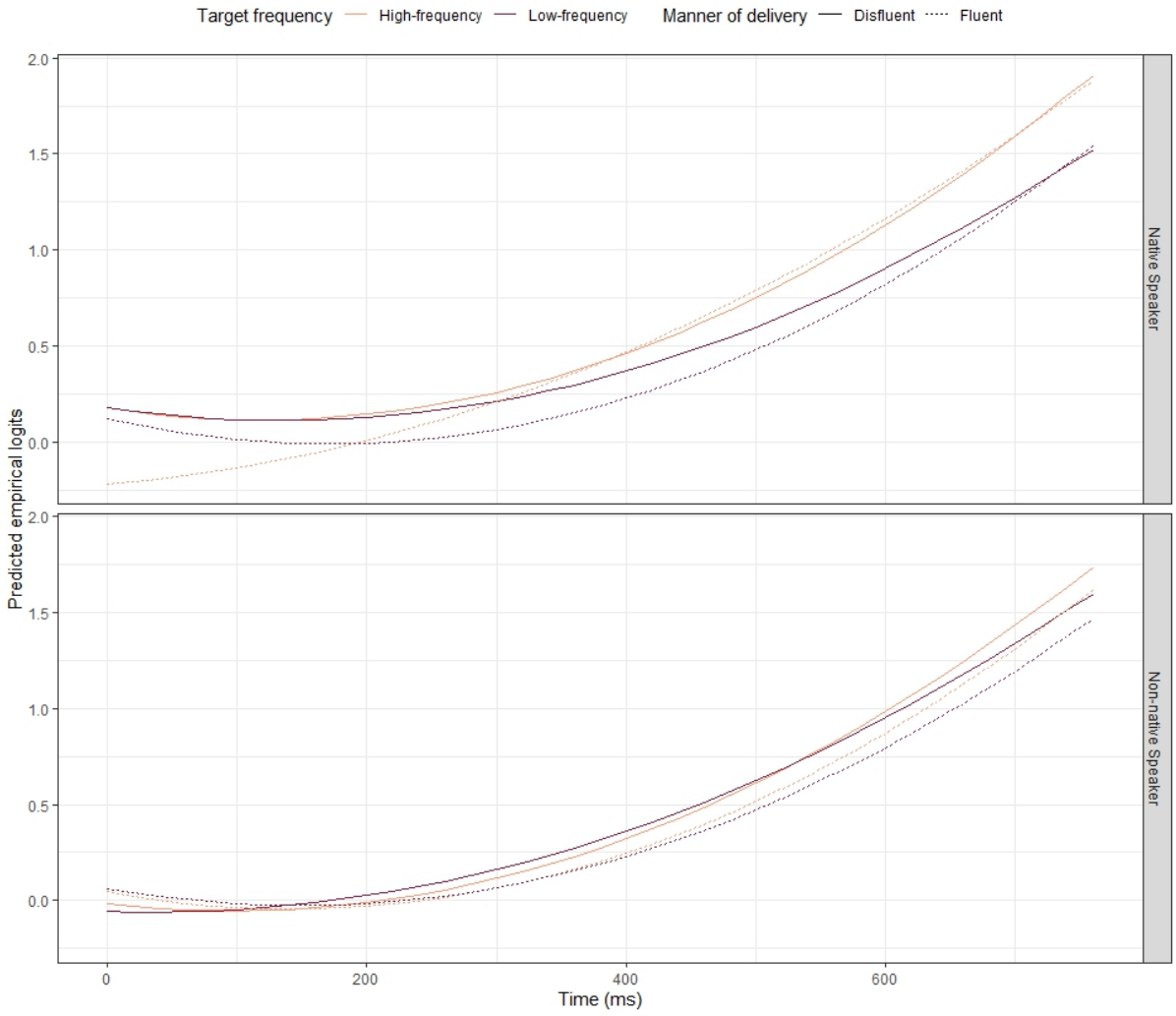

Figure 4.11. Growth Curve Analysis model fit of empirical logits to the low-frequency (red) and high-frequency (orange) fixations as a function of manner of delivery (dashed lines: fluent, solid lines: disfluent) and speaker's linguistic background (native/non-native).

The top panel of Figure 4.11 depicts participants' preference to fixate on high- (orange) and low-frequency (red) targets in the native speaker condition. Parameters A to C refer to the dashed, orange line: The target advantage when it was a high-frequency target produced fluently. The model showed that in this condition, the target advantage increased linearly and quadratically. A filled pause increased the preference to fixate on these targets at the offset (parameter G), although the target advantage decreased linearly in comparison to its fluent counterpart (parameter H). Visual inspection of Figure

4.11 suggests that this is due to the differences at the offset: From 400 ms, both the solid (disfluent) and dashed (fluent) lines appear to follow a similar trajectory. Moving to participants' preference to fixate on low-frequency targets when produced fluently, we find that the increase of the target advantage was less linear and more quadratic than their high-frequency counterparts (parameters E and F). Indeed, Figure 4.11 suggests that the preference to fixate on the target when it was low-frequency, regardless of manner of delivery, increased quadratically, possibly due to an increased word-frequency effect in second-language comprehension (Diependaele et al., 2013). When low-frequency targets were produced disfluently, the target advantage increased more linearly (parameter K). In fact, the comparison of the solid and dashed red lines in Figure 4.11 shows a larger increase in the preference to fixate on low-frequency targets when they were preceded by a disfluency (solid line), which disappears towards the end of the time window. This suggests that the presence of a filled pause aided the recognition of low-frequency targets.

The bottom panel of Figure 4.11 depicts the pattern of fixations in the non-native speaker condition. In this case, we find that the preference to fixate on high-frequency targets produced fluently (parameters M to O, solid orange line) increased less linearly and more quadratically than their native counterparts. Visual inspection of Figure 4.11 suggests that, overall, the recognition of targets produced by the non-native speaker was characterised by a more quadratic, rather than linear, increase - possibly as a consequence of attending to non-native-accented speech and subsequent reduced intelligibility. The preference to fixate on the high-frequency item following a disfluency increased more linearly than its native counterpart (parameters S to U). This is illustrated by comparing the differences between the solid (disfluent) and dashed (fluent) orange lines in the right and left panels of Figure 4.11. The target advantage for low-frequency targets produced fluently was characterised by a decreased preference at the offset (parameter P), followed by a linear increase (parameter Q). The target advantage for low-frequency targets produced disfluency was characterised by an increase at the offset (parameter V), followed by a linear decrease in the target advantage in contrast to the native counterpart (parameter W). Visual inspection of Figure 4.11 suggests that the

presence of a disfluency still increased the preference to fixate on low-frequency targets, i.e., the difference between the solid and dashed red lines, depicting the preference to fixate on low-frequency targets in the disfluent and fluent conditions respectively. In sum, this pattern suggests that disfluencies produced by a non-native speaker led to an increase in the preference to fixate on the target, regardless of its frequency.

4.3.3 Discussion

Experiment 2 extended Experiment 1 to a sample of non-native listeners who followed instructions provided by either a native or a non-native speaker who referred fluently or disfluently to one item out of low- and high-frequency pairs. We failed to find any anticipatory eye-movements towards low-frequency items upon encountering a native disfluency, while a non-native disfluency led to fewer fixations on the low-frequency target. Analysis of fixations post-target onset, which we take as reflecting word recognition, suggests that native disfluencies facilitated the recognition of low-frequency targets, while non-native disfluencies facilitated all targets, regardless of their frequency. However, this facilitation did not translate into participants' reaction times, where we only found an effect of target frequency. These patterns are at odds with those of Watanabe et al. (2008), who reported an effect of manner of delivery in the identification of harder-to-name objects in non-native listeners. We divide the discussion of results into those prior to target onset, reflecting predictive processes, and those after listeners had encountered phonological information about the target, reflecting word recognition.

In Experiment 2, we failed to find any anticipatory eye-movements towards low-frequency items in either speaker condition. Following the results of Experiment 1, it is possible that only non-native disfluencies could yield anticipatory eye-movements towards objects whose labels are low-frequency: Non-native listeners, just like our native participants, did not have enough evidence about the upcoming filled pause in the native speaker's instructions. The lack of expectations towards low-frequency items following a non-native disfluency (and, in fact, a slight dispreference to fixate on these items) could be

attributed to a general delay in predictive processing in second language comprehension (Chun & Kaan, 2019; Dijkgraaf et al., 2019).

Kaan (2014) argued that previously reported differences in prediction between first- and second-language comprehension can be attributed to disparities between native and non-native comprehenders with regard to the target language, such as the exposure to it. In the present study, the property of the targets that was associated with a filled pause (and that the presence of the latter leads to an anticipation of the former) was word frequency. It has been shown that the word-frequency effect is larger in non-native comprehenders due to their reduced exposure to the language (Gollan et al., 2008; see also Diependaele et al., 2013; Whitford and Titone, 2012). Consequently, the pre-activation of these words may require more evidence and will be slower, explaining the lack of anticipatory eye-movements in Experiment 2.

To the best of our knowledge, this is the first experiment directly exploring the effects of filled pauses on predictive processing in second language comprehension. While Watanabe et al. (2008) framed their results as reflecting listeners' anticipations following how the speaker delivered the instructions (i.e., fluently or disfluently), we refrain from taking their results as such due to the measurement employed (see Section 3.2). However, we do take their results as partially demonstrating that non-native listeners' comprehension can be guided by manner of delivery: Specifically, comprehension of one's second language can be benefited by cues that have been mastered (Clahsen & Felser, 2006; Futrell & Gibson, 2017). If the disfluency bias is a learnt inference between disfluencies and trouble in speech production, then it is a transferable cue between languages. In the case of Watanabe et al. (2008), listeners are aware of the correlations between disfluency and referring to objects that might lead to filled pauses due to problems at the level of conceptualising the message (i.e., due to language-extrinsic problems). For example, a filled pause is likely to precede a harder-to-name object. In contrast, the present study manipulated a language-intrinsic property: The accessibility of the label to refer to the

object, as filled pauses are likely to precede low-frequency words (see Corley & Hartsuiker, 2003, for a similar discussion). Our results suggest that anticipations informed by (dis)fluency in second-language comprehension may not be as specific as previously reported for native listeners (Bosker et al., 2014), or may be delayed, potentially suggesting that these latter types of predictions are more heavily tied to listeners' linguistic knowledge.

The pattern of fixations on the target once listeners encountered phonological information shows differences depending on the speaker's linguistic background. For the native speaker, we found that disfluencies aided in the recognition of low-frequency words. In contrast to Experiment 1, the presence of a filled pause led to a preference to fixate on high-frequency words at the beginning, which then decreased. This pattern aligns with both the fillers-as-time and the fillers-as-tokens accounts. Firstly, the initial increase in fixations on high-frequency targets preceded by a filled pause can be taken as evidence of disfluencies easing word recognition (Corley & Hartsuiker, 2011) and possibly re-orienting listeners' attention to the speech signal (Fox Tree, 2002). The fact that the benefit of disfluency was larger for low-frequency targets can be explained by listeners' awareness of the correlation between disfluencies and low-frequency words, facilitating their lexical access. This thus suggests that the disfluency bias towards low-frequency words in second-language comprehension may arise, but is likely delayed due to non-natives' lexical access disadvantage. However, whether this awareness is due to listeners' learning of the distributional properties of the linguistic input (i.e., following the tokens-associative account) or their tuning to the speaker's trouble in production (i.e., following the tokens-inference account) is unclear.

The effects of non-native disfluencies on word recognition fully align with the fillers-as-time account. In this case, disfluencies eased the recognition of both low- and high-frequency targets, suggesting that listeners did not have any expectations about any specific linguistic unit upon encountering a disfluency (or that there was no benefit of

encountering an element that is associated with disfluency). Filled pauses may aid second-language comprehension by virtue of delimiting word boundaries. These properties may be beneficial when listeners are attending to speech produced in a non-native accent, which may have reduced intelligibility. For example, non-native listeners are better at recognising previously heard words when they are produced with a native versus a non-native accent, although accuracy improves as non-native listeners' proficiency in the target language does (Lev-Ari et al., 2017; Weber et al., 2014).

An interesting possibility is that the perception that a non-native speaker struggles with language production may have re-oriented non-native listeners' attention. Fox Tree (2002) argues that listeners might attribute a filled pause to problems in speech production, and subsequently, listeners re-orient their attention to the speech signal. We previously argued that listeners' awareness of what may pose trouble in production leads them to anticipate certain elements over others upon encountering a filled pause (e.g., a low-frequency word over a high-frequency word), and that when this expectation is not met (i.e., a filled pause precedes a high-frequency word), there is a prediction cost. Speculatively, non-native comprehenders' increased awareness of their interlocutor's linguistic background (Rubio-Fernández & Glucksberg, 2012) might have halted all predictive processes, or at least those triggered by perceptions of difficulty, because non-native listeners are more aware of the general difficulties of word-finding in second-language comprehension (cf. Lev-Ari, 2015, for similar arguments). However, the present study does not allow us to answer this question and thus leaves the door open to whether the comprehension of one's second language produced with disfluencies by a non-native speaker relies on an inference mechanism.

In fact, previous research has suggested that non-native listeners' disfluencies can orient listeners' attention to the speech signal. Bosker et al. (2015) reported that native listeners were more accurate at detecting changes in the transcription of a previously listened recording if the words changed had been accompanied by a disfluency, regardless of whether it was produced by a native or a non-native speaker. These findings align with

the idea that disfluencies re-orient listeners' attention to the speech signal. It is possible that for disfluencies to benefit the recognition of low-frequency words, listeners need to have some form of expectation (as in Experiment 1 and the native speaker condition of Experiment 2). Given that processes underlying second-language comprehension appear to be delayed when speech is produced by a non-native speaker (e.g., Grey et al., 2019), the lack of interaction between manner of delivery and target frequency in word recognition can be putatively put down to a lack of resources to engage in predictive processing (even in the forms of associations), and non-native listeners' increased reliance on the bottom-up signal.

To sum up, the results of Experiment 2 speak to both the fillers-as-time and the fillers-as-tokens account. We interpret participants' word recognition as supporting the idea that the fillers-as-tokens account may scaffold on listeners' re-oriented attention to the speech signal. In scenarios where listeners are not in a cognitively demanding situation (i.e., attending to native-accented speech as opposed to non-native-accented speech), there can be further effects of disfluency following the fillers-as-tokens view: Listeners' word recognition is eased when these words are congruent with manner of delivery.

Whether this benefit can be accounted for by a tokens-associative or a tokens-inference account cannot be disentangled by Experiment 2, due to the lack of interaction between manner of delivery and word frequency in the non-native speaker condition. However, we proposed that investigating the differences between native and non-native listeners can shed light on the mechanisms that underlie the disfluency bias. Specifically, exploring whether participants' pattern of fixations differs depending on the speaker's linguistic background is a first step towards disentangling the tokens-associative and the tokens-inference account. While both native and non-native listeners might have experienced a benefit from a disfluency to recognise low-frequency words, this benefit might have taken different forms between listeners.

4.4 Comparison across populations

The findings from the identification task suggest that native and non-native listeners may have perceived fluency differently. To statistically assess whether the findings in Experiment 1 and Experiment 2 differed, we combined data from Experiment 1 and 2 to investigate the patterns of results in the prediction and integration time windows.

4.4.1 Eye movements: Prediction Time window

We conducted two linear mixed models for the preference to fixate on the low-frequency target via empirical logits (Barr, 2008) for fluent and disfluent utterances. As in Experiments 1 and 2, the time window of analysis covered from *'the'* onset to 200 ms post-target onset. Both models included a speaker's linguistic background (native coded as 0, non-native as 1) and experiment (Experiment 1 coded, i.e., native listeners, coded as 0, Experiment 2 coded as 1), their interaction, a linear and a quadratic time component, and the interactions with the factors of interest. For ease of convergence, we divided the linear and quadratic terms by 200. We included random intercepts by-participant, by-item and by-sentence template. We considered results significant at $|t| > 2$ (Baayen, 2008). The full depiction of the results of these models can be found in Appendix A, Section A.3.1.

Analysis of fluent utterances showed that native participants' fixations showed no preference towards either item, regardless of the speaker's background. This pattern did not differ for participants in Experiment 2. The low-frequency advantage in disfluent utterances increased linearly for non-native disfluencies in Experiment 1 (β = 0.11, SE = 0.03, t = 4.22), while in Experiment 2 it decreased (β = -0.18, SE = 0.04, t = -5.12)

4.4.2 Eye movements: Word recognition

To explore whether participants in Experiment 1 and 2 differed in the ease with which they recognised upcoming elements in speech following a disfluency, we conducted a linear mixed model for the preference to fixate on the target via empirical logits (Barr,

2008), where positive values index a preference to fixate on the target, and negative values reflect a preference to fixate on the distractor. The time window of analysis spanned from target onset to 760 ms post-target onset, and included fixed effects of speaker's linguistic background (native coded as 0, non-native as 1), manner of delivery (fluent coded as 0, disfluent as 1), frequency (high-frequency coded as 0, low-frequency as 1), experiment (Experiment 1 coded as 0, Experiment 2 as 1), a linear and a quadratic time component, the interactions between the variables of interest, and their interactions with the time components. For ease of convergence, we divided the linear and quadratic terms by 200. The model included random intercepts by-participant, by-item and by-sentence template. Due to the model's complexity, we will focus on the differences between Experiment 1 and 2, but the model's results are reported in Appendix A, Section A.3.2. We considered results significant at $|t| > 2$ (Baayen, 2008).

Starting with those who listened to the native speaker, the model showed that the growth of the target advantage increased similarly between participants in Experiment 1 and 2 when the target was a high-frequency word produced fluently. The target advantage for these words when they were produced disfluently showed that while in Experiment 1 the pattern did not differ from the fluent counterpart, the trajectory in Experiment 2 was characterised by a preference to fixate on the target at the offset and a decrease in the target advantage over time. This aligns with the idea that initially, disfluencies attracted non-native listeners' attention, but in the case of disfluencies produced by a native speaker, the presence of a filled pause impaired the recognition of high-frequency words. In contrast, the target advantage for low-frequency words was characterised by a reduced linear increase in Experiment 2, likely reflecting the exacerbated word-frequency effect in one's second language (Gollan et al., 2008). However, the linear and quadratic growth of the target advantage for low-frequency targets produced disfluently followed a similar pattern in Experiments 1 and 2.

Moving to the target advantage for those listening to the non-native speaker, the results showed that disfluencies by the non-native speaker led to different trajectories

in Experiments 1 and 2. In Experiment 1, we find that disfluencies preceding a high-frequency target led to a linear decrease and a quadratic increase of the target advantage, while in Experiment 2 the presence of a disfluency led to a linear increase of the target advantage. For low-frequency targets, the model showed that fluent utterances led to a linear increase in the target advantage, which differs from the results of Experiment 1. There was a target advantage at the offset in Experiment 1: for disfluencies preceding a low-frequency target which did not differ in Experiment 2. The linear increase of the target advantage did not differ between experiments. Overall, this corroborates the idea that while disfluencies produced by a native speaker benefited word recognition when filled pauses were a congruent cue (i.e., they preceded a low-frequency word), disfluencies produced by a non-native speaker led to different benefits depending on the listeners' linguistic background: For native listeners, they benefited when they were a congruent cue, while for non-native listeners, they increased the recognition of words regardless of their frequency.

4.5 Identification task

One possibility for the results reported here has to do with the auditory stimuli. In studies that have reported a disfluency bias, the differences between fluent and disfluent auditory materials commonly go beyond the presence of a filled pause itself: For example, the prosody of the segment prior to the disfluency differs from its fluent counterpart (Arnold et al., 2003). In fact, Arnold et al. (2003) posited that the disfluency bias might not be solely due to the presence (or absence) of a filled pause, but the combination of a series of factors that make up the perception of disfluency. In a post-experimental identification task, Bosker et al. (2014) showed a new set of participants the auditory stimuli of the experiment minus the disfluency (i.e., 'Click on') and asked them to categorize segments to belonging either to a fluent or a disfluent utterance. The authors found that participants were 50% and 51% (for the native and non-native speaker, respectively) accurate in identifying disfluent segments as such, suggesting that participants' ability to detect

disfluency in their materials was not above chance. Participants' ability to detect a disfluent segment, even when they have been explained the manipulation, contrasts with Bosker et al. (2014) reported pattern of eye movements, where participants were already looking at the low-frequency item before it was uttered (i.e., while listening to 'Click on'). We previously argued that a potential explanation for the lack of anticipatory eye movements could be put down to the qualities of our auditory stimuli. It could be possible that a filled pause is the most overt mark of a speaker's hesitation, but for the disfluency bias to emerge (at least, prior to encountering the predicted linguistic item), the system requires more cues associated with disfluency.

In order to explore this possibility, we conducted an online two-choice forced identification task similar to that of Bosker et al. (2014). In this task, participants were presented excerpts of our auditory stimuli prior to encountering a filled pause (i.e., 'Click on') and were asked to discriminate each as belonging to a fluent or disfluent utterance.

We tested 100 self-reported native English listeners (50 by speaker condition) and 100 self-reported non-native listeners (50 by speaker condition) from both the University of Edinburgh student pool and the portal Prolific. All participants gave their informed consent, as approved by the University of Edinburgh PPLS Ethics Committee (reference number: 314-2122/2). To ensure that participants were similar to those in Experiments 1 and 2, native English listeners had to be born and raised in the United Kingdom. Non-native listeners had to report whether English was their first language, and their country of origin. All participants had to be currently residing in the United Kingdom to take part in the study. Participants were reimbursed, either with an economic reimbursement (£1) or with university credit.

We employed the 12 sentence templates of Experiments 1 and 2 (six per manner of delivery, six per speaker's linguistic background) as auditory stimuli. We cut the recordings for both fluent and disfluent utterances to only include "Click on". Following Bosker et al.'s (2014) design, we presented each version of the stimuli produced by one

speaker 5 times. Participants only heard the native or the non-native speaker, thus each session only included 30 items, presented in a random order.

At the beginning of the session, participants were informed about the manipulation of the audio. They were told they would listen to a previously recorded native or non-native speaker who produced fluent and disfluent utterances. Participants were explained that the audio they would listen to was edited, in that we had removed the filled pause, and that their task was to identify whether the audio belonged to a fluent or a disfluent utterance. Prior to the task, participants listened to a fluent and a disfluent utterance produced by the speaker, to ensure that participants knew how (dis)fluency sounded like in that particular speaker. After the task, participants reported demographic information, including their country of origin, first language, and self-reported proficiencies in English, to ensure that they met the participation criteria.

Table 4.10 depicts participants' accuracy as a function of their linguistic background, manner of delivery, and speaker condition. The data showed that, in general, native listeners performed at chance, although they were slightly better at recognising fluent excerpts as such when they belonged to a native speaker. In contrast, non-native listeners seemed to have a fluency bias, whereby disfluent excerpts were likely to be categorised as fluent.

Table 4.10

Mean accuracy of participants' recognition task by manner of delivery and speaker condition.

	Native speaker	Non-native speaker
Native participants		
Fluent	65%	54%
Disfluent	52%	41%
Non-native participants		
Fluent	68%	66%
Disfluent	42%	34%

We investigated whether participants had correctly identified whether the stimuli belonged to a fluent or disfluent utterance in a logistic mixed effects model. We included speaker's linguistic background (native coded as 0, non-native as 1), manner of delivery (fluent coded as 0, disfluent coded as 1), participant's linguistic background (native coded as 0, non-native as 1) and their interactions as fixed effects. We included random intercepts by-participant and by-audio, and a random slope for manner of delivery by-participant.

The model showed that native participants were more likely to miscategorise a disfluent utterance (β = -0.72, SE = 0.20, z = -3.59, p <.001), regardless of the speaker's linguistic background (β = 0.28, SE = 0.26, z = 1.06, p = .29). This suggests that disfluent utterances produced by the native and the non-native speaker were more likely to be misidentified as fluent. Additionally, the model showed an interaction between manner of delivery, speaker identity, and listeners' linguistic background. Specifically, native participants were more likely to miscategorise fluent utterances produced by the non-native speaker (β = -0.47, SE = 0.18, z = -2.55, p = .01), suggesting a tendency to perceive this speaker as disfluent. In contrast, non-native participants were slightly more likely to miscategorise disfluent recordings of the native-speaker (β = -0.59, SE = 0.26, z = -3.29, p < .001), suggesting a fluency bias. Overall, this pattern of results that there may have been a bias towards perceived fluency in the segments prior to the filled pause.

4.6 General Discussion

Disfluencies have been shown to affect language comprehension. For example, upon encountering a filled pause, comprehenders seem to prefer certain elements over others (Arnold et al., 2004; Arnold et al., 2007; Bosker et al., 2014; Heller et al., 2015) and to integrate them more easily (Corley et al., 2007), what we have referred to as the *disfluency bias*. Three accounts were put forward to account for this reported bias, based on three properties of filled pauses, namely (1) the benefits of the addition of time in the signal (i.e., the fillers-as-time account, Corley & Hartsuiker, 2011), (2) the distributional

properties of disfluencies (i.e., the token-associative account), and (3) the reasons for a speaker to be disfluent (i.e., the token-inference account, Arnold et al., 2004).

In this chapter, we argued that exploring this bias cross-linguistically would not only inform our understanding of second language comprehension in more naturalistic settings but also shed some light on the mechanisms underlying the disfluency bias. Watanabe et al. (2008) demonstrated that non-native listeners can anticipate upcoming elements in the speech signal when a disfluency cues an element that is hard to describe (i.e., the speaker struggles with conceptualizing speech). In this chapter, we tested whether the disfluency bias can also lead to expectations about trouble at the formulation stage of speech production and cue elements whose label is less accessible (e.g., low-frequency words), and if these anticipatory eye-movements are dependent on individuals' knowledge of and exposure to a language and their cognitive resources. Therefore, we replicated and extended Bosker et al.'s (2014) eye-tracking experiments to a sample of native and non-native English listeners. In the original study, native Dutch listeners followed instructions provided by either a native or a non-native Dutch speaker regarding one out of two items that had opposing frequencies (i.e., a high- and a low-frequency item) on a screen. In their study, participants displayed anticipatory eye movements towards the low-frequency word when the native but not the non-native speaker produced a filled pause.

In Experiment 1, we replicated Bosker et al. (2014) in a sample of native English listeners. Contrary to Bosker et al. (2014), we failed to find an effect of filled pauses in listeners' fixations towards a low-frequency item for native-accented speech, but instead, we found anticipatory eye movements for non-native-accented disfluencies. Further, filled pauses produced by either speaker eased the recognition of low-frequency items, and arguably impaired the recognition of the high-frequency target. In Experiment 2, we extended the paradigm to a sample of non-native English listeners. We found no effects of prediction for either speaker, but the native speaker's disfluencies aided in the recognition of the low-frequency item, whilst the non-native's aided in the recognition of both items.

A follow-up analysis comparing the results of Experiments 1 and 2 indeed showed that disfluencies produced by a native speaker led to similar eye-movements patterns in native and non-native listeners; in contrast, non-native listeners preferred to fixate more on high-frequency targets when they were produced disfluently by a non-native speaker in comparison to native listeners. A post-experimental identification task showed that native and non-native listeners were more likely to take the segment right before a filled pause as belonging to a fluent instruction for native and non-native speakers, suggesting that participants in Experiments 1 and 2 did not have evidence of an upcoming disfluency until they encountered it.

The discussion of this chapter is organised as follows. We will first review the disfluency bias and the proposed mechanisms, in light of the findings of our two experiments. We will then discuss the findings reported here with regard to the auditory stimuli employed in this study and in comparison to the qualities of disfluencies employed in previous studies. We will conclude this chapter by highlighting how future experiments could address the issues of the ones presented here.

4.6.1 The disfluency bias

At the beginning of this chapter, we put forward three mechanisms that could account for the effects previously reported of disfluencies in the face of prediction and integration: the fillers-as-time and the fillers-as-tokens views, with the latter subdivided into tokens-associative and tokens-inference. We argued that under a fillers-as-time account, disfluencies should always benefit the comprehension of subsequent items. In contrast, for the fillers-as-tokens account, the comprehension of upcoming items in speech should be benefited inasmuch they align with what the system anticipates. Specifically, for the tokens-associative account, anticipations are made on the basis of comprehenders' previous experiences of elements that contextually co-occur with disfluency. For the tokens-inference account, the disfluency bias is the outcome of reasoning about the presence of a filled pause. We further argued that while most of the literature suggests that the presence of a disfluency can lead the system to anticipate elements that co-occur with

disfluency due to the problems they pose at the conceptualisation level (e.g., harder-to-name entities), the only evidence for more detailed anticipations (i.e., problems at the formulation level, e.g., retrieving low-frequency lemmas) is that of Bosker et al. (2014).

The pattern of fixations prior to target onset, reflecting predictive processes, partially aligns with the fillers-as-tokens account. We argued that the lack of effects for disfluencies produced by a native speaker could be attributable to the characteristics of our stimuli, which was further corroborated by our identification task (what we will discuss in Section 4.6.1). The pattern of eye movements of our participants suggests that non-natives' disfluencies cued low-frequency items for native listeners, while we did not find such a pattern in non-native listeners. We take this as potentially reflecting native listeners' exposure to non-native-accented speech: While they may still hold the belief that non-native speakers are more disfluent, their experience with this property of speech (i.e., an accent), the predictions are still triggered, in line with a large body of research suggesting that whether individuals predict when attending to non-native-accented speech is partially influenced by their previous exposure to it (e.g., Porretta et al., 2017). In the case of non-native listeners, albeit they can be equally exposed to non-native-accented speech, the combination of a foreign accent with comprehending their non-native language may have prevented any predictive processing from emerging.

The time course of fixations, once participants had encountered phonological information about the target, supports an interplay between the fillers-as-time and the fillers-as-tokens views. Firstly, the results of our experiments do not fully support the fillers-as-time account as the only mechanism underlying the disfluency bias. Were this to be the case, disfluency should have aided in the recognition of the target, regardless of its frequency and the speaker's linguistic background. Crucially, listeners for whom comprehension is more difficult, such as non-native listeners, should have benefited more from the presence of a pause, and the identification of either object should have been faster (Corley & Hartsuiker, 2011). The fact that, for native disfluencies, item recognition was impacted by whether the item's frequency matched manner of delivery, i.e., the

signal was congruent (e.g., a low-frequency item produced disfluently, a high-frequency item produced fluently) aligns with the predictions of the fillers-as-tokens account. Non-native disfluencies aided in the recognition of either item regardless of its frequency for non-native listeners, what we take as evidence for the fillers-as-time account. However, it is important to note that for native disfluencies, non-native listeners benefited when low-frequency words were preceded by a filled pause. This suggests that the different effects of disfluencies in non-native-accented speech cannot be put down to a lack of association between disfluencies and low-frequency words. This opens the question of whether increased recognition of either item is solely due to increased attention to the signal.

The results described here also highlight an important point of the conceptualisation of the tokens-associative account: That of the nature of these links. Elsewhere in the literature, this account has been described as the outcome of learning the distributional properties of filled pauses. As filled pauses tend to occur with specific items (e.g., discourse-new elements, difficult-to-name objects, low-frequency words), they become a probabilistic cue that guides comprehension (Arnold et al., 2007). This opens up the question of whether disfluency is linked to several different sets of objects or to a more general set where all these objects align (namely, difficulty in production). As Arnold et al. (2007) state, conceptualising an association in such a manner requires considering in situ what is hard to produce, thus blurring the divide between the tokens-associative and the tokens-inference accounts.

One likely reason for the tokens-associative account to be conceptualised as such is because previous studies (1) have only contrasted elements that differed in one property (e.g., given versus new discourse elements, easy versus hard to name objects, high- versus low-frequency words) and (2) had visually presented these minimal pairs to participants. These designs likely give the perception that individuals have associations between disfluency and specific sets of properties. Specifically, without any visual presentation, a listener can expect any of these sets upon encountering a disfluency. Arnold et al. (2007)

argue that it is unlikely that listeners generate possible descriptions of objects in situ to then attribute disfluency and anticipate accordingly.

An alternative possibility is that, in real speech comprehension, disfluency halts predictive processing - because the filled pause signals difficulty in speech production, regardless of the reason, and this may entail that what comes is also difficult to comprehend - and thus facilitates comprehension by facilitating the integration of what follows the filled pause. This possibility allows for the tokens-associative account to describe a link between disfluency and a broad set of difficulties in speech production, as disfluencies would not trigger any prediction (and thus avoiding the question of what is predicted). The only study conducted without the presentation of visual stimuli is that of Corley et al. (2007). In an EEG study, the authors found that unpredictable words were easier to integrate if they were preceded by a filled pause, as reflected in a decreased N400 in comparison to the same words when produced fluently. Although due to the nature of their stimuli statistical comparisons where not possible (because of a lack of an appropriate baseline), visual exploration of their data suggested that predictable words preceded by a filled pause elicited a negative response compared to their fluent counterparts. This pattern could be explained by the halting of predictive processes: A predictable word is harder to integrate because it was not predicted. Due to the experimental design of Experiments 1 and 2, we cannot argue whether our pattern of results can be explained by similar mechanisms.

In sum, the results reported here provide partial support for both the fillers-as-time and the fillers-as-tokens account. Particularly, our findings suggest that one feature that can give more relevance to the tokens-associative view (in contrast to the tokens-inference) is listeners' previous exposure to certain signals, in a way that can override their beliefs about speakers. However, in cognitively demanding situations or in adverse listening situations, fillers may just aid the system by segmenting the speech signal.

Characteristics of the signal

It is important to note that the auditory stimuli of studies where predictive processes were impacted by filled pauses diverge in substantial ways from ours (cf. discussion of Experiment 1). The results of the identification experiment partially supports this idea. Participants in the present experiments, whether native or non-native listeners, may not have been aware that the native speaker was being disfluent before they encountered the filled pause itself. Previous experiments have employed disfluencies ranging from 900 ms (Bosker et al., 2014) to 1.3 s (Arnold et al., 2004), where the segment prior to the filled pause was characterized by changes in length and prosody prior to the production of the filled pause (e.g., Arnold et al., 2004; Bosker et al., 2014). In contrast, the filled pauses in the present study were relatively shorter (see Table 4.1), particularly, for the native speaker. Further, the characteristics of the disfluent segments were more likely to be taken as belonging to fluent utterances. In consequence, the availability of information about the speaker's hesitation might have been bigger in the non-native speaker condition, explaining why in this latter case, Experiment 1 did find evidence for a preference towards the low-frequency item prior to target onset.

This highlights the importance of the prosodic environment in which a disfluency occurs. In most previous work the stimuli prior to the filled pause differed in terms of prosody and length from that embedded in fluent utterances. For example, Bosker et al. (2014) reported that listeners' fixations on the low-frequency item increased prior to encountering the filled pause. Notably, in their study, participants had encountered a lengthened 'the' prior to the filled pause, which already can trigger the perception that the speaker is experiencing trouble in speech production by changing the prosody of the utterance just before the filled pause. Arnold et al. (2003) proposed that, in fact, the disfluency bias could be due to something beyond the filled pause itself, e.g., the prosody of the utterance.

This proposal opens up two venues. Firstly, it is possible that the prosodic information listeners encounter before a filled pause impacts how easy it is for them to expect

a disfluency. Secondly, the effects reported in these studies are not attributed only to a filled pause itself, but rather, to the perception of (un)confidence on the speaker's end: The properties of unconfident voices are akin to the characteristics of utterances that contain a disfluency. Confident voices are characterised by a faster rate and are lower in pitch than those perceived as unconfident (Brennan & Williams, 1995) and listeners identify doubtful voices as soon as 100 ms (Jiang & Pell, 2017, Jiang et al., 2019), and experience cue conflict when vocal confidence is accompanied with unconfident hedges (e.g., 'maybe', Jiang & Pell, 2017). In fact, there are prosodic contours that also co-occur with cognitive effort (Goupil et al., 2021). Taking this into consideration entails that a filled pause is the *overt* signal of a speaker's potential difficulties in speech production, but that the prosody of a disfluent utterance in itself could likewise trigger similar predictions to those attributed to filled pauses. This suggests that a better conceptualization of the disfluency bias is not rooted in the presence or absence of a filled pause, but rather, as reflective of listeners' tracking of the speaker's confidence, akin to epistemic vigilance (Sperber et al., 2010). The variability in stimuli employed in experiments exploring the effects of disfluency in speech comprehension points out that further investigation where the properties of stimuli are explored is warranted.

4.7 Conclusion

Speech comprehension is scaffolded on comprehenders' anticipations of what the speaker would say next. Most research on what guides anticipations has placed a large focus on the content of speech; in this chapter, we explored the modulation of anticipations depending on how the speaker talks: Do filled pauses constrain predictions? What are the likely mechanisms underlying the biases exerted by filled pauses? Across two experiments, we found evidence of filled pauses affecting the recognition of upcoming linguistic elements, although listeners for whom comprehension is more difficult (i.e., non-native listeners) did not exhibit a benefit in the same way as native listeners. While our stimuli's characteristics prevent us from drawing firm conclusions about the disfluency bias, our

findings suggest that whether a disfluency exerts an effect on comprehension depends on listeners' experience with the characteristics of the signal (e.g., accents). In the case of non-native comprehenders, disfluencies might be a more difficult cue to guide comprehension. These results partially support the idea that associations between disfluency and elements that contextually co-occur with it may drive the effects previously reported, which is potentially scaffolded by the attention-grabbing effects of disfluencies.

We present these results as a first step towards a more comprehensive understanding of how disfluencies might affect predictive processes underlying speech comprehension. Future studies should explore what properties of disfluent speech itself can guide listeners' anticipations (e.g., prosody).

Part II:

Interpretation in speech comprehension

In Part I, we discussed how disfluent speech affects one mechanism put forward to aid the efficient comprehension of speech: prediction. The results of Experiment 1 and 2 suggest that comprehenders benefit from the presence of a filled pause, particularly when it appears with elements that contextually co-occur with filled pauses, such as low-frequency words. However, this benefit depends on the speaker's and the listener's linguistic background: The benefit of a filled pause produced by a non-native speaker simply aids recognition of the upcoming linguistic item, regardless of its frequency, for non-native listeners. However, speech comprehension entails more than the processing of the words uttered: Communication involves interpreting the meaning of what we, as listeners, are told.

Experience tells us that interpreting an utterance's meaning may not be as straightforward as simply comprehending the actual meaning of a word. Imagine a friend saying 'Such lovely weather today' on a summer day when you live in Scotland, i.e., it is likely to be pouring with rain. Unless it is a rare occasion and the sun is out, it is likely that you would make the interpretation that the weather is in fact awful and our friend is frustrated by that: They were being ironic. This difference between what is said and what is interpreted is an example of a distinction between *literal* and *non-literal* meanings. In brief, the literal meaning of an utterance refers to the conventional significance of words (i.e., their semantics), and thus is bounded. Non-literal meanings, in contrast, are unbounded: There are many interpretations of the same utterance. They arise from an interaction between the utterance's conventional meaning, the listener's knowledge of a language and language use, the wider context, and even perhaps the listener's reasoning about the utterance. In a sense, non-literal meanings are what could have been meant but were not said.

Part II explores the non-literal interpretations listeners can derive from an utterance. Specifically, it asks the question: Can disfluent speech yield an interpretation different from that triggered by fluent speech? If so, what are the factors that can bias listeners' interpretations?

Chapter 5

Meaning interpretation in speech comprehension

Grice (1975) postulated that individuals share a tacit agreement on how communication ought to be: They have knowledge about the conventions of a conversation and expect others to adhere to them. This overarching agreement is captured in the Cooperative Principle: "Make your conversational contribution such as is required, at the stage in which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" (Grice, 1975, p. 45). When a message fails to adhere to these norms, it is said to convey an *implicature*, which listeners interpret via Gricean reasoning. This inference leads to an enrichment of the utterance's meaning, possibly by leading the comprehender to derive a non-literal interpretation. Consider the case of *some*, with its literal meaning said to be *some-and-possibly-all*. When asked by my family how come I am not done yet with my studies, I may explain that 'some students submit on time'. As my family allegedly expects me to be as honest and as informative as I can be (because of the Cooperative Principle), they would expect me to say that *all* students submit their thesis on time if that were the case. Therefore, they very kindly understand that some, *but not all*, students submit on time.

The case of the scalar term *some* is an example of how speakers' linguistic choices influence the meaning that a comprehender interprets. There is now a wealth of research suggesting that *how* something is said (i.e., manner of delivery) has consequences for listeners' interpretations. Consider these two examples, where the capital letter denotes emphasis:

(1a) it looks like a ZEbra

(1b) it LOOKS like a zebra

Sentence (1a) places the emphasis on the noun, while in (1b) the emphasis is on the verb. Kurumada et al. (2012) investigated whether differences in stress placement would lead listeners to different interpretations of the speaker's intended referent. In their study, participants were shown a scene with a set of four potential elements the speaker could refer to. Crucially, the element the speaker would refer to (i.e., the referent, for example, a *zebra*) was similar to another element in the display (i.e., the distractor, e.g., an *okapi*). In sentences like (1a), participants were more likely to click on the target as the speaker's intended referent. In contrast, sentences like (1b) were more likely to bias participants to click on the distractor as they interpret that the speaker meant something that looks like a zebra, but is not one. In a follow-up Visual World Paradigm (VWP) eye-tracking experiment, Kurumada et al. (2014) reported that utterances with a prosodic contour like (1b) were more likely to bias participants' fixations towards the distractor before the noun was encountered, suggesting an early constraining effect of prosody in meaning interpretation. This sensitivity to prosody as a disambiguating factor of speakers' intended meaning has even been reported when listening to non-words (Hellbern & Sammler, 2016). Overall, these results suggest that comprehenders' interpretation of meaning can be guided by factors beyond what is said, such as prosody.

In Part I, we explored how filled pauses can affect speech comprehension by guiding listeners' anticipations of what the speaker might say next. We speculated that these effects may reflect listeners' interpretations of perceived difficulty in speech production on the speaker's behalf. In Part II, we are interested in whether this perception can affect

the pragmatic comprehension of a message. To this end, in this chapter we review the evidence suggesting that, indeed, (dis)fluency can bias listeners' interpretations, with a focus on deception - a form of non-literal meaning where monitoring speakers' mental states may play a crucial role in its interpretation. To explore how disfluent speech can give rise to this interpretation, the chapter will consider theories put forward for how comprehenders interpret an utterance's meaning. Specifically, we will discuss how interpretations may be the outcome of considering different factors. In our case, we will focus on the listeners' and speakers' linguistic backgrounds. This discussion will suggest that individuals' expectations about a non-native speaker's speech and the difficulties associated with second-language comprehension offer a window to explore how disfluent speech might be taken as deceptive. This will set the scene for Chapter 6, where we will explore the interpretation of deceit as a function of these two factors in Experiments 3 and 4 respectively.

5.1 Meaning interpretation in spoken language: The role of manner of delivery as (dis)fluency

Disfluent speech has been shown to impact how a speaker is evaluated. Disfluent speakers are perceived as less intelligent (Christenfeld, 1995), less competent (Norton-Ford & Hogan, 1980) and less assertive (Schleef, 2019). Brennan and Williams (1995) demonstrated that filled pauses are taken as indices of speakers' confidence in their own knowledge. In Experiments 2 and 3, participants listened to previously recorded answers and non-answers to trivia questions, and were asked to rate how likely they thought the answer was correct (in case of answers) and how likely a given speaker was to recognise the correct answer to each question (in the case of non-answers), i.e., what the authors coined as *feeling of another's knowing* (FOAK). Answers riddled with disfluencies were more likely to receive lower FOAK ratings, suggesting that filled pauses were taken as reflective of speakers' reduced certainty about their knowledge; in contrast, non-answers (i.e., *I don't know*) were more likely to receive a higher rating if the speaker filled their

pause, i.e., listeners interpreted that the speaker was sure they did not know the answer. The authors took this as evidence that listeners are sensitive to the surface form of delivery, and in particular, to the cues displayed by speakers when they do not know (or cannot remember at the moment of being asked) the answer to a question. In fact, listeners are attentive to speakers' displays in both the visual and auditory modalities (Swerts & Kramer, 2005), and this sensitivity to cues associated with a speaker's certainty seems to arise from an early age (Hübscher et al., 2017; Krahmer & Swerts, 2005; Mori & Pell, 2019). Importantly, Brennan and Williams (1995) demonstrated that listeners' sensitivity to speakers' displays of (un)certainty were rather accurate: Listeners' evaluations were closely associated with speakers' ratings of their own feeling of knowing (FOK). This aligns with the idea that speakers may want to display their metacognitive states via paralinguistic cues such as manner of delivery (Smith & Clark, 1993, cf. Section 2.2.1).

If manner of delivery in the form of fluency is taken as reflective of the speaker's mental state, then disfluent speech may impact listeners' subsequent behaviours. For example, children prefer to learn new labels from fluent speakers (White et al., 2020, Exp 1., although labels are learnt equally well when they are produced fluently and disfluently, Exp. 2; see also Libersky et al., 2022). Similarly, Barr (2003) demonstrated that adults' learning of new labels, as well as object categorisation, was improved when a speaker's display of uncertainty was a good indicator of the prototypicality of the item.

A more extreme case would be when the speaker's perceived metacognitive state biases listeners' interpretation of their intended meaning. One example is when a speaker may want to 'save face', i.e., they want to preserve a sense of positive identity and public self-esteem (Brown & Levinson, 1987). Imagine you find your snack box open and your flatmate standing next to it with crumbs all over them. If your flatmate were to admit that they ate all of the crackers, they would lose face (e.g., perceptions of greediness). To avoid this undesirable outcome, they have at their disposal a set of strategies involving ambiguous language such as the use of *some* (Bonnefon & Villejoubert, 2005): Upon being

asked what happened, they may reply 'I ate *some* crackers', to invoke the ambiguity that *some* may entail *some, and possibly all* (its literal meaning) or *some, but not all* (its non-literal meaning). In the face of these speaker strategies, you may be more likely to interpret *some* literally (*some and in fact all*), because admitting that they ate all the crackers would make them lose face. Bonnefon et al. (2015) found that producing a silent pause before uttering *some* increased the likelihood of an interpretation where the ambiguity helped soften a negative event. In the utterance 'Some people hated your idea', listeners were more likely to take it as 'All people hated your idea'; in contrast, 'Some people loved your idea' was more likely to be interpreted as 'Not all people loved your idea'.

Loy et al. (2019) took Bonnefon et al.'s (2015) findings one step further by investigating the online interpretation of *some* as a function of disfluency. In their eye-tracking study, participants were first shown a set of snack items (i.e., several plates, each with seven snacks) to establish the number of snacks that there were available to be eaten. Participants then heard a speaker refer either fluently or disfluently to the snacks they had eaten (e.g., 'I ate *uh some* cookies') while looking at a scene where the meaning of some could be ambiguous (i.e., an empty plate, compatible with *some and possibly all*, and a plate with a number of snacks compatible with *some but not all*, that is, fewer than seven snacks) or where it could only mean some but not all (i.e., two plates only compatible with *some but not all*, that is, fewer than seven, but not empty). Participants' task was to select the image that best matched the event described by the utterance. Fluent utterances led to more non-literal interpretations (i.e., in ambiguous scenes, participants were more likely to click on the plate with a few snacks); in contrast, disfluent utterances yielded more clicks on the plate conveying *some and possibly all* (i.e., the empty plate). This click distribution suggests that disfluencies were taken as indices of the speaker's attempt to save face. In fact, the presence of the disfluency constrained the utterance's interpretation very quickly: Participants fixated on the corresponding plate shortly after encountering *some*.

This line of research suggests that disfluency is a cue that listeners can take into consideration when interpreting the meaning of an utterance. Across the board, the results map onto an idea where disfluencies may signal a speaker's difficulty in production, which in turn is interpreted by listeners as evidence of limited knowledge (Brennan & Williams, 1995), or even as concealment of the truth. In the next section, we turn to the most blatant case of a speaker hiding their knowledge: deception.

5.1.1 The case of deception

Deception involves the intention of making the interlocutor believe something known to be false by the speaker (Hala et al., 1991). On the listener's end, interpreting deceit and accurately interpreting a message's meaning involves recognising speakers' intentions, including reasoning about whether the speaker is being cooperative, their goals, and the conversational context (Franke et al., 2012) and then re-adjusting to negate the literal meaning of the sentence.

Detecting deception is a challenging task. Firstly, because listeners tend to believe speakers (Vrij & Baxter, 1999). Meta-analyses of individuals' reported abilities to accurately detect when they are being deceived demonstrate performance only slightly better than chance (Bond & DePaulo, 2006), with individual studies reporting accuracy rates between 40% and 60% (DePaulo & Pfeifer, 1986; Kraut, 1980; Vrij et al., 2000). Indeed, even individuals whose profession involves some form of deception detection (e.g., police officers) are no better than unskilled individuals (Vrij, 2004). These findings may partially reflect individuals' truth heuristic, whereby they tend to interpret information as truthful following the Cooperative Principle (Grice, 1975). Further, the lack of increased accuracy for those said to receive appropriate training may be due to the instructions they are given (Colwell et al., 2006; Frank & Feeley, 2003). Most training programmes designed to improve accuracy in deception detection assume that there is a set of behaviours (verbal and nonverbal) that liars exhibit and that individuals can learn and subsequently look for. However, some of the behaviours said to be signals of deception

have been shown to have a low correspondence to the behaviours liars actually display (DePaulo, Lindsay, et al., 2003).

Although it is hard to find reliable indicators of deception, this thesis relies on a perhaps surprising corollary: Listeners are remarkably consistent in the cues they associate with deception, whether or not those cues facilitate accurate detection. Among the common elements consistently reported to trigger interpretations of deceit are disfluencies. For example, disfluent speakers are rated as less honest (Fox Tree, 2002). Indeed, filled pauses are amongst the elements that individuals commonly reported as evidence of speakers being deceptive (Arciuli et al., 2010).

Disfluency has been shown to constrain interpretations of deceit shortly after it is encountered. Loy et al. (2017) explored the time course of the interpretation of deceit triggered by filled pauses in two Visual World Paradigm eye-tracking experiments. Participants were confronted with a potentially deceptive speaker who referred, either fluently or disfluently, to the location of a treasure as one image (out of two) on a screen. Participants' task was to select the object they believed actually concealed the treasure. Disfluent utterances were more likely to be interpreted as deceptive, as reflected in participants' preference to click on the unmentioned item, regardless of whether the filled pause was utterance-initial (Exp. 1) or utterance-medial (Exp. 2). Crucially, participants' fixations were biased towards unmentioned elements following a disfluency shortly after the target was produced, suggesting that disfluency constrained the interpretation of an utterance in the early stages of comprehension. Follow-up studies have consistently replicated this finding (King et al., 2018, Li et al., 2022), supporting the idea that the interpretation of disfluent speech seems to reflect a form of social reasoning. For example, King et al. (2018) reported a delay in listeners' fixations towards the unmentioned element (i.e., reflecting an interpretation of deceit) when there was an alternative explanation for the speaker to produce a filled pause (e.g., a car horn). Likewise, individual differences in theory of mind seem to impact the emergence of the interpretation of deceit, with those who score higher in the Autistic Spectrum Quotient (i.e., individuals with more traits

associated with autism) showing a delay in their fixations towards unmentioned elements following a disfluency (Li et al., 2022).

One naturally arising question is whether this social reasoning occurs every time a listener encounters a disfluency or if it reflects a heuristic. The former mechanism entails a slow and cognitively-demanding process, specific to each instance when a filled pause is encountered. The latter represents a faster and relatively cost-free mechanism that is applied indiscriminately (i.e., context insensitive). To understand how this bias in the interpretation of deceit operates, we review the accounts put forward to explain how comprehenders interpret meaning. Although these accounts are not typically applied to deception as a form of non-literal meaning, and have not been developed for disfluency, they propose specific properties for the emergence of non-literal meaning that may guide our understanding. In particular, they highlight how different factors can modulate what comprehenders interpret: For example, comprehenders may weigh different cues available in a conversation to interpret a speaker's intended meaning. Specifically, we will review how differences in speakers' linguistic backgrounds can lead listeners to interpret a message differently, arguably due to stereotypes about speakers' abilities to produce speech. Given that interpreting a message is the outcome of the consideration of different cues, we will review how differences in listeners' linguistic backgrounds may lead them to consider different cues. Overall, this discussion will yield contrasting predictions for the emergence of the interpretation of deceit, and importantly, it predicts different time courses when additional factors are included in the communicative context.

5.2 Accounts of meaning interpretation in language comprehension

One way in which we can approach how meanings are interpreted is by assuming that there is a meaning that takes precedence. The standard pragmatic model (Grice, 1975) posits that comprehenders first interpret the literal meaning of a sentence. If the comprehender recognises that there is a violation of a Gricean maxim, they compute the non-literal

meaning. Consequently, non-literal meaning arises as a secondary step in comprehension. In this sense, pragmatic meanings occur in a secondary step. In sharp contrast, some authors have proposed that non-literal meanings are immediately available for comprehenders i.e., there is only one step. Each approach makes contrasting predictions about the emergence and costs associated with inferring meaning: Specifically, two-step models expect non-literal meanings to arise later and to be costly, while one-step models predict that non-literal meanings can arise as quickly as literal ones, without necessarily incurring cognitive costs.

There is experimental evidence supporting both approaches. Bott and Noveck (2004) had participants rate sentences' felicity. In sentences like 'Some elephants are mammals', a non-literal interpretation (i.e., *some but not all*) is false, while its literal interpretation is true (because *all* elephants are mammals). In line with the standard pragmatic model account, the authors found that accurately responding to sentences whose non-literal meaning was false led to longer reaction times. This delay, compared to those sentences where the literal meaning was sufficient to respond accurately, is taken to be reflective of the additional computation associated with processing non-literal meaning (see also Bott et al., 2012; Tomlinson et al., 2013).

It is important to note that these studies confounded non-literal meanings with violations of world knowledge: The delays and costs found could instead mirror participants' understanding that the non-literal meaning does not match their knowledge about the world, and thus they take longer to respond to these sentences. Nonetheless, in scenarios where the interpretation of *some* is not confounded with world knowledge, listeners still take longer to respond when an appropriate response involves a non-literal meaning. In a series of eye-tracking experiments, Huang and Snedeker (2009) had participants select, from a set, the scene that best match a description. For example, for the utterance 'Point to the girl that has some of the socks', participants saw a display with a girl that had two out of four socks (and thus matched both a literal and a non-literal interpretation of *some*) and a girl with three of three socks (and thus matched the literal interpretation

of *some*). Until participants encountered the noun (i.e., *socks*), the appropriate scene was ambiguous in that either girl could be the referent: Disambiguation depended on how the listeners interpreted *some*. Huang and Snedeker (2009) found listeners showed delays in choosing scenes representing a non-literal interpretation of *some* when a literal (*all*) scene was also available.

The evidence supporting the standard pragmatic model falls short to explain the effects reported by Loy et al. (2017, see also King et al, 2018; Li et al., 2022). Bott and Noveck's (2004) and Huang and Snedeker's (2009) findings would predict that the interpretation of deceit should be costly following a disfluency. In Loy et al.'s (2017) measures, this should be reflected in participants' delayed preference to fixate on the unnamed object as the location of the treasure - a more fine-grained measure than reaction times. However, the authors reported that participants' time course of eye movements were biased towards the element on screen that reflects an interpretation of deceit shortly after encountering a disfluency - specifically, visual inspection suggests that this bias emerged 400 ms post-noun onset. The speed with which the constraining effect of disfluency emerged runs against the predictions of the standard pragmatic model.

In contrast, Grodner et al.'s (2010) findings support a one-step model of deriving pragmatic meaning. In a VWP study, participants were first introduced to the total number of available objects (e.g., four balloons, four balls, four planets). Participants were then presented a visual display depicting six figures, three males and three females: Per gender, each character could have all the tokens of an object (e.g., all the balloons), another could have a subset (e.g., two out of four balloons), and another had no items. Participants were asked to click on the appropriate referent following the instructions, which could direct them to click on a character *summa* (replacing *some of*), *alla* (replacing *all*), or *nunna* (replacing *none*). Upon encountering *summa*, participants fixated shortly after on the character representing a *some but not all* interpretation. The speed with which non-literal meaning emerged in Grodner et al. (2010) parallels that reported by

Loy et al. (2017), supporting the idea that non-literal meanings can be quickly computed, without the necessity of deriving the literal meaning first, in line with one-step models.

In line with one-step models, the speed with which a meaning is derived is sensitive to the factors present that support either. Breheny et al. (2006) manipulated the salience of either meaning of *some*. For example, when the sentence 'John replied that he intended to host some of his relatives' is used as an explanation as to why John is cleaning his apartment, the non-literal meaning of *some* is enough to form an informative argument: Only one guest is needed for someone to decide to clean. In contrast, when John says the same sentence as a reply to a question about whether he intends to host all of his relatives in his apartment, *some and possible all* is also a possibility. Breheny et al. (2006) found that contexts where the interpretation of *some but not all* was informative ('host all?'), participants were faster at computing the non-literal interpretation. Conversely, in scenarios where *some and possibly all* was more informative (reason for cleaning), interpreting *some* literally was faster.

The notion that context, and in particular, the cues available in context to support an interpretation, can explain what meaning is interpreted first is best exemplified by constraint-based models (Degen & Tanenhaus, 2015, 2019). In this account, a context provides a certain set of factors that are considered when interpreting meaning. Each factor receives a different weight as a function of its relevance and usefulness in context. As a consequence, the probability of a particular interpretation arising, as well as the speed with which it is computed, are dependent on the support from the multiple cues that are considered. Constraint-based models predict that costs associated with meaning interpretation are not dependent on what meaning is interpreted. Rather, costs arise in situations of cue conflict, or for scenarios of unbounded inferences (Degen & Tanenhaus, 2019). In this way, constraint-based models offer an elegant solution to how both literal and non-literal meanings can be computed, although identifying the constraints that are involved in meaning interpretation is one of the challenges of constraint-based models (cf. Degen & Tanenhaus, 2019). There is now an incipient body of research exploring

different factors that can bias what meaning is interpreted, ranging from the salience of the alternative meaning (Bott & Frisson, 2022) and its relevance to the discourse (Politzer-Ahles & Fiorentino, 2013), to the role of task instructions (Degen & Goodman, 2014).

Thus far, we have paralleled interpreting deception with quantity implicature (i.e., interpreting *some*). It is worth noting that, as discussed in section 5.1.1, interpreting deceit is a special case because the interpreted meaning is the opposite of what the speaker says and thus requires recognising the speaker's deceitful intentions, which are not intentionally communicated. This differs from scenarios where interpretation meaning requires disambiguating between two options (e.g., the literal and non-literal meanings of *some*) and, specifically, scenarios where speakers' intended meaning is communicated. However, the literature reviewed can serve as an anchor for us to understand whether and how deceit is interpreted. Specifically, Loy et al.'s (2017) findings align with a large body of research suggesting that non-literal meanings are rapidly computed, and in particular, that manner of delivery (in the form of disfluency) can be easily integrated to infer a speaker's intentions.

In the case of interpreting deceit, we have only considered one source of information: disfluency. However, understanding language usually requires combining different sources of information. A constraint-based model can also shed light on studies exploring the combination of disfluency with additional sources of information. In King et al. (2018), the presence of an alternative cause for a speaker to be disfluent (i.e., a car horn) led to a delay in the emergence of the interpretation of deceit, as reflected in participants' fixations. This aligns with findings suggesting that linguistic elements (or choices) are contrasted against other cues to interpret meaning, including speaker identity. For example, Bergen and Grodner (2012) reported that the perceived knowledgeability of a given speaker impacted the interpretation of *some*. Specifically, a knowledgeable speaker uttering *some*, when expected to say *all* (because they arguably know), was harder for listeners to process. Further studies have shown that other properties of the speaker, such

as their perceived reliability (Grodner & Sevidy, 2011) or their particular uses of language (Kurumada et al., 2014a; Yildirim et al., 2016), can affect meaning interpretation.

Across the board, this review suggests that interpreting pragmatic meanings is neither necessarily costly nor relegated to a second step. We have discussed how Loy et al.'s (2017) findings align with a broader body of research suggesting that contextual cues can be rapidly integrated to constrain meaning. However, few studies have explored how different sources of information can interact to interpret meaning. Constraint-based models make an interesting prediction regarding the combination of different cues: In scenarios involving cognitive costs, cue conflict may yield processing difficulty that leads to more interpretations of the meaning supported by the context (Degen & Tanenhaus, 2019). Therefore, one way of exploring what underlies the biases in interpretation exerted by disfluent speech is to investigate the effects that speaker and listener attributes may have.

5.2.1 Speaker's linguistic background: Non-native speakers

In Section 3.2.1, we discussed how speaker identity in terms of linguistic background (i.e., native versus non-native) can have consequences for how speech is processed. Specifically, we reviewed Lev-Ari's (2015) proposal that non-native speakers are stereotyped as having lower linguistic competence, and thus listeners have certain expectations of how these speakers will use language. This a priori attitude was put forward as an explanation of a wealth of research suggesting shallower processing of non-native-accented speech by native comprehenders (e.g., less syntactic revision, Hanulíková et al., 2012; more acceptability of ungrammatical sentences, Brehm et al., 2019; Gibson et al., 2017).

The perceived lower competence of non-native speakers may map onto a perception of unreliability, which has been shown to affect how comprehenders interpret speech. In line with this hypothesis, Caffarra et al. (2018) investigated the perception of ironic statements depending on the speaker's linguistic background. Arguably, different types of irony require more or less linguistic competence: Ironic criticism (e.g., *'You're such a*

great chef' after an individual has burnt a meal) is a much more common construction than ironic praise (*'You're such a horrible chef'* after an individual has won a Michelin star). Consequently, the former form of irony requires less linguistic competence than the latter, as it is more frequent and easier to learn. While ironic criticism was perceived similarly when produced in a native or a non-native accent, ironic praise produced by a non-native speaker was rated as less ironic than its native counterpart (see also Bazzi et al., 2022, for similar findings).

The stereotype of non-native speakers as less skilled can also explain the differences in the interpretation of underinformative sentences. Fairchild et al. (2020) presented participants with a written story where the main character, who could be a native or a non-native speaker, failed to give all the information to another character. Participants were then asked to say why they believed the speaker did not provide all the information. Native speakers who provided less information than necessary were more likely to be rated as unwilling to share the information. The same pragmatic failure by a non-native speaker, however, was taken as a sign of inability to produce the necessary information (Exp. 1), even when participants are not explicitly informed that the non-native speaker could experience language difficulties (Exp. 2). Further, this 'forgiveness' for underinformative statements had consequences for participants' subsequent behaviours: They were more likely to learn new information from a previously encountered underinformative non-native speaker than from a native speaker (Exp. 3 and 4). Similarly, written underinformative statements are more likely to be accepted and rated as trustworthy when reported to have been produced by a non-native speaker (Lorenzoni et al., 2022). As the experimental stimuli in these studies were written materials, these findings cannot be attributed to shallower processing of non-natives' statements due to difficulties in non-native-accented speech comprehension (e.g., Bent & Bradlow, 2003; Munro & Derwing, 1995), and in fact, similar findings are reported when using auditory stimuli (Ip & Papafragou, 2022). These results have been interpreted as showing how social expectations about non-native speakers' linguistic abilities can affect sentence interpretation and subsequently impact comprehenders' behaviours (Fairchild & Papafragou, 2018).

These findings support the hypothesis that speaker identity as conceptualised in their linguistic background is a constraint that can guide meaning interpretation. The evidence reviewed here suggests that a stereotype about non-natives' reduced linguistic abilities modulates the interpretations drawn for their speech. A perception of reduced linguistic competence might encompass how capable non-native speakers are of producing speech fluently, so that their speech is stereotyped to contain more filled pauses. If this is the case, then this perception is a potential factor that could interact with manner of delivery and subsequently yield different interpretations of disfluent speech: Specifically, listeners may be more lenient with disfluent non-native speakers and thus be less likely to take their filled pauses as indicators of deceit.

5.2.2 Listener's linguistic background: Non-native comprehenders

In Section 3.2.2, we saw how second-language comprehension may differ from first-language comprehension in terms of predictive processing. Specifically, we argued that previously found differences in speech comprehension are not due to divergent mechanisms; instead, they arise as a consequence of particular characteristics of non-native comprehenders that tap into factors that can affect predictive processing in general (e.g., experience with and exposure to the target language, cognitive load; see Kaan, 2014). Meaning interpretation is also said to tap into factors such as theory of mind, working memory (Fairchild & Papafragou, 2021), and individuals' ability to handle cognitive load (De Neys & Schaeken, 2007) - characteristics said to differ between native and non-native comprehenders. For example, non-natives are said to have higher theory of mind abilities (Schroeder, 2018; yet cf. Feng et al., 2023), to be more socially aware of their interlocutor (Fernández-Rubio & Glucksberg, 2012) and to display higher executive functioning (Bialystok, 2015; Wu & Thierry, 2013; yet cf. Grundy, 2020). Likewise, second-language comprehension is likely to impose a cognitive load on individuals (Ito & Pickering, 2021). Under a constraint-based account, meaning interpretation in second-language comprehension may thus differ from that of first-language comprehension since these properties

impact the cues considered to interpret meaning, as well as the weight given to each of them.

One way in which non-native comprehenders may differ in how they interpret meaning is whether they show a preference for interpreting either literal or non-literal meanings. Research exploring the interpretation of *some* has reported conflicting results: While some authors have reported that non-native comprehenders are more likely to interpret *some* non-literally (Lin, 2016, Slabakova, 2010), other studies have found no differences in how it is interpreted between native and non-native comprehenders (Antoniou et al., 2018; Dupuy et al., 2019; Snape & Hosoi, 2018). Mazzaggio et al. (2021) argued that the differences between these studies may lie in participants' experienced cognitive load: Slabakova's (2010) participants were immersed in the country of their second language, which may have decreased their experience of cognitive load. To counteract this, Mazzaggio et al. (2021) tested the interpretation of *some* in underinformative statements in a sample of non-native English and Spanish (L1: Italian) listeners, who were residing in the country of their native language (Italy). Additionally, the authors increased task demands by imposing a time limit, to ensure that participants were in a cognitively demanding scenario. Across two experiments, the authors reported that non-native listeners were more likely to interpret *some* literally as they were more likely to accept underinformative statements.

Schulz (2021) points out the fact that the difference between studies that found an effect of listeners' non-nativeness (regardless of the direction) and those that did not is the presence of visual context. The former experiments employed visual contexts, while the latter did not. It is possible that this factor can partially account for the contrasting findings. Additionally, stimuli in both Slabakova (2010) and Mazzaggio et al. (2021) employed sentences that also violated world knowledge, while Snape and Hosoi (2018), Antoniou et al. (2018) and Dupuy et al. (2019) used stimuli in which underinformativeness referred to a context set by the experiment (e.g., how many cars are hidden in a visual scene, Dupuy et al., 2019). Therefore, a priori differences may

have more likely been attributed to processing costs associated with violations of world knowledge in second-language comprehension (cf. Foucart et al., 2016; Romero-Rivas, 2017). In contrast, studies investigating other kinds of non-literal meanings such as irony have found that native and non-native comprehenders interpret meaning similarly, and at similar speeds (Bromberek-Dyzman et al., 2021, Tiv et al., 2021).

Another manner in which second-language comprehension can yield different interpretations from first-language comprehension is in what cues are considered, as well as the probabilities attached to them. Indeed, an interesting proposal of constraint-based models is the possibility that a factor can be assigned different weights, which may be comprehender-specific. The associated cognitive costs to second-language comprehension may influence comprehenders' strategy (Clahsen & Felser, 2006) so that their processing is guided by cues that do not impose an additional cost (Futrell & Gibson, 2017).

Starr and Cho (2022) investigated how the Question Under Discussion (QUD) affected the interpretation of *some* in native and non-native comprehenders. In an acceptability judgement task, where participants had to evaluate the answer to a question given a scene, native comprehenders' judgements depended on how the question was formulated: The use of *some* for questions including *all* when all targets met a condition led to lower acceptability ratings than when the question included *any*. In contrast, non-native speakers did not show any effects of QUD. One potential explanation for this is that non-native comprehenders do not show the same degree of sensitivity to structural factors (e.g., differences in how the QUD was posed). In contrast, factors that are well-mastered and language-independent, such as speaker identity, seem to constrain interpretations alike in native and non-native comprehenders. Zhang and Wu (2022) followed Bergen and Grodner's (2012) study where participants read narratives that set expectations about the speaker's knowledgeability. The authors found that native and non-native speakers behaved similarly, depending on the speaker's ascribed knowledge in how they interpreted *some*. This aligns with findings whereby non-native listeners display

earlier responses than native listeners to utterances where the content mismatches the speaker's stereotypes (Foucart et al., 2015).

The research reviewed here suggests that non-literal interpretation in second-language comprehension may differ as a function of what cues are preferred by these listeners, following a constraint-based account. Specifically, they may have a preference to interpret meaning following cues that they have already mastered (e.g., speaker identity), arguably to compensate for the cognitive costs associated with second-language comprehension. In the case of deception, the evidence presented here suggests that non-native listeners are likely to develop interpretations in a fashion similar to natives as disfluency is arguably a contextual cue. However, the fact that they are more interlocutor-oriented may entail that in the presence of alternative reasons for the speaker to be disfluent (e.g., a non-native speaker), the consideration of these two cues (disfluency and speaker identity) not only may lead to a delay in the emergence of the interpretation, but also they may be more likely to prompt a consideration of the speaker's identity.

5.3 Conclusion and Chapter aims

In this chapter, we have discussed how disfluent speech may not only affect predictive processing but also the interpretations that listeners derive from an utterance. There is evidence suggesting that the presence of a filled pause leads comprehenders to derive different interpretations than those that arise for fluent speech: For example, a disfluent utterance is more likely to be taken as evidence of the speaker's attempt to save face (Loy et al., 2019) or as the speaker's attempt to deceive (Loy et al., 2017; King et al., 2018; Li et al., 2022).

In this thesis, we have taken the interpretation of deceit following disfluent speech as a test bed to investigate how disfluent speech is interpreted. The evidence thus far suggests that it can be explained by listeners' social reasoning about the causes for the speaker to be disfluent. In the face of potential deceit, listeners attribute the disfluency to the cognitive costs associated with producing a lie. The speed with which these

effects emerge supports accounts where non-literal meaning is readily available for comprehenders as a function of the context, such as constraint-based models. As a speaker's perceived intention is a possible constraining cue (Degen & Tanenhaus, 2019), it can interact with the presence of a disfluency to yield an interpretation of deceit - suggesting that this interpretation reflects social reasoning on the listeners' end.

One interesting property of constraint-based models is their probabilistic stance on meaning interpretation. In this chapter, we have reviewed two potential factors that can impact the interpretation listeners derive: The speaker's and the listener's identity as a function of their linguistic background. This review showed that comprehenders interpret speech produced by a second-language speaker differently, and commonly under the light of a stereotype where these speakers are believed to be less competent. We also discussed how non-native comprehenders seem to weigh factors differently, depending on their dexterity with specific cues in their second language.

Understanding the interpretation of disfluent speech as deceitful under a constraint-based model offers a window to explore how it occurs. If listeners take into consideration relevant cues to interpret meaning, then the perception of a speaker who may produce more disfluencies by virtue of who they are (i.e., a non-native speaker) may interact with the interpretation of deceit. The presence of conflicting cues (i.e., speaker identity, disfluency) may have an effect on how listeners who are in a cognitively demanding situation (i.e., non-native listeners) interpret disfluent speech. Exploring these parameters is particularly relevant because the social reasoning account for this interpretation relies on a form of mentalising (Goodman & Stuhmüller, 2013) that is bounded by the conversational context in the experiments described (i.e., the expectation that the speaker may lie at some point, in Loy et al., 2017).

In Chapter 6, we explore how disfluent speech can be interpreted as deceiving when there are multiple sources of information. Chapter 6 describes two eye-tracking

experiments set up following Loy et al. (2017), where non-native speaker identity and non-native language comprehension were used to investigate the flexibility of the interpretation of disfluent speech as deceitful.

Chapter 6

Disfluency as Deception

Not every word we utter is truthful. In fact, most people admit to lying more than once a day (Serota et al., 2010; Serota & Levine, 2014). As we discussed in Section 5.1.1, comprehenders are not particularly good at detecting deceit (Bond & DePaulo, 2006). Their no-better-than-chance ability to distinguish the truth from deception can be partially explained by pre-determined biases: For example, listeners tend to assume that their interlocutor is truthful (Grice, 1975; Vrij et al., 2000), while at the same time, they hold mistaken beliefs about what deception sounds and looks like (DePaulo & Pfeifer, 1986; Henningsen et al., 2005; Nahari & Ben-Shakhar, 2013; Zuckerman, Koestner, et al., 1981). One relevant feature ascribed to deceptive behaviour that can decrease listeners' accuracy in identifying deception is the stereotype of deception leading to 'flawed speech', what we will refer to as the *disfluency-as-deception bias*. Liars are expected to use filled pauses such as *uh*, stutter more, and produce longer pauses (Global Deception Research Team, 2016), although, in actuality, this feature is not strongly associated with deceit (DePaulo, Lindsay, et al., 2003; Loy et al., 2018; Sporer & Schwandt, 2006).

In scenarios where deceit is expected, there are two ways in which cues might trigger inferences of deception. On the one hand, the presence of a cue such as a filled pause could immediately trigger an interpretation of deceit: For example, if a speaker hesitates, a listener may just rely on a stereotype whereby liars are disfluent and interpret that

the speaker is lying. We will refer to this as an *associative account*. On the other hand, because the interpretation of a speaker's intention requires social reasoning, the presence of a cue might be compared against listeners' expectations, and lead them to reason about the presence of said cue. In the case of disfluencies, a listener might believe that disfluency is the outcome of language production under cognitive load, and if the speaker is experiencing trouble in speech production, then a possible explanation is that they are trying to be deceptive. We will refer to this as an *inference account*. Importantly, these two accounts have consequences regarding the emergence of the disfluency-as-deception bias across contexts.

In this chapter, these two accounts are explored in two eye-tracking experiments, in which speakers name treasure locations, possibly deceptively and possibly disfluently. Each of them exploits a difference in the two accounts that should yield different time courses; namely, the flexibility of the disfluency-as-deception bias (Exp. 3) and the cognitive costs associated with its emergence (Exp. 4). In Experiment 3, we will test whether the disfluency-as-deception bias is flexible by manipulating the speaker's linguistic background (i.e., whether they are producing speech in their first or second language). In Experiment 4, we will test whether the disfluency-as-deception bias is a cognitively costly interpretation by testing participants who are comprehending speech in their second language and thus are expected to be in a cognitively demanding task (Segalowitz & Hulstijn, 2005). Across the board, these studies will show that in situations where deceit is expected, interpretations of deceit in the presence of disfluency are pervasive - even in the face of alternative reasons for its occurrence and increasing task demands, in line with an associative account. It is important to note that in the experiments presented here, participants are not only forewarned that the speaker will be deceitful, but they expect them to do so at least half of the time, and thus this expectation might have influenced how they relied on the cues available to inform their decisions.

The structure of this chapter is as follows. In the first section, we will review the evidence for the disfluency-as-deception bias. Of particular interest is its time course:

When do listeners integrate an interpretation of deceit in the face of disfluency? From there, we will discuss how an associative and an inferential account can account for the disfluency-as-deception bias and discuss how these two accounts differ in terms of the flexibility of the bias, and present Experiment 3. From these findings, we will discuss a second factor in the bias: that of cognitive resources, and present Experiment 4. To further explore the disfluency-as-deception bias, we will compare the results of these two experiments as well as present an exploratory analysis of the precise time course of the bias.

6.1 The disfluency-as-deception bias

The evidence that filled pauses can bias listeners to believe the speaker is being deceptive can be found throughout the literature (cf. Section 5.1.1). Most of these studies involve asking comprehenders to rate a speaker's honesty or deceit after listening to pre-recorded audio. Fox Tree (2002) reported that recordings of disfluent speakers were rated as less reflective of what the speaker truly thought. Research done on deceit interpretation and detection shows that filled pauses have been consistently reported as the elements that bias individuals towards interpretations of deceit (Arciuli et al., 2010), which is particularly salient when listeners expect deceit (Akehurst et al., 1996; Sporer & Schwandt, 2006; Zuckerman, Koestner, et al., 1981).

A growing body of research is turning its attention to the real-time emergence of interpretations triggered by filled pauses. Notably, Loy et al. (2017) explored the real-time processing of disfluent speech when it is interpreted as deceitful. In their eye-tracking study, participants were instructed to select one out of two items depicted on a screen as the location of some hidden treasure. Importantly, the only information participants had to inform their decision came from a potentially deceitful speaker who referred either fluently or disfluently to one of the two items as the location of said treasure. Upon encountering a disfluent utterance, participants were less likely to select the named location as the actual location of the treasure, regardless of whether the disfluency was

utterance-initial (e.g., ***Uh**, the treasure is behind the* [item], Exp. 1) or utterance-medial (e.g., *The treasure is behind **thee uh*** [item], Exp. 2). Eye-movement analyses showed that, compared to fluent instructions, disfluent ones were promptly followed by fewer fixations to the named location. The fact that the effect was closely time-locked to the disfluency suggests that it was that specific aspect of the speech which caused listeners to doubt the speaker's truthfulness, rather than a more general difficulty in understanding disfluent speech.

6.1.1 Deception bias as an association

One explanation for the effects of manner of delivery relies on the fast application of associations. In Section 3.1.1, we discussed that when predicting upcoming linguistic units, the presence of a filled pause constrains the set of expectations for upcoming elements in the linguistic signal: Shortly after encountering a filled pause in speech, listeners display anticipatory eye movements towards harder-to-name objects (Arnold et al., 2007), discourse-new entities (Arnold et al., 2004), and low-frequency items (Bosker et al., 2014). Under an associative account, these biases are explained as the outcome of individuals' learning of what co-occurs with disfluency (e.g., harder-to-name objects, discourse-new entities, low-frequency words) via exposure to the distributional pattern of a language.

The disfluency-as-deception bias reported by Loy et al. (2017) can likewise be explained by this mechanism. Individuals hold a priori theories on how deception sounds, whereby filled pauses are to be expected in deceptive speech. The origins of this stereotype are unclear (interlocutors rarely get to learn whether they were being deceived or not), but a potential explanation involves individuals' experience with speakers' filled pauses and difficulty in production. As deception is believed to be cognitively costly (Vrij et al., 2000; Vrij et al., 2008), one can expect that a speaker's increased cognitive load leaks onto their behaviours (Ekman & Friesen, 1969). As a consequence, individuals stereotype disfluent speech as deceptive. In experimental designs such as that of Loy et

al. (2017), where deceit is expected, this interpretation is computed automatically when encountering a filled pause.

Understanding the disfluency-as-deception bias as the outcome of an association has important consequences for its emergence. It entails an automatic computation, in that manner of delivery is the sole cue that the system considers when interpreting deceit. As no other cues are considered, the computation is relatively fast and thus the interpretation emerges quickly shortly after encountering a filled pause, and without requiring cognitive resources from the system. Further, understood as the by-product of a stereotyped association, it is technically inflexible: Upon encountering a disfluency in a context where deceit is expected, individuals should always exhibit this interpretation; in this regard, the associative account posits that manner of delivery occupies a privileged position for the interpretation of deceit.

6.1.2 Deception bias as an inference

The effects of filled pauses in deriving an interpretation can be alternatively seen as the outcome of social reasoning (cf. Goodman & Stuhlmüller, 2013). In the case of predictive disfluencies, in Section 3.1.1 we reviewed how the constraints exerted by the presence of a filled pause are subject to contextual factors; specifically, the question of who produces the disfluency has consequences for what is (or is not) expected. Anticipatory eye movements towards harder-to-name entities following disfluencies are attenuated when there are alternative explanations for the speaker to be disfluent (Arnold et al., 2007, Exp. 2; Bosker et al., 2014, Exp. 2). These modulations can be understood under an inference account: Listeners combine speakers' identity (or perspective) with manner of delivery to guide their comprehension. This account highlights the role of speaker identity in language comprehension (e.g., Berkum et al., 2008; Kleinschmidt and Jaeger, 2015).

Inferences about the speaker could equally explain the disfluency-as-deception bias. This mechanism, prima facie, would be similar to that underlying the associative account. Listeners' awareness of the correlation between problems in language production

and filled pauses in a context where cognitive load can be explained by speakers' deceitful intentions leads them to reason that the presence of a disfluency is reflective of speakers' deceptive intentions. In contrast to the associative account, an inferential one proposes that, all things being equal, disfluency will be a cue favouring the interpretation of deceit when it is the only cue present in the context to support this interpretation. However, because listeners are expected to reason about the speaker, one important prediction of this account is that the disfluency-as-deception bias can be modulated.

In Loy et al. (2017), listeners only had to reason about why the disfluency was there against a context of expected deceit. Take, however, a situation in which there is another explanation for the speaker to produce a disfluency: Under an inference account, manner of delivery would be considered alongside the perceived cause for the speaker to produce speech in such way. A by-product of this consideration of different cues is an increase in cognitive demands, because the set of alternatives that can be activated requires working memory resources (Chierchia et al., 2001; Gotzner & Spalek, 2017). Consequently, the emergence of the bias in the time course of speech comprehension may be delayed, reflecting the system's consideration of different, opposing cues at the same time.

We have reviewed two potential mechanisms for the disfluency-as-deception bias emergence. These two accounts represent two contrasting ways in which cues would be weighed to interpret deceit in disfluent speech (Degen & Tanenhaus, 2015, 2019). Under an associative account, manner of delivery heavily influences the interpretation of disfluent speech and overrides the influence of any other potential cues because it is rapidly available for listeners as it is the output of a stereotyped association. Under an inference account, manner of delivery influences the disfluency-as-deception bias because listeners reason about why the speaker was disfluent. Consequently, it may not override other cues available in the signal from the early stages of comprehension. It is important to note that it is likely that disfluent speech will be interpreted as indexing deceit. What is of interest for us is the underlying processes i.e., the *timing* of an emerging interpretation:

The associative and inference accounts herein described entail that this emergence will differ depending on whether listeners are solely relying on a stereotyped association or are considering different cues equally.

6.2 Experiment 3: The role of speaker identity

Experiment 3 was designed to test one of the key distinctions between the two proposed mechanisms: their flexibility. Specifically, we were interested in the emergence of the disfluency-as-deception bias depending on the presence of alternative cues (beyond the possibility of deceit on the speaker's behalf) for the speaker to produce disfluent speech.

Of course, speakers can produce *um*s and *er*s for many reasons which have nothing to do with deception. One reason for a speaker to be disfluent is if they are producing speech in their second language (i.e., they are non-native speakers). Speaking a second language is cognitively demanding (Gregersen, 2005) and many of the difficulties associated with language production in a second language (e.g., word-finding difficulties, Pivneva et al., 2012) can be associated with disfluency (e.g., Beattie & Butterworth, 1979; Schachter et al., 1991). Indeed, non-native-accented speech is associated with increased disfluency rates overall (Davies, 2003), and is perceived as such by first-language listeners (Pinget et al., 2014).

In Section 5.2.1, we discussed how speech produced by a non-native speaker can be interpreted differently from that produced by a native speaker. The literature reviewed suggests that these differences can be accounted for by a stereotype of non-native speakers as being linguistically less competent than native speakers, so their speech is interpreted through this lens (Lev-Ari, 2015). Disfluency can be taken as evidence of lower linguistic competence and, consequently, disfluent speech would align with listeners' expectations of non-native speakers as being less linguistically skilled and thus influence the disfluency-as-deception bias.

It is also important to note that a speaker's linguistic background in itself can bias veracity assessments. The deception literature consistently reports a bias against non-native speakers: Some authors have reported a lie-bias against non-native speakers (Da Silva & Leach, 2013; Elliot & Leach, 2016; Leach & da Silva, 2013; Leach, Da Silva, et al., 2017; Leach et al., 2020), while others report a truth-bias towards native speakers that non-native speakers do not benefit from (Castillo et al., 2014; Cheng & Broadhurst, 2005; Evans & Michael, 2013; Evans, Michael, et al., 2013; Evans & Michael, 2014; Evans, et al., 2017). These biases have been explained in terms of processing fluency (i.e., the ease with which a stimulus is processed, Oppenheimer, 2008; see Dragojevic et al., 2017, Lev-Ari & Keysar, 2010), group membership (Turner et al., 1979), and a combination of both factors (Mai & Hoffmann, 2013; see also Foucart & Hartsuiker, 2021). Importantly, these biases might just reflect the fact that non-native speech happens to contain more elements that listeners attribute to deception (Castillo et al., 2014).

In order to explore these issues, Experiment 3 replicates and extends Loy et al.'s (2017) eye-tracking treasure-hunt paradigm, in which participants had to guess the locations of treasure based on the instructions provided by a potentially deceitful speaker. As in Loy et al. (2017), manner of delivery was manipulated by presenting half of the critical utterances fluently (*The treasure is behind the...*) and the other half disfluently (*The treasure is behind* ***thee uh****...*) within (listening) participants. In addition, we manipulated the nativeness of the speaker between-participants: Half of the participants heard a native English speaker, while the other half heard a non-native (L1: Spanish) speaker of English. In each trial, participants were instructed to click on the location (out of two) that they thought concealed the treasure, and their eye movements were recorded throughout.

To the extent that judgements are driven by the specific cue of disfluency, disfluent utterances should be more likely to be judged as deceitful than their fluent counterparts, regardless of who the speaker is. In this case, the disfluency-as-deception bias can be

attributed to associations that listeners hold between disfluency and deceit: Although alternative causes of disfluencies have been shown to exert an influence on prediction (e.g., Arnold et al., 2007; Bosker et al., 2014), it may be that they are not of much consequence in the context of deception because the bias emerges as a consequence of an association. Alternatively, it may be the case that listeners model the speaker more closely: If listeners attribute non-native speakers' disfluencies to general production difficulties rather than any cognitive effort associated with lying, then disfluency-driven evaluations of deceit should be less evident for speech from non-native than that from native speakers. Specifically, if the disfluency-as-deception bias emerges due to the fast application of an association, there should be no differences in the time course depending on the speaker's linguistic background. However, if listeners are reasoning about why a speaker is disfluent, then the time course of the emergence of this bias should differ between speaker conditions, with a delay for the non-native speaker (cf. Roettger & Franke, 2019).

6.2.1 Methods

All materials, including experimental script and analysis can be found at https://osf.io/9s6jv/.

Participants

Eighty-three self-reportedly monolingual speakers of British English, born and raised in the United Kingdom, took part in the study. Participants were recruited from either the general population or from the University of Edinburgh student pool. Participants were aged between 18 and 30 years, had no hearing impairments and had normal or corrected-to-normal vision. All participants gave their informed consent, as approved by the PPLS Ethics Committee (ref no. 294-1718/9). Participants were reimbursed with either £5 or university credit, and debriefed after the experiment.

Participants were excluded from the analysis if they did not meet the participation criteria (e.g., disclosed not being from the UK, N = 6). A further 13 participants were removed because they rated the auditory stimuli as unnatural (N = 11), or guessed the

study's aim (N = 2). These participants were removed to reduce the likelihood that any findings might be affected because participants noticed something unusual about the auditory stimuli. The final sample comprised 64 participants (males = 22).

Materials

120 black and white line drawings (Weldon & Roediger, 1987) were selected as visual stimuli. These images were grouped into 60 pairs, with no phonological overlap within pairs. Each pair included an image consistent with the item named by the speaker (the *referent*) and an alternative possibility (the *distractor*). 20 pairs were used as critical stimuli and the remaining 40 pairs were used for filler trials. All referent names had a log frequency higher than 3.5 in the SUBTLEX-UK database (Van Heuven et al., 2014). Pairs were matched for ease of naming ($H < 1$) and familiarity ($F > 3$; Griffin & Huitema, 1999) to avoid biases toward unfamiliar items which have been shown to arise in response to disfluent speech (Arnold et al., 2007).

A native British English speaker and a non-native English speaker (Spanish-English accent) recorded the auditory stimuli. Sentences were recorded in their entirety by both speakers. To create the critical stimuli, we followed Loy et al. (2017): A prolonged article followed by a filled pause from a disfluent utterance was cross-spliced into the fluent utterances, before the mention of the referent. This manipulation ensured that participants reacted to the same utterance across conditions, with only the substitution of "thee, uh" in the disfluent condition (Loy et al., 2017). To ensure that the auditory stimuli resembled natural speech, we presented the materials to two raters blind to the manipulation and who did not take part in the study. These raters heard the stimuli and were asked to report whether there was anything unnatural in the audio; if they reported anything odd, they were asked to elaborate. Neither of the raters reported noting the manipulation or considered the stimuli unnatural. The native and non-native speakers' filled pauses were each approximately 800 ms in duration.

Critical referents were counterbalanced across two lists, such that they were presented in both fluent and disfluent contexts across the experiment. Each list contained 40 additional filler utterances. These filler trials included 19 fluent and 21 'disfluent' recordings, with disfluent trials including several different indicators of uncertainty. Each speaker produced similar filler items. Table 6.1 below shows a breakdown of filler stimuli.

Table 6.1

Breakdown of sentence types for filler trials for the native and non-native speaker condition.

Filler type	**Description**	**No. of utterances**	**Example**
Fluent	None	19	The treasure is behind the
Disfluent	Filled pause	3	*Um*, the treasure is behind the
	Elongation	5	The treasure is behind *thee*
	Repetition	2	The treasure is behind *thee-thee*
Other	Discourse marker	5	*Ok*, the treasure is behind the
	Modal verbs	4	The treasure *might be* behind the
	Combination	2	*Now* the treasure *could be* behind the

Procedure

The stimuli were presented using OpenSesame v.3.2.7. (Mathôt et al., 2012) on a 21" CRT monitor screen, with the participant seated at a viewing distance of approximately 80 cm. Visual stimuli were presented vertically centred to the left and right sides of the screen, and audio was presented at a comfortable loudness level from two loudspeakers. Participants' eye movements were recorded using a Tower-mounted SR Research Eyelink

1000 eye-tracker, sampling the right eye at 500 Hz. Mouse responses were collected using a standard computer mouse.

At the beginning of the experimental session, the experimenter gave a cover story to explain the origin of the auditory stimuli and to introduce the element of deception. Participants were told that they would hear a previously recorded participant who had played against another participant and had been asked to lie half of the time. To engage participants and keep them motivated, they were told they would compete against the rest of the participants in the study and that they would get points for every time they guessed the location of the treasure correctly, with some "bonus rounds" earning them double the score. Participants were told they could enter their names into a high-score table if they won. They were shown this table at the beginning of the experiment. Participants were randomly assigned to either a group that heard the first-language British English speaker (n=32) or a group that heard the second-language English speaker (n=32).

Eye-tracking began with a nine-point calibration and validation procedure. Participants then performed a five-trial practice session. Trials began with a drift correction procedure, followed by a red fixation dot which remained on the screen for 500 ms. Images were then presented for 2000 ms, after which the audio began. The referent and distractor's positions on the screen were counterbalanced across trials. Participants could click on either object at any time once the speaker finished talking, but responses that took more than 5 additional seconds after the audio had stopped playing were discarded. Critical and filler trials were presented in random order. In order to motivate participants, 25% of filler trials were followed by a "bonus" feedback screen indicating that the participant had clicked on the item that concealed the treasure, regardless of where they had actually clicked. Figure 6.1 depicts the trial sequence.

After the experiment, participants filled in a questionnaire to assess whether they had noticed the manipulation. To further ensure that our auditory stimuli resembled natural speech and that participants' responses could not be attributed to noticing the manipulation of the auditory stimuli, participants rated the naturalness of the audio on a

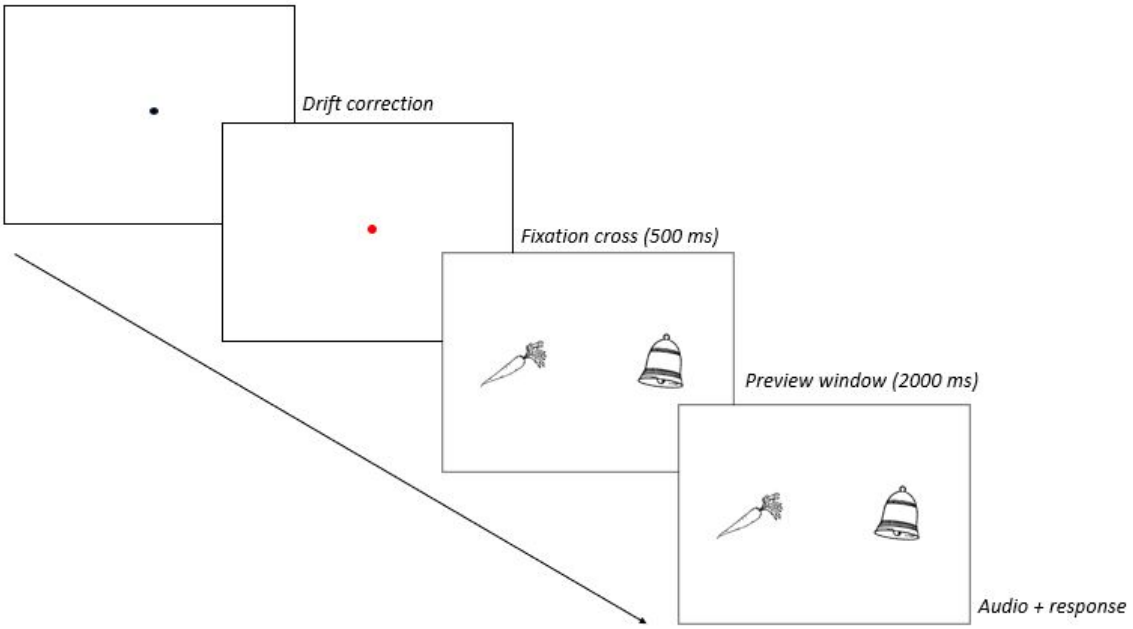

Figure 6.1. Trial sequence of Experiments 3 and 4.

9-point scale (1: not natural; 9: natural) and answered five open-ended questions regarding the experiment's stimuli and aim. Participants in the non-native speaker condition additionally rated on a 9-point scale the perceived fluency (1: very disfluent; 9: very fluent) and accentedness (1: not accented, 9: very accented) of the audio. These participants additionally reported their exposure to foreign-accented English as measured by their daily interactions with second-language English speakers (1: never, 9: always) and their daily exposure to foreign-accented English (1: never, 9: always). After filling in the questionnaire, participants were informed of the actual aim of the experiment and further asked if they had noticed the manipulation.

6.2.2 Results

Analyses were carried out in R version 4.2.0 (R Core Team, 2020), using the *lme4* package (version 1.1.29, Bates et al., 2015). Data wrangling was done with the *tidyverse* (version 1.3.1, Wickham et al., 2019) package, and data visualization was done with the *ggplot2* (version 3.3.6, Wickham, 2016) and *wesanderson* (version 0.3.6., Ram & Wickham, 2018) packages. Data from ten trials in which no click was recorded before cutoff (0.78% of all trials) was discarded prior to analysis. All models were fitted using the *bobyqa* optimiser.

Answers to questionnaire

Participants who listened to the native speaker rated the naturalness of the auditory stimuli an average of 7.56 (SD = 1.28), while the average rating for the non-native speaker was 7.18 (SD = 1.04). Naturalness ratings did not differ significantly between speaker conditions ($t(62) = 1.27$, $p = .21$). Additionally, participants who listened to the non-native speaker rated the audio's fluency on average 7.56 (SD = 1.15) and its accentedness as 7.06 (SD = 1.15). Participants reported their daily exposure to non-native accented speech an average of 6.11 (SD = 2.23), and interacting daily with non-native speakers an average of 5.83 (SD = 2.41). These measures were not found to affect participants' clicking behaviour and are not discussed further.

Object clicked

The objects clicked in each trial were analysed as indices of participants' interpretations of deceit. We modelled the referent disadvantage (i.e., participants' preference to click on the distractor, assumed to indicate a judgement of deceit; 1 = distractor, 0 = referent) using mixed-effects logistic regression. We included a within-participant and within-item effect of manner of delivery (fluent coded as -0.5, disfluent as +0.5), a between-participant but within-item effect of speaker's linguistic background (native coded as -0.5, non-native coded as +0.5) and their interaction as fixed effects. The maximal model justified by the design (Barr et al., 2013), with relevant by-participant and by-item random effects, failed to converge. The random effects structure was therefore simplified such that the final model included by-participant and by-item random intercepts and random slopes for manner of delivery by-participant and by-item.

The overall distribution of clicks in the experiment (i.e., combining critical and filler trials) show a tendency to believe both the native and the non-native speaker. There were 55.05% and 56.47% of trials with clicks on the referent for the native and the non-native speaker respectively, while 44.81% and 43.53% of trials had clicks on the distractor for the native and the non-native speaker respectively. Table 6.2 shows the proportion of

Table 6.2

Proportion of object clicks in critical trials split by conditions. (Raw counts in parentheses.)

		Native Speaker	Non-native Speaker
Fluent			
	Referent	0.79 (251)	0.70 (225)
	Distractor	0.21 (68)	0.30 (95)
Disfluent			
	Referent	0.32 (102)	0.29 (92)
	Distractor	0.68 (218)	0.71 (222)

clicks on each object per condition. The generalised mixed model showed a main effect of manner of delivery, whereby disfluent utterances led to more clicks on the distractor (β = 2.43, SE = 0.31, p <.001), in line with previous findings (Loy et al., 2017; see also King et al., 2018; Li et al., 2022). We found a marginally significant main effect of speaker's linguistic background (β = 0.34., SE = 0.18, p = .05), whereby utterances produced by the non-native speaker were more likely to be taken as deceptive. Crucially, the model showed no interaction between manner of delivery and speaker's linguistic background (β = -0.46, SE = 0.6, p = .44).[1]

Eye movements

The time course of fixations to referent and distractor by speaker's linguistic background is depicted in Figures 6.2 and 6.3 for the fluent and disfluent conditions respectively. For both speakers, listeners initially fixated on the referent. However, starting at around 500 ms after target onset, listeners began to fixate more on the distractor if the speaker was disfluent (Figure 6.3). The patterns of fixations for both native and non-native speakers appear to be similar to the patterns reported by Loy et al. (2017).

Eye-tracking data was averaged into bins of 20 ms (10 samples each) prior to analysis. Proportion of fixations towards either object were calculated per bin and then transformed into empirical logits (Barr, 2008) to measure the preference of fixations

[1]Similar results were obtained when including data from those who reported noticing the manipulation or rated the naturalness of the audio lower than 4: the model showed a main effect of manner of delivery (β = 2.29, SE = 0.27, p < 0.001), with disfluent utterances leading to more clicks on the distractor. Neither speaker's linguistic background (β = 0.3, SE = 0.17, p = .07), nor its interaction with manner of delivery (β = -0.46, SE = 0.54, p = .39), yielded significant effects.

towards the distractor over the referent: Positive numbers indexed a preference for the distractor, and negative numbers indexed a preference for the referent - i.e., the referent disadvantage. By doing so, we modelled the tendency to infer deceit over time depending on manner of delivery and speaker's linguistic background. The time window for analysis spanned from referent onset until 800 ms later, following previous research (Loy et al., 2017). This exceeds the duration of the longest critical referent name for both the native (616 ms) and the non-native (530 ms) speakers.

Following previous studies (e.g., King et al., 2018; Loy et al., 2017), we built a linear mixed model to explore participants' eye movements. Our model included within-participants and within-items effects of time (scaled), within-participants and within-items effects of manner of delivery (fluent coded as -0.5, disfluent as +0.5), and between-participants but within-items effects of speaker's linguistic background (native coded as -0.5, non-native coded as +0.5), and their interaction as fixed effects. We fit the maximal model justified by design (Barr et al., 2013) with by-participant and by-item random effects to match. Effects were considered significant at $|t| > 2$ (Baayen, 2008).

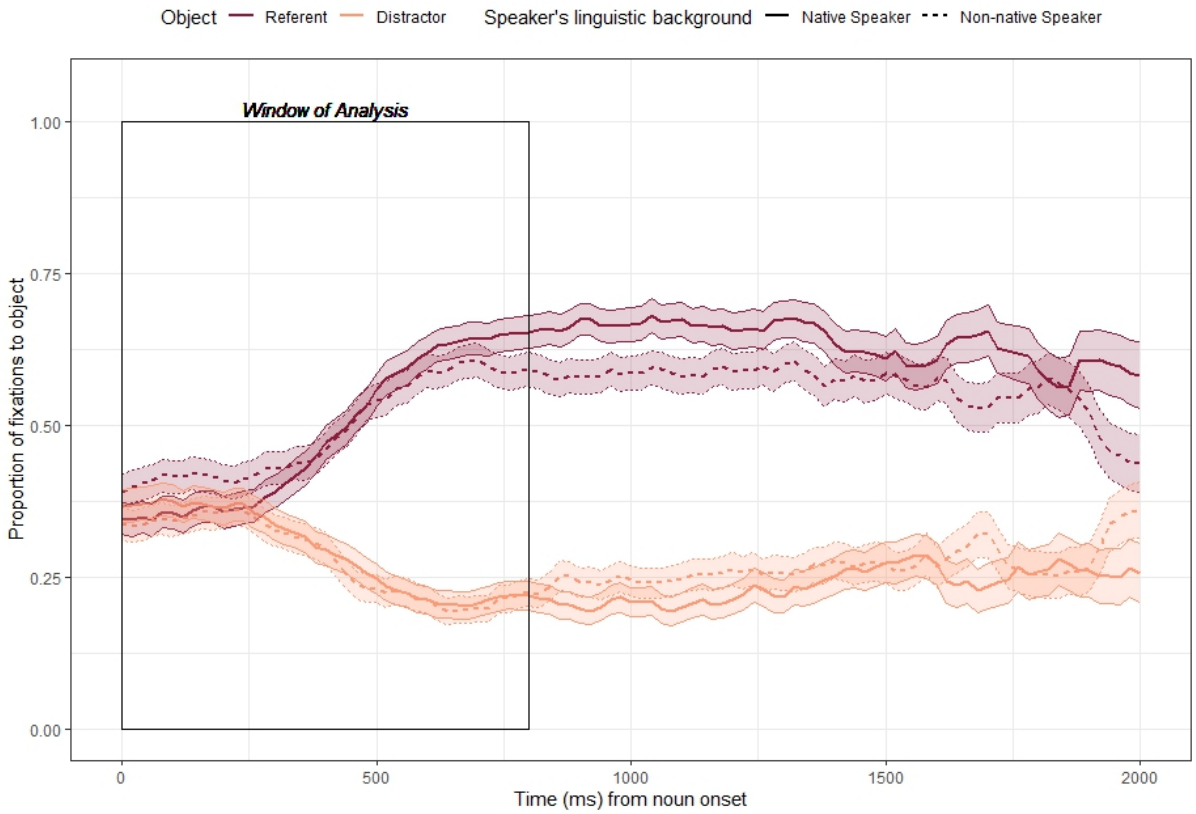

Figure 6.2. Fluent utterances: Mean proportion of fixations to referent and distractor by speaker's linguistic background (native/non-native). Proportions were calculated out of the sum of fixations towards either object for each 20-ms time bin from target onset to 2,000 ms post target-onset. Shaded areas represent ± 1 standard error of the mean.

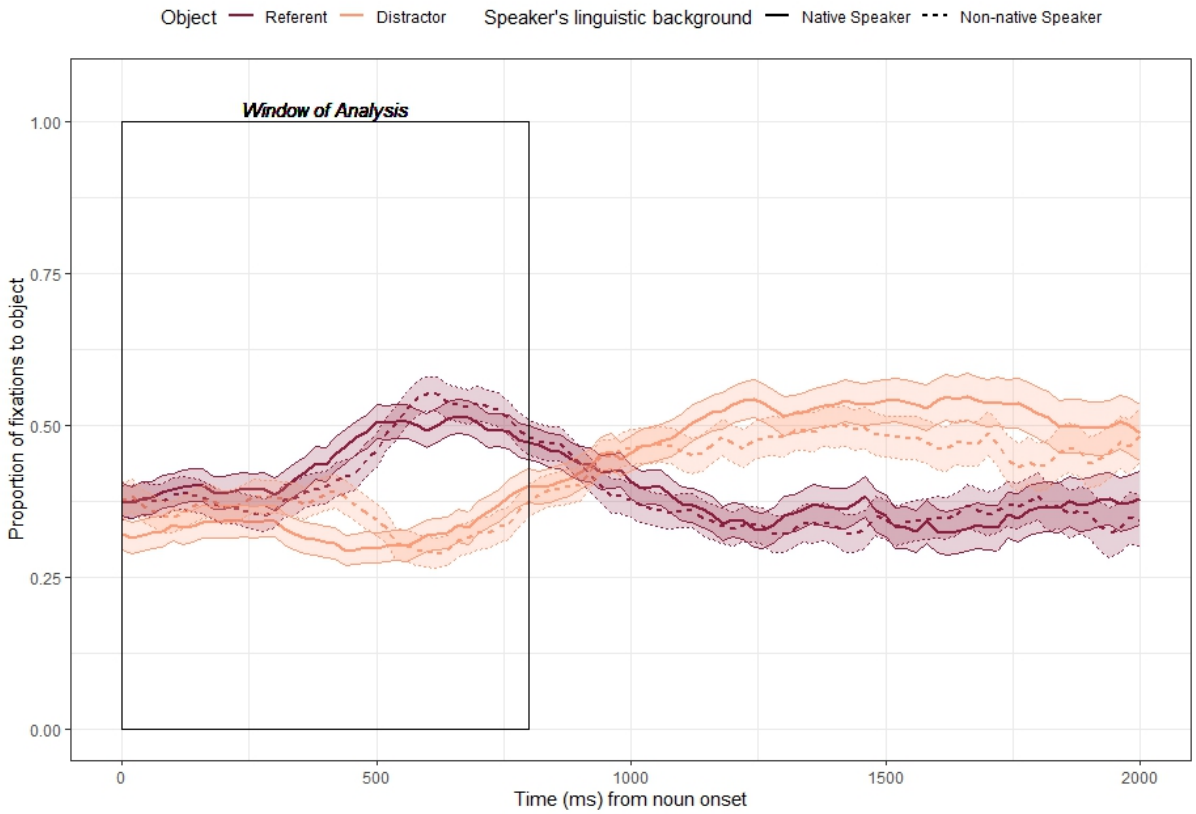

Figure 6.3. Disfluent utterances: Mean proportion of fixations to referent and distractor by speaker's linguistic background (native/non-native). Proportions were calculated out of the sum of fixations towards either object for each 20-ms time bin from target onset to 2,000 ms post target-onset. Shaded areas represent ± 1 standard error of the mean.

This analysis showed an interaction between manner of delivery and time, such that participants' fixations towards the distractor increased over time following disfluent utterances (β = 0.34, SE = 0.15, t = 2.30). The time course of fixations did not differ based on the speaker's linguistic background (β = -0.02, SE = 0.15, t = -0.10). Crucially, the time course of fixations towards either object did not differ based on the interaction between manner of delivery and speaker's linguistic background (β = -0.32, SE = 0.27, t = -1.20), suggesting that disfluent utterances increased the referent disadvantage in the same manner regardless of the speaker.[2]

[2]Similar results were found including data of participants who reported noticing the manipulation or rated the audio's naturalness lower than four. There was a main effect of manner of delivery in the time course of participants' eye movements (β = 0.35, SE = 0.13, t = 2.81), whereby following disfluent utterances participants fixated more on the distractor. The time course of fixations did not depend on

Post-hoc analysis

The results of the eye-tracking data suggest, at face value, that only the manner of delivery, and not the speaker's linguistic background, biased participants' interpretations towards deceit. The lack of interaction with the speaker's linguistic background, and in particular, the lack of delay in the time course of fixations, support an associative mechanism whereby disfluency is in a privileged position in the context of potential deceit.

The analyses conducted shed light on the overall effect of manner of delivery in the time course of the disfluency-as-deception bias. A post-hoc analysis was conducted to investigate whether participants learnt to discriminate the disfluency by comparing participants' fixations in the first and last third epochs of the experiment. The reason for this analysis is twofold. Firstly, investigating participants' behaviour throughout the course of the experiment allows us to test whether participants became sensitive to the manner of delivery during the experiment and solely relied on filled pauses due to the nature of the task. Secondly, the lack of effect of speaker's linguistic background on the emergence of the bias could be masked in an analysis that does not take presentation order into account. On the one hand, participants might have been cognitively taxed at the beginning of the experiment due to the non-native accent. If participants get used to non-native accents via exposure (Porretta, Tremblar et al., 2017; Porretta, Buchanan et al., 2020; Trude et al., 2013; Witteman et al., 2014), it could be that only towards the end of the experiment did participants have sufficient remaining resources to reason about the speaker. On the other hand, participants may have overridden stereotyped biases against non-native speakers throughout the course of the experiment, so that at the beginning of the experiment their behaviours were consistent with the inference account, but they were consistent with the associative account at the end of it.

We thus compared the pattern of fixations in the first and last thirds of the experiment. In this model, we compared participants' eye movements in the first and

speaker's linguistic background (β = 0.0005, SE = 0.13, t = 0.004) or its interaction with manner of delivery (β = -0.32, SE = 0.25, t = -1.29).

third epochs of the experiment. We modelled the referent disadvantage over an 800 ms time window post-target onset with fixed effects of time (scaled), speaker's linguistic background (native coded as -0.5, non-native coded as +0.5), epoch (first epoch coded as -0.5, third epoch as +0.5), fluency (fluent coded as -0.5, disfluent as +0.5) and their interactions. The maximal random structure included by-participant and by-item random intercepts, with random slopes of time, fluency, epoch, and their interaction by-participant, and time, speaker's linguistic background, fluency, and their interaction by-item. As depicted in Figure 6.4, the model showed that experimental epoch did not interact with the critical interaction between manner of delivery and speaker's linguistic background (β = -1.11, SE = 0.39, t = -0.28). The lack of differences between epochs suggests that disfluent utterances increased the referent disadvantage from the beginning of the experiment, which can be seen in Figure 6.4, bottom panel.

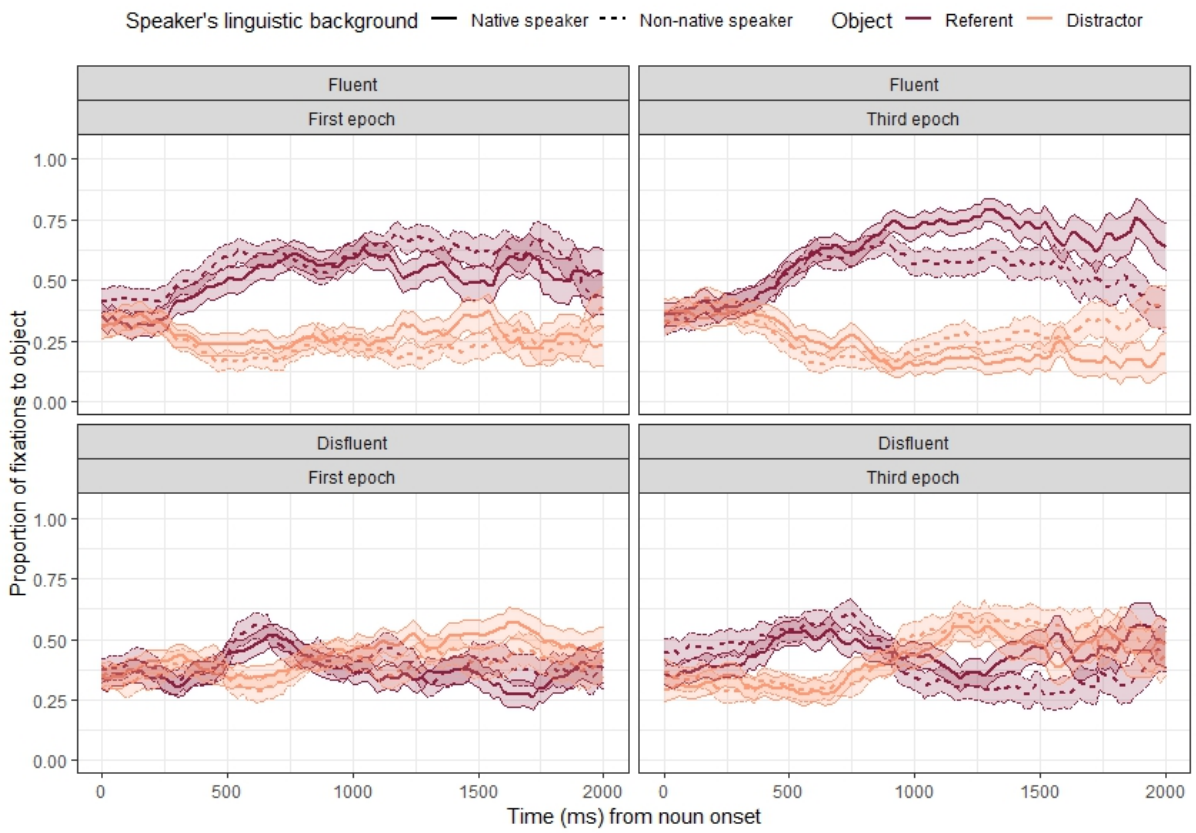

Figure 6.4. Mean proportion of fixations to referent and distractor for utterances produced by the native and the non-native speaker by the manner of delivery (fluent/disfluent) and experiment epoch (first/third). Proportions were calculated out of the sum of fixations towards either object for each 20-ms time bin from target onset to 2,000 ms post target-onset. Shaded areas represent ± 1 standard error of the mean.

6.2.3 Discussion

The findings are consistent with previous suggestions that listeners interpret disfluency as a signal of deceit. When speakers were disfluent, listeners were more likely to select an unmentioned distractor as the supposed location of the treasure. Contrary to previous findings (e.g., Lev-Ari & Keysar, 2010), we only found marginally significant evidence for a lie-bias towards the non-native speaker. These findings were mirrored in participants' eye movements: Fixations to the referent began to reduce soon after the referent onset following a disfluent utterance, replicating previous findings (e.g., Loy et al., 2017; see

also King et al., 2017; Li et al., 2022), in line with the predictions of an associative account.

The eye-movement analyses demonstrate that native and non-native disfluencies in the present study were interpreted in highly similar manners. Starting at around 500 ms after target onset, participants started to fixate more on the distractor following a disfluent utterance. This pattern mimics that reported in Loy et al. (2017), suggesting that the interpretation (that the treasure is unlikely to be in the mentioned location following a disfluency) occurs early during the comprehension process. Given the speed with which these fixation biases emerge, alongside the lack of differences as a function of speaker's linguistic background, it is possible that the effect may be derived by a stereotype listeners hold about the way deceitful speakers sound (i.e., following an associative mechanism), rather than reasoning about why a speaker produces speech in a specific way. The strength of this association may override the value of any other cue available.

One could argue that the imposition of a high-stakes context (i.e., where listeners are aware of speaker's deceitful intentions) may have rendered speaker identity as a useless cue, explaining the lack of effect for speaker identity. However, unless participants have set a 'strategy' prior to the experiment (i.e., decided what cues would be relevant for the task at hand), we should have found a difference in the time course of the disfluency-as-deception between those listening to the native and those listening to the non-native speaker - at least, at the beginning of the experiment. We did not find, however, any statistical differences in participants' patterns of fixations across the experiment, as shown by our post-hoc analysis. Taking these results together, Experiment 3 is best interpreted in terms of the associative account.

These findings are at odds with previous studies where the presence of competing cues for the speaker to produce a filled pause did delay the emergence of the disfluency-as-deception bias. King et al. (2018) used a similar paradigm to that of the experiment reported here: In their study, half of the disfluent items included a potential distraction to the speaker (a car horn) immediately preceding the disfluency. When this potential cause

for a disfluency was present, it took slightly longer for a fixation bias against the referent and for the distractor to emerge, although ultimately disfluent recordings both with and without car horns led to similar numbers of clicks on the distractor item. The authors took this delay in the disfluency-as-deception bias as reflective of the time it takes the system to draw inferences about the speaker's production system and to take into account the speaker's momentary distraction as an alternative explanation for their disfluency. While momentary distractors have not been shown to modulate predictive disfluencies (Arnold et al., 2007, Exp. 3), King et al. (2018) argue that their experimental set-up "requires participants to reason about the speaker and thus may encourage reasoning about the detail of the speaker's utterances" (King et al., 2018, p.133).

The contrast between Experiment 3 and King et al. (2018) suggests that not all disfluency is perceived as deceptive. In King et al. (2018), listeners were able to attribute the speaker's disfluency to a cause that was momentarily present, and extrinsic (i.e., a car horn that both speaker and listener heard). In the present study, any plausible alternative cause of disfluency was constantly present and intrinsic (i.e., disfluency could be put down to reduced second language competence). The differences in the salience of these alternatives could lead to them being weighted differently: Whilst the overall speaker identity might be downplayed, a momentary, time-independent distraction can update the weight given to speaker identity and thus make this cue worthwhile to consider when interpreting deceit, i.e., dynamically update comprehension following a speaker's characteristics.

This leaves open the question of what drives the disfluency-as-deception bias and why modulations are sometimes but not always possible: If the disfluency-as-deception bias is a powerful heuristic that overrides any other inference about the speaker's production system, then previously found modulations could simply be attributed to listeners' comprehension difficulties (a car horn is distracting to the listener too) and the absence of modulation confirms the strength of the disfluency-as-deception bias as an association

(non-native speakers' difficulties don't provide an alternative explanation). On the contrary, if the modulation of the disfluency-as-deception bias depends on the salience of an alternative explanation for the speaker's disfluency, then the absence of modulation may reflect the lack of a sufficiently salient alternative explanation. One way of disentangling this mechanism is by comparing the interpretation and time course of this bias across scenarios that vary in the comprehension difficulties they impose while simultaneously increasing the salience of the alternative cause for the speaker to be disfluent. In the next section, we argue that exploring the time course of the disfluency-as-deception bias in non-native listeners allows us to explore this question.

6.3 Experiment 4: The role of listener identity

In contexts in which disfluencies represent only one of many cues, such as when interpreting deception, the salience of the alternative cause for the speaker to be disfluent may need to be heightened for it to be taken into consideration by listeners. In Experiment 3, listeners would need to consider both the speaker's identity and the conversational context to model the production system of a non-native speaker may be less accessible for native listeners to compute.

In Experiment 4, we argue that non-native comprehenders have two particular properties that may answer whether modulations of the disfluency-as-deception bias are possible, and subsequently, what mechanism underlies it. Firstly, inferences are said to be delayed in this population, arguably due to reduced automaticity and cognitive resources (Ito & Pickering, 2021; see also Corps, Liao et al., 2022). Further, non-native listeners are said to be more aware of their interlocutor's language background (Rubio-Fernández, 2017): This suggests that the alternative reason for the speaker to be disfluent (i.e., non-nativeness) may be more salient for this population.

Regarding the first property, we presented evidence in Sections 3.2.2 and 5.2.2 suggesting that second-language comprehension may be costlier than first-language comprehension, which may affect meaning interpretation. We previously argued that an

inference account for the disfluency-as-deception bias entails a cost. While this cost may still give rise to the relatively quick emergence of the bias in native listeners, in scenarios where listeners have fewer cognitive resources, there can be a delay. Consequently, attending to disfluent native-accented speech in one's second language can give rise to a disfluency-as-deception bias, although arguably in a relatively delayed manner compared to native speakers.

We argued that in Experiment 3, listeners may have not been sensitive to non-native speakers' difficulties in speech production and thus took their disfluencies as evidence of their deceptive intent. Previous research has suggested that second-language comprehension is more sensitive to non-linguistic knowledge that is transferable across languages, such as world knowledge. Foucart et al. (2015a) found that non-native listeners displayed earlier responses to pragmatically inconsistent utterances (e.g., '*Every night I drink a little bit of* ***wine*** *before going to sleep*', said by a child) than their natives' counterparts. Their findings align with proposals suggesting that to compensate for the costs associated with second-language comprehension, non-native listeners rely more on knowledge they have mastered well (Futrell & Gibson, 2017), such as speaker identity. This idea may be particularly salient given that non-native comprehenders are said to be more socially aware of their interlocutor's features (Rubio-Fernández & Gluckberg, 2012) and show an advantage in theory of mind (Schroeder, 2018) compared to native, monolingual comprehenders - which, incidentally, were the participants from Experiment 3.

This speaker-oriented sensitivity may thus pose a challenge when attending to disfluent non-native-accented speech. This scenario entails two competing reasons for a disfluency to be present (i.e., speaker's intention to deceive, speaker's struggle with speech production due to their linguistic background), which weaken the contextual support of an interpretation of deceit (because a non-native speaker inhibits the disfluency-as-deception bias). Consequently, interpreting deceit when confronted with alternative

explanations when there are fewer resources available may be more difficult (Politzer-Ahles & Gwilliams, 2015; Degen & Tanenhaus, 2019). If indeed there is an inference underlying the disfluency-as-deception bias, given non-native listeners' heightened attention to speaker identity, and the cognitive resources available, interpreting disfluent speech produced by a non-native speaker may pose a more dramatic cue conflict, resulting in a delay in the emergence of the disfluency-as-deception bias.

Experiment 4 extends Experiment 3 to a sample of non-native English listeners. A heterogeneous sample of non-native listeners was tested as we had no specific hypothesis about a specific first language impacting the interpretation of deceit for filled pauses. As previously discussed, if the disfluency-as-deception bias is driven by an associative mechanism, then, as in Experiment 3, we should expect a similar time course in both speaker conditions. If there is some form of inference involved in the bias, then non-native participants should exhibit a delay in the emergence of the bias, which might be heightened when listening to the non-native speaker. Nonetheless, regardless of the speaker condition, we do expect that disfluent utterances will lead to more interpretations of deceit: Given the experimental paradigm, ultimately participants will guide their choices based on stereotypes about how deceptive speakers sound.

6.3.1 Methods

All materials and data, including experimental and analysis scripts, can be found at https://osf.io/zfjtw/.

Participants

Eighty-three self-reported non-native English speakers (with different L1s, residing in the United Kingdom) took part in the study for a final sample size of sixty-four. Participation criteria included self-reported age of acquisition, self-reported use of English during childhood (less than 20%), and self-reported ages of achieving competence in English (later than 6 years old). Five participants were not included in the analyses because they were native English speakers. Additionally, we removed those who found the auditory

stimuli unnatural (N = 10), guessed the study's aim (N = 1), or did not understand the instructions (N = 2). Data from one participant was not included in the analysis due to a technical error with the presentation computer. Participants were aged between 18 and 30, had no hearing impairments and had normal or corrected-to-normal vision. All participants gave their informed consent, as approved by the PPLS Ethics Committee (ref no. 294-1718/9). Participants were reimbursed either economically or with university credits and debriefed after the experiment. The final sample size was 64 (males = 15). Table 6.3 depicts collected measures of participants' English proficiency gathered via a language background questionnaire and English proficiency task done at the end of the experimental session. Further measures of participants' linguistic backgrounds can be found in Appendix B.

Materials

For the eye-tracking part of the experiment, the experimental materials were identical to those used in Experiment 3. Because participants in this experiment were non-native English listeners, we included measures of English proficiency and language background as in Experiment 2. We further included measures of participants' sociolinguistic abilities.

English proficiency was measured with a version of the English LexTale (Lemhöfer & Broersma, 2012) implemented in OpenSesame (Mathôt et al., 2012). To measure participants' language backgrounds, we used an adapted version of the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian et al., 2007), which included additional questions on language use (de Bruin, 2019), and was similar to those employed in previous studies on non-native speakers (Foucart et al., 2014; Ito et al., 2018). Participants reported all the languages they spoke in order of acquisition and self-rated their proficiency in speaking, writing, reading, and listening on a 10-point scale. For each of the languages participants listed, they reported the age of acquisition, the age when fluency was acquired, the age when they started using the language for communicative purposes (i.e. outside of a classroom), for how long they thought they had been learning the language, and the mode of acquisition (i.e. classroom, interaction with other people, media, or

Table 6.3

Participants' measured English proficiency mean (and standard deviation) as measured by the LexTale as well as the number of participants in each CPF level of proficiency following their LexTale score, their self-reported proficiency in English (1 = not good at all, 10 = native-like) mean (and standard deviations) by speaker condition, alongside participants' country of origin. Variables where data for four participants is missing are marked with an asterisk [*].

	Native speaker condition	Non-native speaker condition
LexTale score	Mean score = 81 (12.8) C1 - C2: 17 B2: 13 B1: 2	Mean score = 74.6 (14.5) C1 - C2: 15 B2: 10 B1: 7
Self-reported English proficiency in speaking [*]	8.25 (1.38)	7.5 (1.48)
Self-reported English proficiency in reading [*]	8.75 (1.08)	8.22 (1.18)
Self-reported English proficiency in listening [*]	8.54 (1.29)	8 (1.32)
Self-reported English proficiency in writing [*]	7.93 (1.46)	7.5 (1.46)
Length of stay in the UK	Range = 1 month - 9 years Mode = 1 month	Range = 1 month - 14 years Mode = 1 year
First language	Chinese (7), Spanish (5), Mandarin (4), German (3), Italian (2), Norgewian (2), Cantonese (1), Czech (1), Finnish (1), French (1), Greek (1), Iban (1), Romanian (1), Swiss German (1), Ukranian (1)	Chinese (10), Mandarin (4), Polish (4), Cantonese (3), Italian (3), Greek (2), Czech (1), Indionesian (1), Latvian (1), Romanian (1), Spanish (1), Vietnamese (1)

mixed). They additionally reported the percentage of exposure to each language in childhood, adolescence and currently. Further, we measured their current exposure to each language in different contexts (e.g. interaction with relatives, classmates, or media) on a 9-point scale. To evaluate participants' socio-linguistic abilities, we employed five questions from Bachman and Palmer (1989) wherein participants reported their abilities and difficulties in using English for communicative purposes. Demographic questions included country of origin, length of residence in the United Kingdom and in any other English-speaking country, and how commonly they interacted with British English speakers in daily life on a 9-point scale.

Procedure

The same procedure as in Experiment 3 was followed, with the following additions. Once participants finished the eye-tracking experiment, they completed the English LexTale (Lemhöfer & Broersma, 2012) implemented in OpenSesame (Mathôt et al., 2012). Participants then answered the same questionnaire as those in Experiment 3 to assess whether they had noticed the manipulation, and to give their evaluations of the auditory stimuli.

Finally, participants filled in the language background questionnaire. 44 participants completed the questionnaire with pen-and-paper in the lab. Due to COVID-19, testing was resumed 2 years after these 44 participants were tested and the experiment was adapted to meet the new lab protocol. This meant moving the language background questionnaire online (implemented in Qualtrics) to minimise participants' time in the eye-tracking lab. Thirty-five participants completed the questionnaire online before the eye-tracking session[3]. Upon completion of their time in the lab, participants were debriefed and further asked if they had noticed the manipulation.

[3]Four participants did not complete the online questionnaire. Throughout the rest of this chapter, measures that do not include their data are marked with an asterisk.

6.3.2 Results

Analyses of object clicked and eye movements were conducted in the same way as those for Experiment 3. Data from trials in which no click was recorded before cutoff (6 trials, 0.47% of data) was discarded prior to analysis.

Answers to questionnaire

Naturalness of auditory stimuli for the native speaker was rated an average of 8.22 (SD = 1.04), and for the non-native speaker an average of 7.56 (SD = 1.37), which differed significantly ($t(62) = 2.16$, $p = .03$). The non-native speaker's audio fluency was rated an average of 7.22 (SD = 1.36) and the accentedness as 6.69 (SD = 2.05). Those who listened to the non-native speaker rated their daily exposure to non-native accented speech as 6.59 (SD = 2.09) and their daily interactions with non-native English speakers as 6.50 (SD = 2.16). Participants who listened to the native speaker reported interacting daily with British English speakers as 7.04 (SD = 1.70). None of these measures impacted participants' click distributions, and they are not discussed further.

Object clicked

We modelled participants' interpretations of deceit by looking at the object clicked in each trial. The model included fixed effects of fluency (fluent coded as -0.5, disfluent coded as +0.5), speaker's linguistic background (native coded as -0.5, non-native coded as +0.5), and their interaction as fixed effects. The maximal model (Barr et al., 2008) failed to converge. The final model included random intercepts by-participant and by-item, and random slopes for manner of delivery by-participant.

Across the experiment, participants displayed a truth-bias for both the native and the non-native speaker. 53.14% and 58.21% of trials recorded a click on the referent for the native and the non-native speaker conditions, respectively. In contrast, 46.86% and 41.79% recorded a click on the distractor. Table 6.4 shows the proportion of participants' clicks on each object per condition. Our model showed a main effect of manner of delivery,

Table 6.4

Proportion of object clicks in critical trials split by conditions. (Raw counts in parentheses.)

		Native Speaker	**Non-native Speaker**
Fluent			
	Referent	0.70 (224)	0.72 (230)
	Distractor	0.30 (95)	0.28 (89)
Disfluent			
	Referent	0.30 (94)	0.39 (125)
	Distractor	0.70 (223)	0.61 (194)

whereby disfluent utterances led to more clicks on the distractor as the location of the treasure (β = 1.88, SE = 0.26, $p < .001$). We did not find a main effect of speaker's linguistic background (β = -0.30, SE = 0.18, $p = .09$), nor an interaction between manner of delivery and speaker's linguistic background (β = -0.40, SE = 0.49, $p = .41$).[4]

Eye movements

Figures 6.5 and 6.6 depict the time course of eye movements to the referent and the distractor by speaker's linguistic background for the fluent and disfluent condition respectively. The time course of fluent utterances (Fig. 6.5) produced by native and non-native speakers is characterised by an increase in fixations towards the referent that is sustained over time. For disfluent utterances (Fig. 6.6), participants listening to both the native and non-native speaker start off with a fixation bias towards the referent. However, at around 600 ms, participants start to fixate more on the distractor. A similar pattern is seen in both speaker conditions.

We modelled eye movements via empirical logits measuring the referent disadvantage (where positive values indicate a preference to fixate on the distractor, and negative values, a preference to fixate on the referent). Our model included the same predictors as in Experiment 3 in a time window of 800 ms post-target onset. We included manner of delivery (fluent coded as -0.5, disfluent as +0.5), speaker's linguistic background (native

[4]The same pattern of results was found in a model including participants who guessed the aim or rated the naturalness of the audio as low. Participants clicked more on the distractor following a disfluent utterance (β = 1.92, SE = 0.22, $p < .001$). Object clicked did not depend on speaker's linguistic background (β = -0.30, SE = 0.16, $p = .06$) nor its interaction with manner of delivery (β = 0.008, SE = 0.44, $p = .99$).

coded as -0.5, non-native coded as +0.5), time (scaled) and their interactions as fixed effects. The model included random intercepts by-participant and by-item, and random slopes for time, manner of delivery and their interaction by-participant, and for time, manner of delivery, speaker's linguistic background and their interaction by-item. We considered effects significant at $|t| > 2$ (Baayen, 2008).

The analysis showed an interaction between manner of delivery and time: Participants' eye movements towards the distractor increased over time following disfluent utterances (β = 0.32, SE = 0.12, t = 2.62). Eye movements were not biased depending on the speaker's linguistic background over time (β = -0.05, SE = 0.12, t = -0.44). Importantly, the results showed no interaction between manner of delivery and speaker's linguistic background over time (β = -0.34, SE = 0.31, t = -1.14): The bias induced by disfluent utterances towards the distractor did not differ depending on the speaker. For a visualization of the referent disadvantage in empirical logits, please see Appendix B[5].

[5]The same pattern of results held true when participants who were excluded from analysis because they noticed the manipulation were included. The model showed an interaction of manner of delivery and time, whereby disfluent utterances led to more fixations on the distractor over time (β = 0.37, SE = 0.11, t = 3.47). There was no interaction of time and speaker's linguistic background nativeness (β = -0.06, SE = 0.11, t = -0.56), nor an interaction of time, manner of delivery and speaker's linguistic background (β = -0.32, SE = 0.24, t = -1.35).

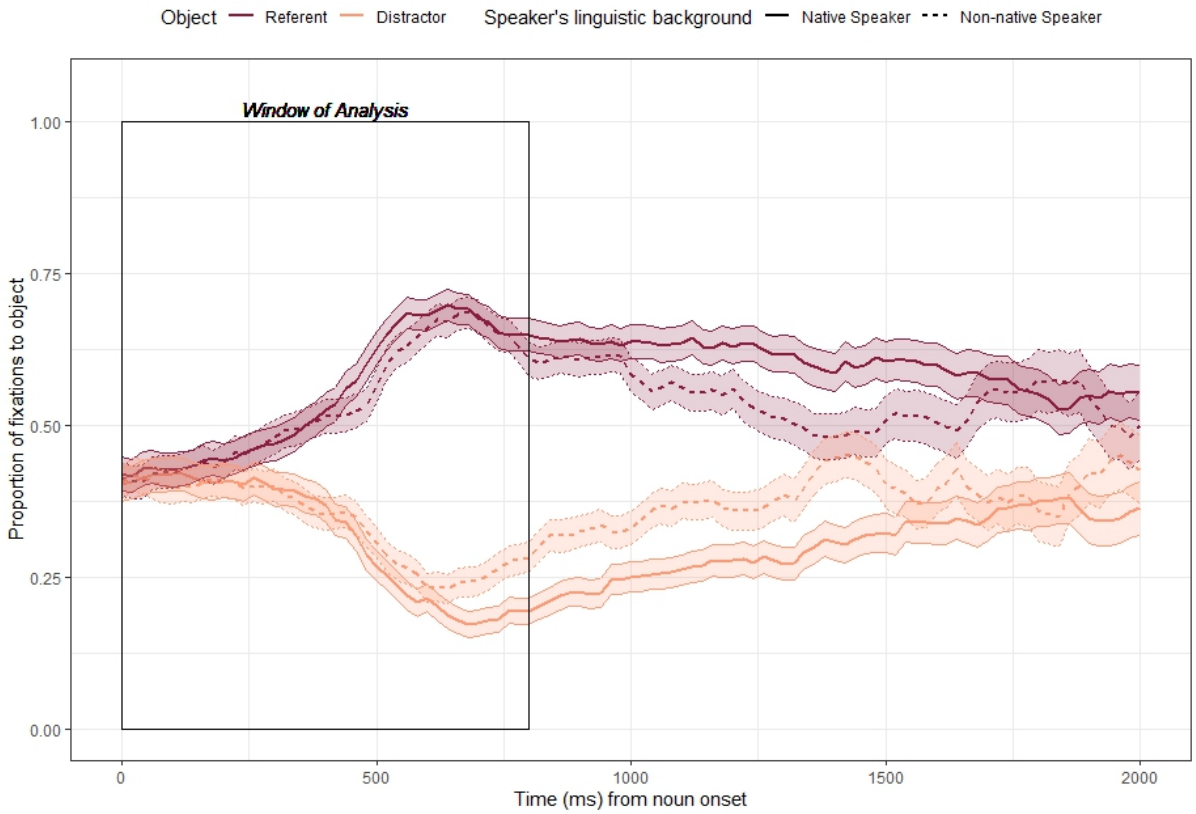

Figure 6.5. Fluent utterances: Mean proportion of fixations to referent and distractor by speaker's linguistic background (native/non-native). Proportions were calculated out of the sum of fixations towards either object for each 20-ms time bin from target onset to 2,000 ms post target-onset. Shaded areas represent ± 1 standard error of the mean.

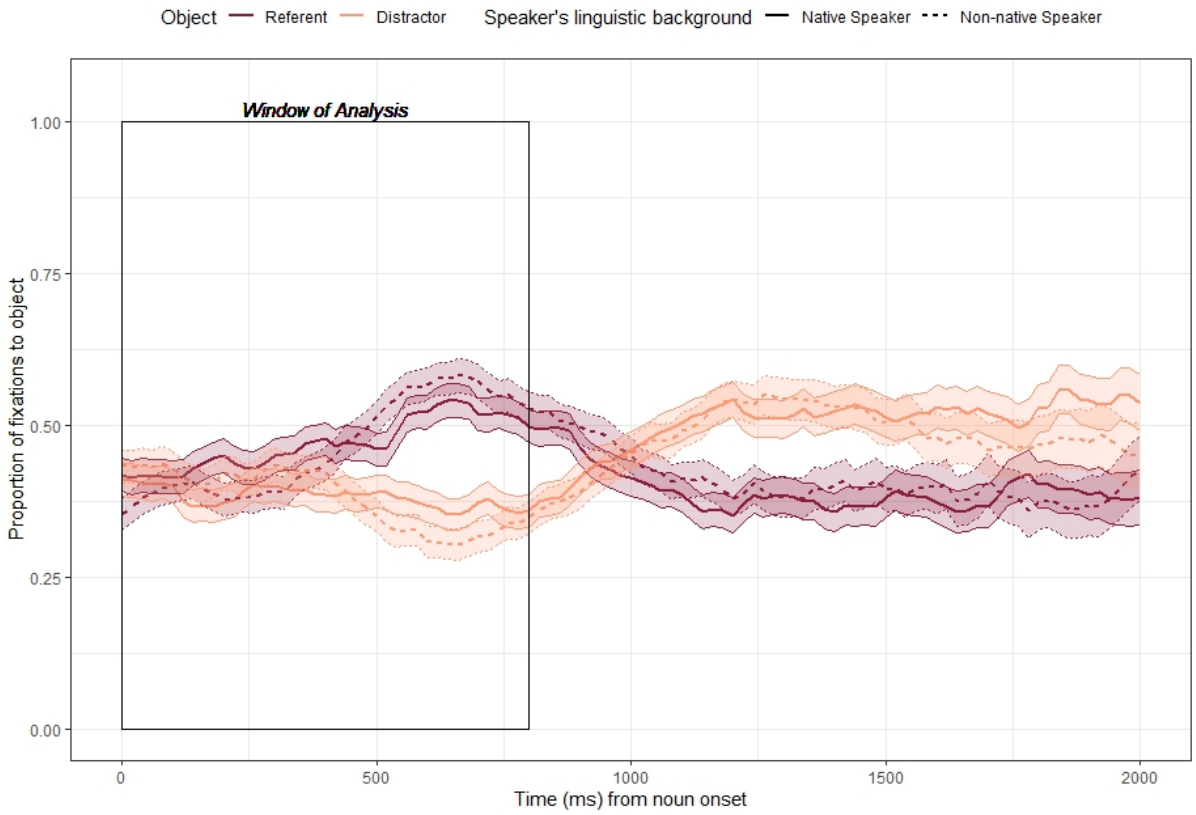

Figure 6.6. Disfluent utterances: Mean proportion of fixations to referent and distractor by speaker's linguistic background (native/non-native). Proportions were calculated out of the sum of fixations towards either object for each 20-ms time bin from target onset to 2,000 ms post target-onset. Shaded areas represent ± 1 standard error of the mean.

Post-hoc analysis

Similar to Experiment 3, a post-hoc analysis was run to investigate whether the disfluency-as-deception bias differed across the experiment. We modelled the referent disadvantage in an 800 ms time window post-target onset with fixed effects of time (scaled), speaker's linguistic background (native coded as -0.5, non-native coded as +0.5), epoch (first epoch coded as -0.5, third epoch as +0.5), fluency (fluent coded as -0.5, disfluent as +0.5) and their interactions, and by-participant and by-item random intercepts, with random slopes for time, fluency, epoch, and their interactions by-participant, and random slopes for time, speaker's linguistic background, fluency, and their interactions by-item.

The model showed that manner of delivery was the main factor in driving the disfluency-as-deception bias time course (β = 0.38, SE = 0.17, t = 2.2). Importantly, participants' time course of fixations did not differ between the first and the third part of the experiment as a function of fluency (β = 0.40, SE = 0.26, t = 1.51), nativeness (β = -0.41, SE = 0.28, t = -1.45) and their interaction (β = 0.69, SE = 0.53, t = 1.31). As in Experiment 3, the disfluency-as-deception bias emerged from the very beginning of the experiment and continued throughout, as depicted in the bottom panel of Figure 6.7.

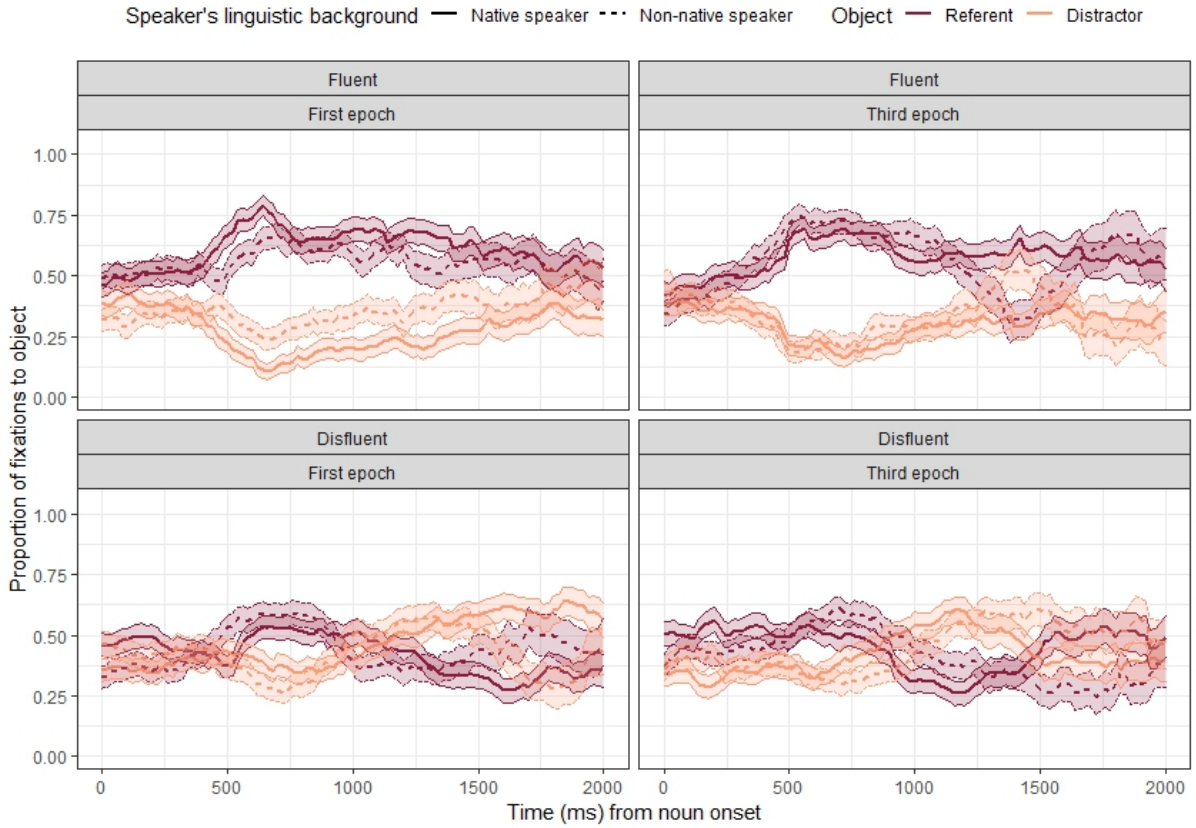

Figure 6.7. Mean proportion of fixations to referent and distractor for utterances produced by the native and the non-native speaker by manner of delivery (fluent/disfluent) and experiment epoch (first/third). Proportions were calculated out of the sum of fixations towards either object for each 20-ms time bin from target onset to 2,000 ms post target-onset. Shaded areas represent ± 1 standard error of the mean.

6.3.3 Discussion

The results of Experiment 4 replicated those of Experiment 3. Disfluencies produced by both the native and the non-native speaker were taken as indices of deceit. Importantly, the emergence of the disfluency-as-deception bias did not differ as a function of speaker's linguistic background: Visual inspection suggests that around 600 ms post-referent onset, participants started to fixate on the distractor upon encountering a disfluent utterance. Participants' pattern of fixations did not change throughout the course of the experiment as suggested by our post-hoc analysis: Disfluency triggered interpretations of deceit from the beginning of the experiment.

We initially proposed that the lack of modulation of the disfluency-as-deception bias due to speaker's non-nativeness in Experiment 3 may have been due to the lack of salience of this alternative explanation for those participants. This argument was based on King et al.'s (2018) findings where the presence of an alternative reason for the speaker to be disfluent, in the form of an environmental sound (i.e., a car horn), did delay the emergence of the bias - which the authors took as evidence for an inference mechanism whereby listeners reason about why a speaker is disfluent. We argued that to test whether the difference between the results of Experiment 3 and those of King et al.'s (2018) can be put down to listeners' different sensitivities to what may pose problems in speech production, we should explore the emergence of the disfluency-as-deception bias in a sample of non-native listeners. Our argument was twofold. Firstly, second-language comprehension is characterised by increased cognitive demands and decreased automaticity (Cunnings, 2017; Ito & Pickering, 2021). This entails that, if an inference mechanism underlies the disfluency-as-deception bias, we should see a delay in its emergence even when attending to native-accented speech. Secondly, non-native listeners are said to be more tuned to their interlocutors' linguistic background (Rubio-Fernández, 2017) and better at mentalising (Schroeder, 2018), which is an operation expected to underlie meaning interpretation

(cf. Woensdregt & Smith, 2017). Consequently, they should be more aware of the difficulties associated with second-language production and thus more likely to consider the speaker's non-nativeness to interpret a filled pause.

The emergence of the disfluency bias was relatively quick post-target onset. Although visual inspection of the time course of the bias initially suggests a delay if compared to that in native listeners (Fig. 6.3 shows that participants' fixations on the distractor started to increase at around 400 ms post-referent onset), this could be attributed to reduced word recognition in second-language comprehension (e.g., due to increased lexical competition, Lagrou et al., 2013). The fact that we found no differences in the time course of the disfluency-as-deception bias as a function of speaker's linguistic background provides further evidence for an associative mechanism. This suggests that interpreting disfluency as deception is insensitive to who produces the disfluency, and thus likely to reflect a lack (or reduced) social reasoning to derive the interpretation.

There are, however, two counterarguments to our proposal. The most contestable assumption is that our non-native participants did not experience cognitive load, as we expected. While working memory demands have been as a potential explanation for second language comprehension difficulties, and have been put forward to account for the differences with first language comprehension (cf. Cunnings, 2017; Kaan, 2014), it is possible that, at least in the case of our participants, they did not experience more cognitive demands than native listeners do. The participants in the present study were immersed in an English-speaking country, a factor shown to ease the cognitive burden of comprehending a second language (Mazzaggio et al., 2021). We assumed also that second-language comprehension of speech produced with a non-native accent would be more taxing than comprehension of native-accented speech. It is important to note that both accents are foreign to our participants; it is just that the native-accented speech resembles more the 'canonical' variety. The fact that there are no differences over the course of the experiment for either speaker would suggest that participants did not adapt

to the specific speaker's accent, suggesting that both accents were equally easy (or difficult) to comprehend for our participants. It is important to note that our participants reported being exposed to native and non-native English speakers and accents rather often in their daily lives. This life experience is brought to the experiment: Previous research has shown that second language comprehension of speech produced with a native and a non-native accent is impacted by listeners' familiarity with the accent (Grey et al., 2019). Consequently, we could interpret the lack of delay as a result of participants' familiarity with non-native-accented English. Nonetheless, this first counterargument does not negate our interpretation of the findings under an association account: It could still be possible that disfluency was the only cue considered to interpret deceit in Experiments 3 and 4.

The second argument involves the assumption that non-native listeners would be more speaker-oriented, and thus more initially 'forgiving' of our non-native speaker's disfluencies. Attention to social cues does play a role in the emergence of the disfluency-as-deception bias: Individuals who score higher on the Autism Spectrum Quotient, and thus display a higher amount of traits associated with autism, display a delay in the emergence of the disfluency-as-deception bias (Li et al., 2022). However,we may have overestimated the implications of theory of mind in second-language comprehenders. Likewise, it is possible that our non-native listeners were indeed more speaker-oriented, but this did not translate onto a heightening of alternative explanations for the presence of a disfluency. As we did not measure our participants' social skills, this is an open possibility.

Finally, it is possible that the task's instructions downplay the necessity of engaging in theory of mind. Specifically, participants were informed that they had to make a relatively fast decision. In turn, this may have favoured relying on any particularly salient cue available to make their interpretation. The fact that participants' pattern of fixations did not change throughout the experiment runs partially against this argument: Unless our participants had decided from the very beginning of the experiment what cues

would be useful to perform the task, then we should have seen different time courses in our post-hoc analysis, which was not the case.

This discussion highlights the need of comparing the pattern of fixations for native and non-native participants. This test would establish whether there were any delays in the emergence of the disfluency-as-deception bias as a function of the language listeners were comprehending (i.e., their first or their second one). As our reasoning lies on the assumption that there should be differences between these two populations that can help disentangle the inference from the associative account, in the next section we discuss the results of an analysis between these two populations. As we will show, a more detailed exploration of participants' pattern of fixations revealed that the emergence of the disfluency-as-deception bias did not differ between native and non-native listeners.

6.4 Comparison across populations

Participants' eye movements from Experiment 3 and Experiment 4 were combined to explore any potential differences between native and non-native listeners; specifically, we wanted to establish if there was any delay in the emergence of the disfluency-as-deception bias in non-native participants are opposed to native ones.

The model was akin to that employed in each experiment, with the difference that participants' linguistic background was added as a fixed effect (native coded as -0.5, non-native as +0.5), alongside its interaction with time, speaker's linguistic background, and manner of delivery. The model additionally included a random slope by-items for participants' linguistic background. The analysis showed that there were no differences between native and non-native listeners over time for manner of delivery (β = -0.02, SE = 0.14, t = -0.12) or speaker's linguistic background (β = -0.07, SE = 0.15, t = -0.46). Importantly, there was no interaction between the manner of delivery, speaker's linguistic background and participants' linguistic background over time (β = -0.01, SE = 0.28, t = -0.04).

A model paralleling that of the post-hoc analysis in Experiments 3 and 4 further corroborated that these two populations did not differ across the experiment in terms of the time course of fixations as a function of speaker's linguistic background and manner of delivery ($\beta = 0.40$, SE = 0.64, t = 0.62).

6.5 Non-linear time course of eye movements

The results of the comparison between the two populations suggest that the emergence of the disfluency-as-deception bias might be more nuanced than what the initial analyses showed. As argued by Stone et al. (2021) and Ito and Knoeferle (2022), eye-tracking experiments whose hypotheses rely heavily on timing differences might benefit from non-linear modelling (e.g., Generalised Additive Mixed Models, GAMM, Wood, 2003, 2004, 2011, 2016, 2017). Therefore, we conducted an exploratory analysis for Experiments 3 and 4 using GAMMs. An explanation of GAMMs in terms of how they are built is reported in Appendix B, as well as the specification of the models built and their results. As these models' results do not change our conclusions, we refer the reader to this Appendix for a detailed description of their results. One of the drawbacks of GAMMs is that they rely on visual inspection of the model's results; hereafter, we will solely discuss the emergence of the disfluency-as-deception bias via the models' visualisation (which were done with the *itsadug* (version 2.4.; van Rij et al., 2020) package).

Figure 6.8 shows the smooths for fluent and disfluent utterances by speaker's linguistic background (on the left), and the difference between these smooths by speaker's linguistic background (on the right) for Experiment 3. In both figures, we see that following disfluent utterances, there is an increase in the fixations towards the distractor compared to fluent utterances. The model showed that the disfluency-as-deception bias emerged at around 315 ms post-target onset following a disfluency produced by a native speaker. The referent disadvantage following a disfluency emerged at around 210 ms for those listening to the non-native speaker.

Figure 6.9 shows non-native participants' smooth function of eye movements for fluent and disfluent utterances by speaker's linguistic background (on the left), and the difference smooth by speaker's linguistic background (on the right). Mapping onto our results, disfluent utterances led to an increase in fixations towards the distractor compared to fluent utterances. Both for the native (A) and the non-native (B) speaker conditions, we find that fluent utterances led to more fixations to the referent compared to disfluent ones. Specifically, this difference in fixations emerges at around 290 ms post-target onset for both speaker conditions. This difference disappears for those who listened to the non-native speaker at around 678 ms.

Figure 6.10 depicts whether the difference between fluent and disfluent utterances in the native and non-native speaker conditions are statistically different. This Figure is equivalent to contrasting the right plots of Figures 6.8 and 6.9. The model showed that there were no differences by speaker's linguistic background in the difference between fluent and disfluent utterances: The disfluency-as-deception bias emerged in a similar fashion in both speaker conditions, and regardless of the listener's linguistic backgrounds.

Taken together, these analyses suggest a relatively early emergence of the interpretation of disfluent speech as deceptive. Importantly, it corroborates the fact that there were no differences in the time course of this interpretation as a function of speaker's linguistic background.

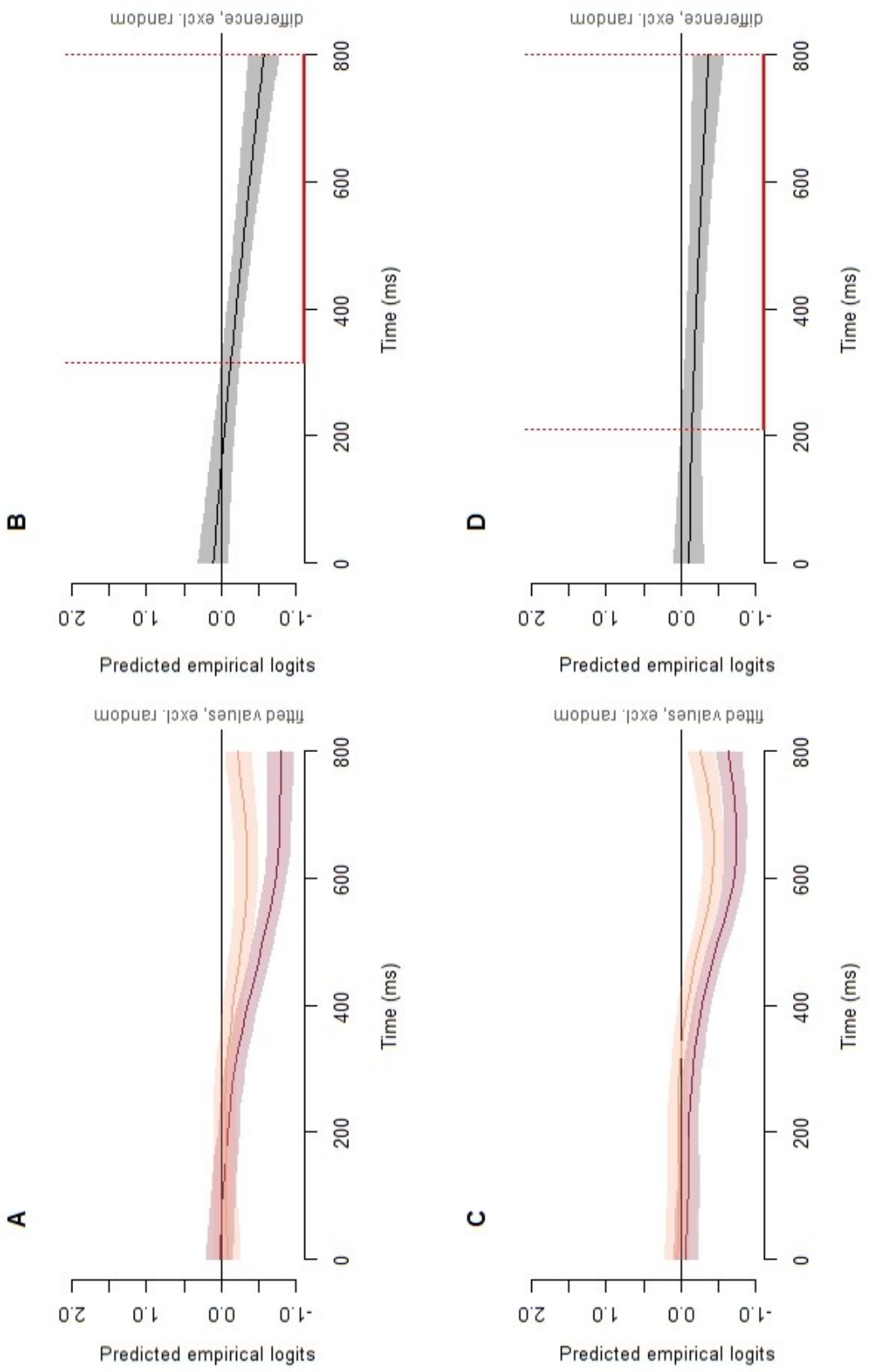

Figure 6.8. Predicted empirical logits for Experiment 3. Figure A and C depict the model predictions for the native and non-native speaker respectively, for fluent (red) and disfluent (orange) utterances. Figures B and D depict the difference between the fluent and disfluent conditions. Area of the x-axis highlighted in red shows when there are significant differences.

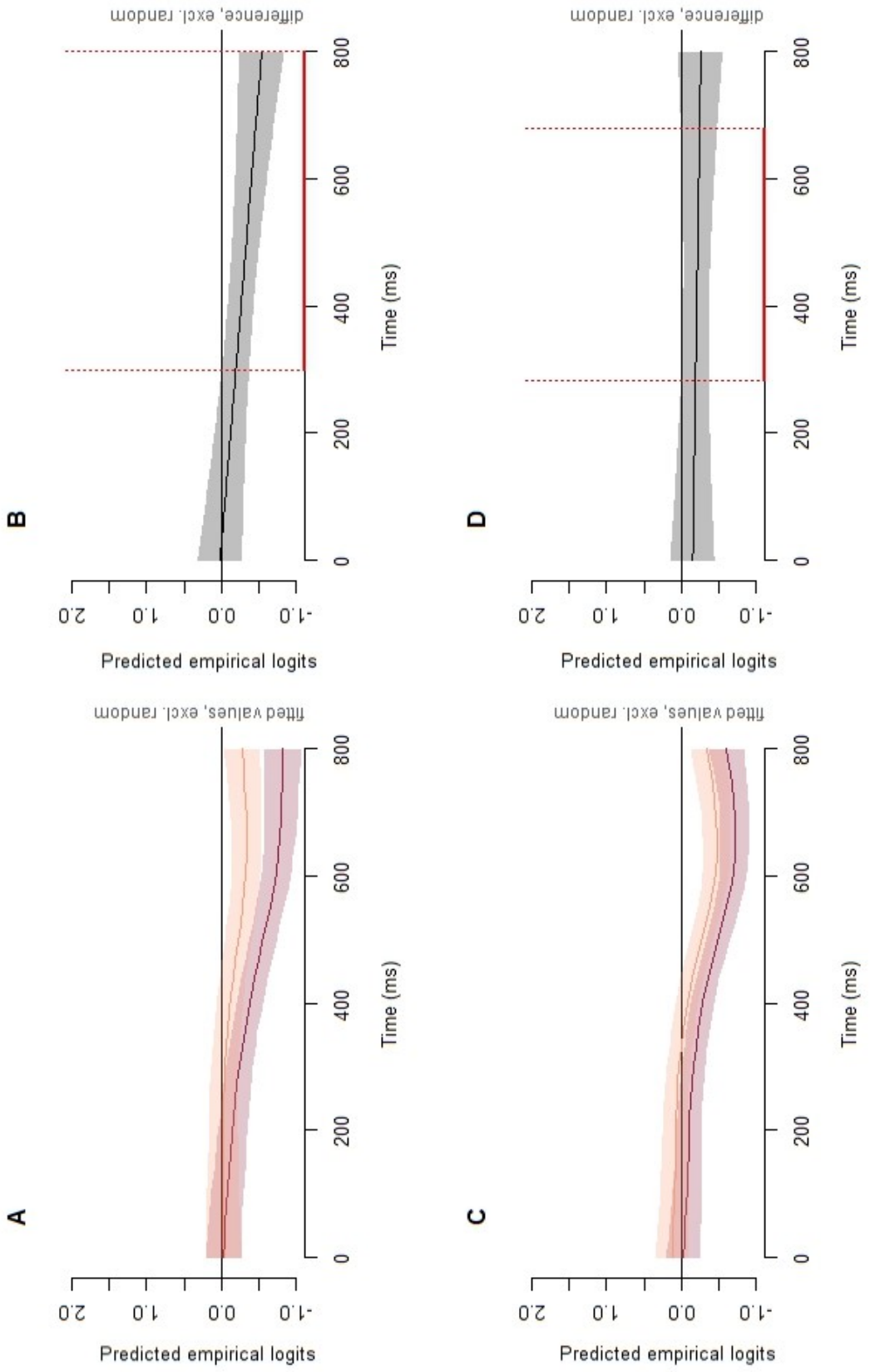

Figure 6.9. Predicted empirical logits for Experiment 4. A and C depict the model predictions for the native and non-native speaker respectively, for fluent (red) and disfluent (orange) utterances. B and D depict the difference between the fluent and disfluent conditions. Area of the x-axis highlighted in red shows when there are significant differences.

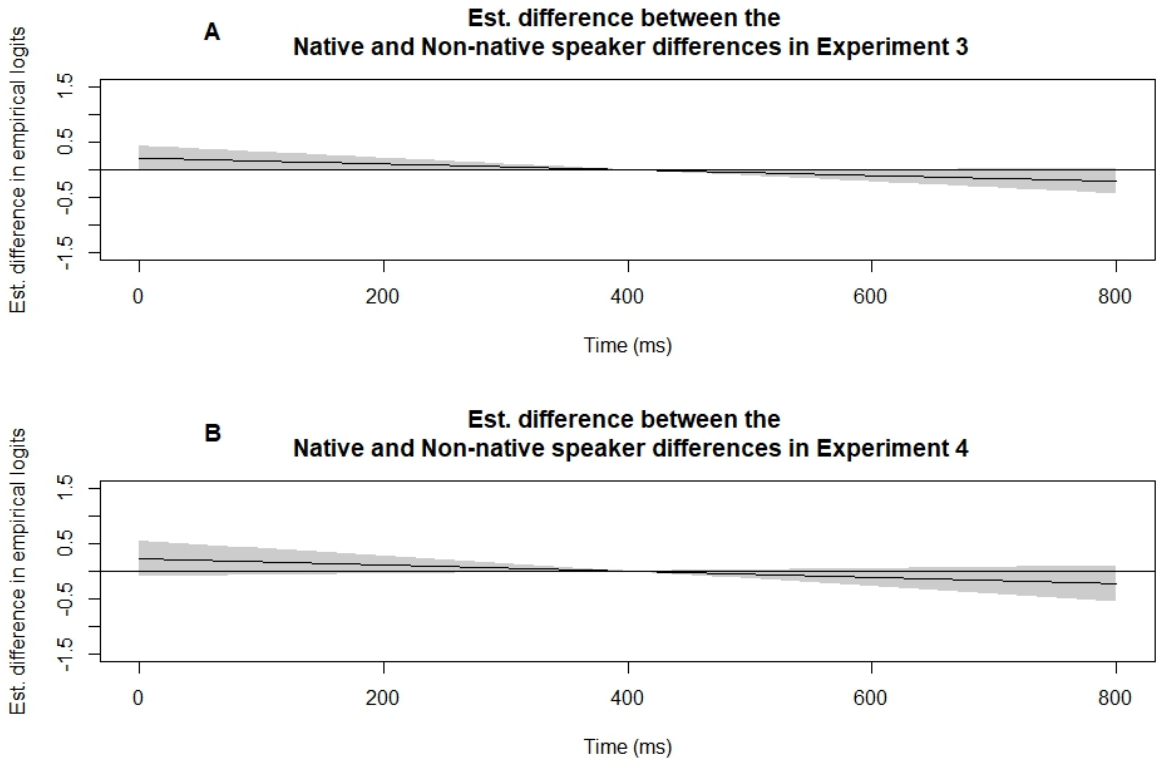

Figure 6.10. Predicted differences in the referent disadvantage for the difference between fluent and disfluent between speaker conditions. A depicts this difference for Experiment 3, while B depicts it for Experiment 4

6.6 General Discussion

Deception is a pervasive phenomenon in everyday communication (Serota & Levine, 2015; Serota et al., 2010). Although individuals are no better than chance at detecting it (Bond & DePaulo, 2006), when put in a situation where deceit is expected, their judgements seem to be biased by a range of cues that range from the situational (such as who is speaking) to the ephemeral (a verbal hesitation). In the two experiments presented here, we contrasted these two types of cues in a paradigm designed to shed light on the processes by which listeners decide whether or not to trust speakers with respect to a specific utterance depending on how it is delivered. Specifically, we were interested in whether this disfluency-as-deception bias is the outcome of listeners' reliance on a priori theories or if it reflects listeners' reasoning about how speech is produced in a given

context. Note that the focus here was on the ways in which these decisions are made, rather than on whether they are accurate or not: What are the cues that listeners take into account when they assess the truthfulness of what they are being told?

We proposed that the disfluency-as-deception bias can be accounted for in terms of an associative or an inference mechanism. These two mechanisms make different predictions for the emergence of the bias; specifically, we were interested in the malleability of the bias and its resistance to cognitive load. An associative mechanism predicts that all disfluencies will be taken as indices of deceit, regardless of the context in which they are produced, and this interpretation is effortless. In sharp opposition, an inference mechanism predicts that some, but not all, disfluencies will be taken as evidence for deception: Listeners consider the relevant cues available in a context to derive an interpretation, so that, in the light of competing cues that explain the presence of a disfluency, the emergence of the bias might be slower and costlier. Consequently, an inference account is compatible with the expectation that the disfluency-as-deception bias emergence might be impacted by the cognitive resources available.

Across two experiments, we found evidence supporting an associative mechanism. Specifically, we aimed to disentangle these two accounts by extending Loy et al.'s (2017) study to include an alternative explanation for the presence of a disfluency: Listeners comprehended speech produced by either a native or a non-native speaker; in the latter case, the presence of a disfluency can be plausibly explained by word-finding difficulties associated with second language production. We extended this paradigm to a sample of non-native listeners. We posited that non-native listeners are in a cognitively demanding situation, which increases when attending to non-native-accented speech. These two experiments demonstrated that manner of delivery was the sole player in the emergence of the bias, with both groups showing a lack of sensitivity to the speaker's linguistic background. Importantly, the time course of this interpretation was similar across both sets of participants: By 300 ms post-noun onset, the effects of disfluency were visible in all participants.

In what follows, we discuss what this means in terms of the mechanisms underlying the disfluency-as-deception bias. Likewise, we will relate these findings to previous studies investigating biases against non-native speakers. Finally, we will conclude with a discussion regarding the broader aims of the thesis.

6.6.1 The disfluency-as-deception bias

At the beginning of this chapter, we put forward two mechanisms to account for the disfluency-as-deception bias: an associative and an inference account. This proposal was partially motivated by the rationales reported in the literature to account for the effects of disfluency in linguistic prediction (cf. Chapter 4). Predictive processes in language comprehension have been shown to be modulated by the presence of a disfluency; specifically, filled pauses seem to bias comprehenders' anticipations towards elements that coincidentally co-occur with disfluency, arguably because the presence of a disfluency provides a cue that something unexpected might come due to previously learnt associations. These modulations are further subject to who produces the disfluency, suggesting that a priori expectations about a speaker's speech can guide whether a disfluency is a reliable cue (Arnold et al., 2007; Bosker et al., 2014). Whereas these prior experiments do not involve judgements of deception, they provide an interesting point of comparison to the effects reported here. Specifically, we argue that the results reported here are evidence of an associative mechanism leading to the present bias towards deception in the face of disfluency. In experiments involving prediction, the effects of disfluencies are largely transitory: The social costs of repairing a failed prediction while comprehending speech is likely to be low. In these scenarios, it might be worthwhile to entrain and contrast different cues available for the presence of a disfluency.

Why do we then say that our findings support an associative account? Our arguments are twofold. As previously discussed, unless our participants have set on a strategy prior to the experiment whereby only manner of delivery would be used as a cue to detect deceit, an inference mechanism (where different cues are contrasted against the context) predicts an effect of speaker identity. Specifically, the non-native identity we selected

for our manipulation has been previously shown to impact how speech is comprehended (Brehm et al., 2019; Gibson et al., 2017; Hanulíková et al., 2012) and interpreted (Caffarra et al., 2018; Fairchild et al., 2020; Ip & Papafragou, 2022; Lorenzoni et al., 2022). These effects can be partially explained by a priori expectations individuals have of non-native speech, namely, as less linguistically competent (Lev-Ari, 2015). Therefore, this prior expectation should have been a cue that listeners could have considered when integrating the disfluency in comprehension. The lack of delay suggests that identity was not considered when taking the disfluency as evidence for deception, and thus fails to support the inference account.

We argued that one possibility for the lack of effect of speaker identity was that the salience of this cue as an alternative explanation for the presence of a disfluency was not sufficiently salient for listeners to consider, especially in the case of our native participants. However, there is a wealth of research showing that native and non-native participants behave differently when their interlocutor is non-native compared to a native one, and, specifically, that these behaviours can be explained by individuals' beliefs about non-native individuals' linguistic competence. For example, native speakers are more likely to employ an unusual label to refer to an object if said object had been referred to as such by a non-native speaker, in contrast to a native speaker (Ivanova et al., 2007; Cai et al., 2021), native speakers tend to produce more redundant labels when interacting with non-native speakers (Tal et al., 2021), and native individuals are more likely to produce utterances with the same syntactical structure as their non-native interlocutor's if they are presented with evidence suggesting a lower linguistic competence (Loy & Smith, 2019). Similarly, non-native individuals are more likely to reuse unpreferred labels when they have been previously produced by a non-native speaker than by a native speaker (Suffill et al., 2021). Unless non-natives' stereotyped reduced linguistic competence does not entail an increase in disfluency, our results are difficult to reconcile with the idea that both native and non-native listeners are not aware of non-native speakers' linguistic abilities, supporting an associative account underlying the disfluency-as-deception bias.

The second argument follows the same reasoning as proposals suggesting that some inferences might become 'routinised' (Mazzone, 2009, 2013). Indeed, individuals consistently report filled pauses as a trait displayed by liars (Arciuli et al., 2010; Global Deception Research Team, 2016; Zuckerman et al., 198). The imposition of a context whereby the potential utility value of following this heuristic is maximised (i.e., being warned beforehand that the speaker would lie half of the time) might just activate it, bypassing any reasoning when the relationship between the context and a cue is perceived to be high enough.

The origins of this routine are hard to pin down. One possibility is that by exposure to a language, users learn its distributional properties and this knowledge guides their comprehension. Importantly, an initial tuning to a speaker's mental state would evolve to constitute an implicit knowledge that disfluencies accompany 'things that are hard to say', which we could consider filled pauses' conventionalised meaning. Although speakers are no more disfluent when they lie (Loy et al., 2018), listeners still hold the belief that lying is cognitively costly and thus falls under the umbrella of 'things that are hard to say'. Subsequently, disfluency is thought to accompany deception, because being deceptive is hard.

The present experiments cannot disentangle these two possibilities. However, this inference mechanism is particularly hard to reconcile with the fact that non-native participants comprehending their second language produced with a non-native accent displayed a similar time course to when speech was produced with a native accent. Specifically, second language comprehension has been shown to be heavily influenced by knowledge that listeners have mastered, with a particular sensitivity to speaker's identity. Therefore, this particular condition (i.e., non-native listeners attending to non-native-accented speech) is where the combination of these two factors should have been particularly salient and thus the most prone to show an effect of inferences in the form of a delay in the time course of the bias. Yet, our findings showed that the pattern of fixations was indistinguishable from that of those listening to the native speakers.

Alternatively, what may have decreased the utility of the speaker identity as a useful cue may have been our participants' (both native and non-native) experience with and exposure to non-native speakers. Our native listeners reported being exposed to non-native speakers an average of 5.83 (SD = 2.41) on a 9-point scale, while non-native listeners reported this exposure as 6.5 (SD = 2.16). In Chapter 4 we argued that a plausible reason for non-native-accented disfluencies to guide comprehension was listeners' experience with non-native-accented speech. Following on the evidence from predictive disfluencies, listeners might move along a continuum between associations and inferences, whereby the latter are more likely when the cue providing an alternative explanation for a disfluency is reasonable following listeners' experience with language use by different speakers. By the same token, the disfluency-as-deception bias can operate as an associative mechanism unless something signals the listener that it is worth engaging in further reasoning. For example, King et al. (2018) found a delay in the emergence of the bias when the alternative reason for the speaker to be disfluent was momentarily present and accessible for the listener (a car horn). This reasoning leaves the question open of whether listeners who are less exposed to non-native-accented speech would show a modulation of the disfluency bias.

Another concern is whether the pattern of results reported here can be, at least partially, attributable to our participants' expectations. Firstly, because our participants expected the speaker to have deceitful intentions. This deviates from other scenarios wherein individuals are uncertain in the first place about whether the speaker would be deceitful. However, the pattern of results here described support the necessity to include this expectation, as our participants demonstrated a truth-bias in both Experiments 3 and 4. Indeed, there is a wealth of research suggesting that there is a tendency to believe our interlocutors, even within the deception detection literature (e.g., Vrij & Baxter, 1999).

Participants in Experiments 3 and 4 also had expectations about how often the speaker would be deceitful: They were informed that the speaker would lie at least half

of the time, following the design of Loy et al. (2017). This instruction might have shaped participants' decision-making, as technically they had some certainty about the distribution of potential deceit in the experiment. It is possible that in scenarios with different degrees of (un)certainty (e.g., a lower or a higher proportion of expected deceit) participants' decision-making could be skewed (e.g., consistently believe the speaker to be truthful or deceitful to maximise their accuracy). For example, whether individuals count all the cues that support either decision and select the one that has more cues in its favour (i.e., a 'tallying heuristic'; Dawes, 1979) or follow a 'take-the-best' heuristic (i.e., rely on the most important cue to make a decision, Gigerenzer & Goldstein, 1996; in this case, the most important cue to discriminate between a truth and a lie) might be affected by how often they expect to encounter a lie, so that the heuristic followed is that which maximises accuracy. However, whether individuals resort to different heuristics in making these decisions as a function of the expected distribution of deceit falls outwith the scope of this thesis.

6.6.2 Biases against non-native speakers

Another interesting finding from these experiments is the lack of an overall effect of speaker identity. We find no evidence to support previous findings in the deception literature suggesting that native and non-native speakers are evaluated differently by both native (e.g., Lev-Ari & Keysar, 2010) and non-native listeners (Hanzlíková & Skarnitzl, 2017). Although previous studies have found that native listeners display, at best, a truth-bias towards native speakers that non-native speakers do not benefit from, and at worst, a lie-bias against non-native speakers, our findings indicate that both speakers were considered equally truthful overall. This could be due to the differences between the experimental stimuli employed in these experiments and those used elsewhere.

The differences in biases depending on speaker identity that have been reported in the deception literature commonly come from studies wherein participants were presented with larger, more heterogeneous, sets of cues from which to infer deception. For example, Da Silva and Leach (2013) showed participants video footage of native and non-native

speakers who were interrogated about whether they had cheated on a test. The speakers were not following a script but were allowed to answer freely, for an average recording duration of 93 s per response, with varying numbers of cues such as disfluencies in each response. In our study, native and non-native speakers recorded brief scripted sentences, and the number of overt (filled pause) disfluencies was kept constant between speakers. These methodological differences leave open the question of whether non-natives might have been previously judged to be more deceitful as a consequence of producing more disfluencies; but it does seem unlikely that such judgements are (solely) based on their identities.

Evidence for the effects of speaker identity on veracity ratings via decreased processing fluency comes from Lev-Ari and Keysar (2010). Importantly, their findings have been recently replicated (Boduch-Grabka & Lev-Ari, 2021, although see Foucart & Hartsuiker, 2021 for a lack of replication). Boduch-Grabka and Lev-Ari (2021) reported that brief exposure to non-native accented speech prior to the ratings (i.e., approximately 10 min. videos) numerically reduced the negative bias towards trivia statements produced by non-native speakers. This brief exposure improved participants' processing of non-native accented speech as evidenced by participants' better transcriptions of non-native speakers' sentences after being exposed to the non-native accented video (as opposed to those who watched native accented videos). Importantly, the effect of exposure on veracity ratings was mediated by participants' improved comprehension of the non-native accent. In fact, although some studies following similar paradigms investigating processing fluency have failed to replicate Lev-Ari and Keysar's (2010) original findings, they have found differences in neural responses between native and non-native accented speech in terms of group attribution (e.g., Foucart & Hartsuiker, 2021; Foucart et al., 2020). Interestingly, our participants' exposure to non-native accented speech was much longer than in Lev-Ari and Keysar (2010) and Boduch-Grabka and Lev-Ari (2021), which may partially explain the differences in the findings. In our case, although participants' reports of exposure and interaction with non-native speakers did not impact their clicking behaviour, it could be the case of a ceiling effect: Our participants may be, overall, exposed to non-native

accented English so much that the potential effects of processing fluency on veracity ratings are diminished, explaining the differences between our findings and those of Lev-Ari and Keysar (2010) and Boduch-Grabka and Lev-Ari (2021).

6.7 Conclusion

Overall, our results are consistent with a frequent finding that disfluencies are amongst the cues that listeners employ to interpret an utterance as deceptive. Across two studies, we demonstrated that this disfluency-as-deception bias is a heavily anchored cue that can override any other cues present in the context, such as speaker identity. Importantly, the bias is resistant to seemingly increasingly harder tasks, such as comprehending a second language. These results support models where non-literal meaning can be interpreted and integrated quickly and rather cost-free. However, we argue that when individuals are warned about the speaker's intention beforehand, and they hold a priori theories about how this intention might be manifested, deriving an utterance's meaning might not require any form of reasoning: Instead, we believe that some inferences might become 'routinised' and implemented if the context supports such a heuristic as the most useful one to follow.

Chapter 7

General Discussion

Spontaneous speech is characterised by the presence of disfluencies, such as repeated and elongated words, repairs and pauses filled with *um* and *uh*. This thesis explored the last type of disfluency: filled pauses. Previous research has shown that the presence of these elements in an utterance impacts both the underlying processes involved in speech comprehension as well as listeners' interpretations of the utterance's meaning. For example, the presence of filled pauses biases comprehenders' eye movements towards certain elements over others (e.g., discourse-new entities, Arnold et al., 2004; Heller et al., 2015; hard-to-describe objects, Arnold et al., 2007; Watanabe et al., 2008; low-frequency words, Bosker et al., 2014; Bosker et al., 2019) and skews listeners' evaluations of both the speaker and their communicative intent (e.g., uncertainty, Brennan & Williams, 1995; deception, Loy et al., 2017; see also Arciuli et al., 2010; King et al., 2018, Li et al., 2022; saving face, Loy et al., 2019).

Beyond any effect due to the interruption to the speech signal that filled pauses present (what we coin as the fillers-as-time account, see Section 4.1), explanations put forward to account for the comprehension of disfluent speech have fallen into two broad camps: an association and an inference account. Each of these accounts capitalises on different cognitive abilities. An associative account has its roots in individuals' ability to learn the statistical properties of linguistic input. An inference account is based on

individuals' capacity to reason about the speaker and use social information to guide their comprehension of speech. These two mechanisms are not mutually exclusive, and it has been argued that different properties of a communicative context (e.g., properties of the signal, Bosker et al., 2019) can give relevance to either mechanism.

Throughout this thesis, we have juxtaposed these two accounts. We have described how these two accounts can explain the biases exerted by disfluencies as originating from different properties of filled pauses (distribution and origins, respectively) and how they differ across a set of factors (experience, cognitive resources, speaker-orientation). An associative account can be thus characterised as emerging from users' experiences with and exposure to a language, leading to a habitual, fast, but also inflexible system. An inferential account proposes that listeners reason about the causes for a speaker to be disfluent, and therefore is a context-dependent, slow, but flexible system. We proposed that these differences have consequences for how filled pauses will affect speech comprehension as a function of speakers' and listeners' characteristics, specifically in the time course of comprehending disfluent speech.

In this thesis, we attempted to distinguish these accounts by exploring the effect of linguistic background as an attribute of speakers and listeners across two contrasting communicative contexts: following instructions and interpreting deceit. We argued that exploring the comprehension of disfluent speech cross-linguistically and across participants would shed light on the mechanisms involved. By having listeners attend to speech produced by a native or a non-native speaker, we explored the flexibility of the effects of disfluent speech. Further, we proposed that the properties of second-language comprehension could help us understand how disfluent speech is comprehended. This is due to (1) the possibility that it imposes an additional cognitive load on comprehenders, thus allowing us to test the predictions of the associative and the inference accounts in terms of costs, and (2) the fact that second-language comprehension may rely on cues that comprehenders have mastered well, thus allowing us to test whether the reduced experience with a language may lead comprehenders to revert to social reasoning. We

investigated the interplay of these two factors across two communicative scenarios. We used the prediction of lexical items and the interpretation of deceit as test beds: The former is a context that may give rise to disfluent speech due to difficulties in producing a label while the latter capitalises on disfluencies as a potential index of deceit.

We explored how the comprehension of disfluent speech varies across the combination of these factors in four eye-tracking experiments. Chapter 4 described two experiments exploring whether the presence of a filled pause biased listeners' expectations towards objects with low-frequency (versus high-frequency) labels, depending on the linguistic background of the speaker and the listener. Chapter 6 presented two experiments where the interpretation elicited by disfluent speech was investigated by examining the emergence of the pragmatic inference of deception, again depending on the speaker and listeners' linguistic backgrounds. This present chapter briefly summarises the main findings of these experiments, in relation to the aims of this thesis and to the broader context of disfluencies, and suggests future avenues for research.

7.1 Disfluency as difficulty

Part I of this thesis explored the mechanisms proposed to underlie prediction in speech, with a focus on how disfluencies can guide anticipations about upcoming linguistic units. We tested whether what listeners can expect can be best accounted for in terms of individuals' learning of the speech regularities or by individuals' reasoning about speakers by exploring what listeners anticipate following a disfluency produced by a native or a non-native speaker. We also explored whether non-native listeners, for whom the word-frequency effect is potentially larger, could equally inform their comprehension following a disfluency.

In Chapter 4, we replicated and extended Bosker et al.'s (2014) eye-tracking experiments, where native Dutch listeners were presented with two pictures, one of which had a high- and one a low-frequency name, and were asked to follow the instructions of either a native or a non-native Dutch speaker who could refer to either object fluently or

disfluently. Bosker et al. (2014) reported that listeners displayed anticipatory eye movements towards low-frequency items following a filled pause produced by a native speaker, but not for those produced by a non-native speaker. In Experiment 1, we attempted to replicate Bosker et al. (2014) in a sample of native English listeners, while Experiment 2 extended this paradigm to a sample of non-native English listeners. Our experiments failed to replicate Bosker et al.'s (2014) findings. Instead, in Experiment 1 we found that native listeners displayed anticipatory fixations towards the low-frequency item following a disfluency produced by a non-native speaker, rather than by a native speaker. In Experiment 2, we did not find any evidence for prediction following a disfluency for either speaker. We argued that the lack of anticipatory eye movements for our disfluent native speaker is likely a consequence of the (lack of) perception of hesitation in this speaker, a point we will address in Section 7.3.

Separate from predictive processing, additional analyses suggested that the presence of a filled pause aided in the recognition of the named object, as reflected in a larger increase in looks over time on the target in comparison to when these targets were produced fluently post-target onset. For native listeners (Exp. 1), the increase in fixations following a filled pause (in comparison to when the speaker was fluent) was more pronounced for low-frequency labels than for high-frequency ones - which, in fact, decreased participants' fixations to the target. This pattern of fixations was similar across speaker conditions. For non-native listeners (Exp. 2), while a similar pattern to that of Experiment 1 emerged for those who listened to a native speaker, filled pauses produced by a non-native speaker benefited the recognition of both low- and high-frequency items.

This pattern of results partially aligns with an associative account. Starting with Experiment 1, Bosker et al. (2014) interpreted the lack of prediction following non-native disfluencies as evidence for an associative account. Listeners' experience with the properties of non-native-accented speech (whereby the distribution of filled pauses is more arbitrary) diminishes the predictive value of non-natives' filled pauses. Bosker et al.'s (2014) participants, however, reported being exposed to non-native-accented speech at

an average of 3.83 (SD = 2.13) on a 9-point scale. In contrast, in the present experiment, native participants' exposure to non-native-accented speech and experience with non-native speakers was relatively higher (6.08 (SD = 2.34) and 6.13 (SD = 2.36)). Given previous findings suggesting that exposure to non-native-accented speech can ease its processing and revert it to that of native-accented speech (Porretta et al., 2017; Porretta et al., 2020), it is possible that Bosker et al.'s (2014) findings are attributable to a stereotype about non-native speakers' speech while our participants' experience in daily life with non-native speakers decreased the reliability of those stereotypes.

Due to the nature of our auditory stimuli, where the native speaker's disfluency may not have been sufficiently salient, only the non-native speaker's disfluency was a likely candidate to increase fixations towards the low-frequency object before a label was encountered (similar to Experiment 1). In Experiment 2, however, we failed to find such anticipatory fixations. Previous research has suggested that predictive processes in second-language comprehension may emerge later than in first-language comprehension, due to reduced automaticity (Ito & Pickering, 2021) and increased cognitive load (Munro & Derwing, 1995) and, in the present study, with the addition of an increased word-frequency effect (Duyck et al., 2008). We take the pattern of fixations post-target onset as partially supporting an associative account. In this window, we found that filled pauses produced by a native speaker facilitated the recognition of the low-frequency label compared to when it was produced fluently. In contrast, filled pauses produced by a non-native speaker facilitated the recognition of both high- and low-frequency items. This latter finding, in fact, aligns with the fillers-as-time proposal: Filled pauses may 'just' ease word segmentation, and potentially re-orient listeners' attention to the speech signal, so that the integration of upcoming linguistic elements is facilitated.

Overall, these patterns can be understood as reflecting listeners' experience of language. Whether a disfluency aids speech comprehension (either by predictive processing or word recognition) depends on how used individuals are to the specifics of the signal (i.e., accent). If comprehenders are not used to it, comprehension is likely to be informed

by the beliefs listeners hold about speakers (i.e., following an inference account), as Bosker et al. (2014) found. When listening conditions are adverse, such as attending to one's second language produced with a non-native accent, the benefits of disfluency might not go beyond the interruption of the speech signal.

7.2 Disfluency as deception

In Part II, we explored how the presence of a disfluency induced listeners to infer meaning that differs from the content of the utterance, for example, when the listener believes they may be being deceived. We tested whether this interpretation is due to the fast application of a heuristic, whereby listeners associate disfluent speech with deceit, or if it reflects listeners' reasoning about the speaker. To do so, we explored whether the presence of an alternative cause for the speaker to be disfluent (i.e., producing speech in their second language) could modulate this interpretation. Further, we explored whether this interpretation depended on listeners' cognitive resources and the strategies they use to comprehend speech by testing a sample of non-native listeners.

In Chapter 6, we extended and replicated Loy et al.'s (2017) eye-tracking experiment, in which native English listeners were presented with a potentially deceitful speaker who would refer to the location of some treasure either fluently or disfluently (*The treasure is behind* ***the***/***thee um***). Experiment 3 replicated and extended Loy et al. (2017) to test whether listeners interpret language in this task differently when it is produced by a native versus non-native speaker. Experiment 4 extended this paradigm to a sample of non-native English listeners.

Manner of delivery has consistently been reported as the main factor impacting listeners' interpretations of deceit (Arciuli et al., 2010; King et al., 2018; Li et al., 2022), and the results of Experiments 3 and 4 provide further evidence of this. Both native and non-native listeners were more likely to interpret disfluent utterances as deceitful, regardless of the speaker's linguistic background. Crucially, this interpretation emerged similarly across speakers and listeners: At around 400 ms post-target onset, participants'

fixations were reflective of their interpretation of deceit. These patterns did not differ between listeners, nor did they become significantly enhanced towards the end of the experiment as a result of learning.

We take this pattern of results as supporting an associative account too. We initially argued that if social reasoning drove the interpretation of disfluent speech as deceitful, then attending to a speaker who may produce disfluencies for reasons other than deception (e.g., producing speech in their second language) might delay its emergence (see King et al., 2018). However, the time course of fixations did not differ as a function of the speaker's nativeness. We argued that there is one alternative possibility: The alternative reason for the speaker to be disfluent might not have been particularly salient for native listeners. Nonetheless, when listeners are cognitively taxed, but also may be more likely to consider speaker identity to compute an interpretation (i.e., when listeners are comprehending speech in their non-native language) the time course of this disfluency-as-deception interpretation did not differ from that found when listeners are attending to their first language. This suggests that this interpretation of disfluent speech as deceptive has its roots in a stereotype (or heuristic) of how deception sounds. One explanation considers 'habits of speech' (Gershman & Goodman, 2014): Upon exposure and language use learning, some interpretations are 'routinised' and thus reduce the necessity of the listener to model the speaker, easing the computational burden (Elman et al., 2004). For example, a repeated inference can yield an associative relationship (cf. Mazzone, 2009, 2013) so that its interpretation is eased and thus computed relatively quickly and cost-free.

7.3 Accounts for comprehending disfluent speech

As we have reviewed, the findings of this thesis largely support an associative account. Throughout four eye-tracking studies, we found a relatively early effect of filled pauses in the integration and interpretation of speech. However, throughout this thesis, we have argued that an associative and an inference account are not mutually exclusive. As Hanus

(2016) argued (within learning): "False dichotomies create false battles (...) Instead of asking what the circumstances are that allow or trigger associative and cognitive capacities in (human and nonhuman) animals" (Hanus, 2016, p. 246). He advocates that the question should be whether learning is due to blind associations or some form of inference. Translating this proposal onto our topic, a better question is what underlies the associative account of disfluent speech comprehension.

Speculatively, what listeners may have learnt is an association between filled pauses *and things that are hard to say.* This association may have its roots in the ability to keep track of our interlocutor's mental state (i.e., theory of mind, or mentalising). Some authors have proposed that we are equipped with a set of mechanisms to monitor our interlocutors' reliability, known as epistemic vigilance (Sperber et al., 2010), which allows us to evaluate both our interlocutors' competence and honesty (Breheny et al., 2013; Mazzaggio et al., 2021). As individuals' experience with language use increases, inferences about an interlocutor's mental state following a filled pause become routinized, leading to an association - in this sense, the association is a form of 'crystallised' social learning. What is hard to say depends on the context, so this association is mapped onto listeners' expectations: The context disambiguates the filled pause. An interpretation between the properties of the signal (e.g., different speakers), the comprehender (e.g., their experience with language) and the context (e.g., perceived communicative intent) may lead the system to revert back to an inference. In this regard, comprehension of disfluent speech follows an associative account unless there is a signal that these associations might not be relevant. The evidence presented here may be thus taken as showing how listeners' increased experience with different signals (in the form of native and non-native speakers) translates into the system relying on an association.

Understanding the comprehension of disfluent speech as a component of a general system that monitors speakers' mental states might have consequences for the amount of evidence the system requires to perceive the speaker as uncertain. Throughout this thesis, we have been discussing disfluencies from the perspective that they are individual

phenomena, such as filled pauses, that form part of the speech signal. However, Segalowitz (2010) identifies this *utterance fluency* as one of three types of fluency. The others are *cognitive fluency*, the speaker's ability to efficiently produce language by bringing to play the relevant cognitive processes, and *perceived fluency*: 'the inferences listeners make about a speaker's cognitive fluency based on their perception of utterance fluency' (Segalowitz, 2010, p.48). The crucial component for the experimental work reported here is the difference between utterance and perceived fluency: Utterance fluency is a fact, whilst perceived fluency is subjective. Segalowitz (2010) states that listeners do not normally treat every pause and hesitation as evidence of disfluency - some degree of hesitation is acceptable and even expected.

The latter point reflects another interesting aspect of the present thesis: The evidence required to consider that the speaker is disfluent may depend on the task at hand. In Experiments 1 and 2, we argued that one possibility for a native disfluency not to have impacted predictive processing may have been due to the lack of evidence before the filled pause that the speaker was going to be disfluent. Arnold et al. (2003) argued that the effects of disfluency in predictive processing might be attributable to factors beyond the presence of a filled pause, such as the prosody of an utterance. Interestingly, in Experiments 3 and 4, listeners only had the filled pause as evidence, and this was enough for listeners to interpret deceit. This suggests that an important component of the signal is how prototypical it sounds, in terms of how much evidence is needed to perceive (lack of) confidence.

7.4 Limitations and future directions

The work presented here is just a first step into deepening our understanding of the comprehension of disfluent speech. Many of the conclusions drawn in this chapter are rather speculative, and future research should address the predictions presented here as well as work on the limitations of the experiments reported here.

Foremost, it would be important to run similar experiments with more carefully designed materials for Experiments 1 and 2. The lack of prior evidence that the speaker was going to be disfluent, alongside the differences in the production of the filled pause between the native and the non-native speaker, prevent us from drawing clear conclusions about the mechanisms that underlie the effects of disfluency in predictive processing. Given the variability in auditory stimuli employed in disfluent speech comprehension research, future research could explore what is the threshold of perceived fluency for a filled pause to exert an effect on predictive processing.

It is worth noting that Experiments 3 and 4 capitalised on the experimental task: Participants were instructed to judge the speaker's veracity. Testing comprehenders' pragmatic interpretations outside of this paradigm may shed light on the circumstances in which filled pauses' conventional meaning is enriched. One prediction of the proposal presented here is that the enrichment (or the lack thereof) depends on the costs comprehenders may incur if they do not compute the most relevant interpretation. Previous research has shown that non-native speakers' pragmatic failures, such as being underinformative, are more likely to be forgiven (Fairchild et al., 2020; Lorenzoni et al., 2022). In scenarios where not accurately decoding the speaker's intent does not entail social costs for the comprehender, it may be possible to find different interpretations for disfluent speech depending on the speaker's characteristics. Likewise, different characteristics of the communicative situation may also change how listeners attribute uncertainty. What would happen if, for example, participants had been shown pairs of high- and low-frequency items in the treasure-hunt task paradigm? Would there be a delay in the emergence of the bias in comparison to only seeing items that are arguably easy to produce? Exploring these questions would clarify what further characteristics may lead listeners to infer *in situ*, as opposed to relying on an association.

Future research should explore the idea that filled pauses belong to a larger class of elements that are taken as signs of uncertainty. For example, languages are equipped with a set of verbal elements that signal a speaker's degree of commitment to the utterance

(e.g., *probably*, *perhaps*; Roseano et al., 2016) as well as to mark contextual expectations (e.g., *indeed*, *actually*; van Bergen & Bosker, 2018; Rasenberg et al., 2020) that listeners are responsive to. In Dutch, the discourse marker *er* commonly precedes less predictable (yet plausible) elements in the sentence context. Grondelaers et al. (2009) demonstrated that the reading time of sentences with unpredictable endings was reduced when they were preceded by *er*, suggesting that the presence of *er* eased the integration of unpredictable endings. Grondelaers et al. (2009) argue that this reflects listeners' abandonment of the predictions they may have initiated given the sentence context following this discourse marker. It is possible that perceived uncertainty yields similar responses.

Finally, an important question that arises from the claim that individuals have an association between filled pauses and the elements with which they co-occur is whether speakers intentionally produce them to enrich the meaning of an utterance. For example, a speaker might want to express that an event was unpleasant but do it politely by adding a filled pause. Interlocutors' knowledge that a filled pause signals something hard to say may lead the speaker to intentionally produce them to enrich their message. Future research should examine whether there are differences in the production of filled pauses that clearly are said to enrich meaning to investigate whether listeners are actually sensitive to this display. One venue to test this hypothesis is by exploring written disfluencies. Given that the written modality is not the natural medium of disfluencies, this would test whether written disfluencies only make sense when produced to imply something different from the actual content of the message.

7.5 Conclusion

The experiments carried out in the present thesis are an attempt to comprehensively investigate the different effects of filled pauses at the processing and interpretation levels of speech comprehension. The novelty of this research was to consider both the characteristics of who speaks and who listens (via linguistic background) as additional parameters in the effects of disfluent speech comprehension both at the level of predictive processing

and meaning interpretation. The results of the experimental work presented here support an account whereby these effects are the outcome of language users' experience with language: Exposure to disfluency leads to the creation of an association between filled pauses and elements that contextually co-occur with them, learning that may have its roots in comprehenders' ability to monitor their interlocutors. In this sense, this association is just the routinisation of a repeated inference about the speaker.

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*(2), 249–306.

Agmon, G., Jaeger, M., Tsarfaty, R., Bleichner, M. G., & Golumbic, E. Z. (2022). "um. . . , it's really difficult to. . . um. . . speak fluently": Neural tracking of spontaneous speech. *bioRxiv*.

Akehurst, L., Köhnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behaviour. *Applied Cognitive Psychology*, *10*(6), 461–471.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*(4), 583–609.

Altmann, G. T. (2011). The mediation of eye movements by spoken language. In S. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The oxford handbook of eye movements* (pp. 979–1003). Oxford University Press.

Altmann, Y., Gerry T.M.and Kamide. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518.

Antoniou, K., Veenstra, A., Kissine, M., & Katsos, N. (2018). The impact of childhood bilingualism and bi-dialectalism on pragmatic interpretation and processing. *BUCLD 42: Proceedings of the 42nd annual Boston University Conference on Language Development*, *1*, 15–28.

Arciuli, J., Mallard, D., & Villar, G. (2010). "Um, I can tell you're lying": Linguistic markers of deception versus truth-telling in speech. *Applied Psycholinguistics*, *31*(3), 397–411.

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

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 914–930.

Arnold, J. E., & Tanenhaus, M. K. (2007). Disfluency effects in comprehension: How new information can become accessible. *The processing and acquisition of reference*, 197–217.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new. *Psychological Science*, *15*(9), 578–582.

Arslan, B., & Göksun, T. (2022). Aging, gesture production, and disfluency in speech: A comparison of younger and older adults. *Cognitive Science*, *46*(2), e13098.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using r*. Cambridge University Press.

Bachman, L. F., & Palmer, A. S. (1989). The construct validation of self-ratings of communicative language ability. *Language testing*, *6*(1), 14–29.

Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies influence parsing of garden path sentences. *Journal of Memory and Language*, *49*(2), 183–200.

Bailey, K. G., & Ferreira, F. (2005). The disfluent hairy dog: Can syntactic parsing be affected by nonword disfluencies? In *Approaches to studying world-situated language*

use: Bridging the language-as-product and language-as-action traditions (John C. Trueswell and Michael K. Tanenhaus, p. 303). MIT Press.

Bailey, K. G., & Ferreira, F. (2007). The processing of filled pause disfluencies in the visual world. In *Eye movements: A window on mind and brain* (pp. 487–502). Elsevier.

Baker, C., & Love, T. (2022). It's about time! time as a parameter for lexical and syntactic processing: An eye-tracking-while-listening investigation. *Language, Cognition and Neuroscience*, *37*(1), 42–62.

Barr, D. J. (2001). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. *Oralité et gestualité: Interactions et comportements multimodaux dans la communication*, 597–600.

Barr, D. J. (2003). Paralinguistic correlates of conceptual structure. *Psychonomic Bulletin & Review*, *10*, 462–467.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474.

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.

Barr, D. J., & Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes*, *25*(4), 441–455.

Bazzi, L., Brouwer, S., & Foucart, A. (2022). The impact of foreign accent on irony and its consequences on social interaction. *Journal of Multilingual and Multicultural Development*, 1–13.

Beattie, G. W., & Butterworth, B. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, *22*(3), 201–211.

Bellinghausen, C., Gósy, M., Rauh, R., Schröder, B., Fangmeier, T., van Elst, L. T., & Riedel, A. (n.d.). Disfluency patterns in speech of children with and without

autism spectrum disorder. *SPPL2020: 2nd Workshop on Speech Perception and Production across the Lifespan*, 22.

Belz, M., & Reichel, U. (2015). Pitch characteristics of filled pauses. *The 7th Workshop on Disfluency in Spontaneous Speech (DiSS)*.

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *114*(3), 1600–1610.

Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1450.

Bergmann, C., Sprenger, S. A., & Schmid, M. S. (2015). The impact of language co-activation on L1 and L2 speech fluency. *Acta Psychologica*, *161*, 25–35.

Berkum, J. J. A. V., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591.

Bialystok, E. (2015). Bilingualism and the development of executive function: The role of attention. *Child Development Perspectives*, *9*(2), 117–121.

Blau, E. K. (1991). More on comprehensible input: The effect of pauses and hesitation markers on listening comprehension. *Annual Meeting of the Puerto Rico Teachers of English to Speakers of Other Languages, San Juan, Puerto Rico.*

Bloomfield, A., Wayland, S. C., Rhoades, E. A., Blodgett, A., Linck, J., & Ross, S. J. (2010). What makes listening difficult? factors affecting second language listening comprehension.

Blumenfeld, H. K., & Marian, V. (2013). Parallel language activation and cognitive control during spoken word recognition in bilinguals. *Journal of Cognitive Psychology*, *25*(5), 547–567.

Boduch-Grabka, K., & Lev-Ari, S. (2021). Exposing individuals to foreign accent increases their trust in what nonnative speakers say. *Cognitive science*, *45*(11), e13064.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234.

Bonnefon, J.-F., Hopfensitz, A., & De Neys, W. (2015). Face-ism and kernels of truth in facial inferences. *Trends in cognitive sciences*, *19*(8), 421–422.

Bonnefon, J.-F., & Villejoubert, G. (2005). Communicating likelihood and managing face: Can we say it is probable when we know it to be certain? *Proceedings of the Annual Meeting of the Cognitive Science Society*, *27*(27).

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

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2015). Both native and non-native disfluencies trigger listeners' attention. *Proceedings of Disfluency in Spontaneous Speech 2015*, 1–4.

Bosker, H. R., Badaya, E., & Corley, M. (2021). Discourse markers activate their, like, cohort competitors. *Discourse Processes*, *58*(9), 837–851.

Bosker, H. R., Quené, H., Sanders, T., & De Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language*, *75*, 104–116.

Bosker, H. R., Van Os, M., Does, R., & Van Bergen, G. (2019). Counting 'uhm's: How tracking the distribution of native and non-native disfluencies influences online language comprehension. *Journal of Memory and Language*, *106*, 189–202.

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, *66*(1), 123–142.

Bott, L., & Frisson, S. (2022). Salient alternatives facilitate implicatures. *PLoS One*, *17*(3), e0265781.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437–457.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729.

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

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition*, *126*(3), 423–440.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? an on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*(3), 434–463.

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

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, *44*(2), 274–296.

Brennan, S. E., & Williams, M. (1995). The feeling of another s knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*(3), 383–398.

Bromberek-Dyzman, K., Jankowiak, K., & Chełminiak, P. (2021). Modality matters: Testing bilingual irony comprehension in the textual, auditory, and audio-visual modality. *Journal of Pragmatics*, *180*, 219–231.

Brown, C., & Hagoort, P. (1993). The processing nature of the n400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*(1), 34–44.

Brown, P., Levinson, S. C., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge University Press.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in german. *Experimental Psychology*, *58*, 412–424.

Buck, G. (2001). *Assessing listening*. Cambridge University Press.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, *30*(5), 542–579.

Caffarra, S., Michell, E., & Martin, C. D. (2018). The impact of foreign accent on irony interpretation. *PLoS One*, *13*(8), e0200939.

Cai, Z. G., Sun, Z., & Zhao, N. (2021). Interlocutor modelling in lexical alignment: The role of linguistic competence. *Journal of Memory and Language*, *121*, 104278.

Cai, Z. G., Zhao, N., & Pickering, M. J. (2022). How do people interpret implausible sentences? *Cognition*, *225*, 105101.

Carney, N. (2022). L2 comprehension of filled pauses and fillers in unscripted speech. *System*, *105*, 102726.

Castillo, P. A., Tyson, G., & Mallard, D. (2014). An investigation of accuracy and bias in cross-cultural lie detection. *Applied Psychology in Criminal Justice*, *10*(1).

Cevasco, J., & van den Broek, P. (2016). The effect of filled pauses on the processing of the surface form and the establishment of causal connections during the comprehension of spoken expository discourse. *Cognitive Processing*, *17*, 185–194.

Cheng, K. H. W., & Broadhurst, R. (2005). The detection of deception: The effects of first and second language on lie detection ability. *Psychiatry, Psychology and Law*, *12*(1), 107–118.

Cheng, L. S., Burgess, D., Vernooij, N., Solıs-Barroso, C., McDermott, A., & Namboodiripad, S. (2021). The problematic concept of native speaker in psycholinguistics: Replacing vague and harmful terminology with inclusive and accurate measures. *Frontiers in Psychology*, 3980.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond* (pp. 39–103). Oxford, UK.

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry*, *37*(4), 535–590.

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., Meroni, L., et al. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures.

Proceedings of the 25th Boston University conference on language development, *25*, 157–168.

Christenfeld, N. (1995). Does it hurt to say um? *Journal of Nonverbal Behavior*, *19*, 171–186.

Christenfeld, N., & Creager, B. (1996). Anxiety, alcohol, aphasia, and ums. *Journal of Personality and Social Psychology*, *70*(3), 451.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, e62.

Chun, E., & Kaan, E. (2019). L2 prediction during complex sentence processing. *Journal of Cultural Cognitive Science*, *3*(2), 203–216.

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*(12), 564–570.

Clark, H. H. (1994). Managing problems in speaking. *Speech communication*, *15*(3-4), 243–250.

Clark, H. H. (1997). Dogmas of understanding. *Discourse Processes*, *23*(3), 567–598.

Clark, H. H. (2002). Speaking in time. *Speech communication*, *36*(1-2), 5–13.

Clark, H. H. (2006). Pragmatics of language performance. In *The handbook of pragmatics* (pp. 365–382). Blackwell Publishing Ltd Oxford, UK.

Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, *84*(1), 73–111.

Collard, P., Corley, M., MacGregor, L. J., & Donaldson, D. I. (2008). Attention orienting effects of hesitations in speech: Evidence from ERPs. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(3), 696–702.

Colwell, L. H., Miller, H. A., Lyons Jr, P. M., & Miller, R. S. (2006). The training of law enforcement officers in detecting deception: A survey of current practices and suggestions for improving accuracy. *Police Quarterly*, *9*(3), 275–290.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84–107.

Corley, M., & Hartsuiker, R. J. (2003). Hesitation in speech can... um... help a listener understand. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *25*(25).

Corley, M., & Hartsuiker, R. J. (2011). Why um helps auditory word recognition: The temporal delay hypothesis. *PLoS one*, *6*(5), e19792.

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, *105*(3), 658–668.

Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of um. *Language and Linguistics Compass*, *2*(4), 589–602.

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

Corps, R. E., Liao, M., & Pickering, M. J. (2022). Evidence for two stages of prediction in non-native speakers: A visual-world eye-tracking study. *Bilingualism: Language and Cognition*, *26*(1), 231–243.

Cossavella, F., & Cevasco, J. (2021). The importance of studying the role of filled pauses in the construction of a coherent representation of spontaneous spoken discourse. *Journal of Cognitive Psychology*, *33*(2), 172–186.

Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1057.

Crible, L., Degand, L., & Gilquin, G. (2017). The clustering of discourse markers and filled pauses: A corpus-based french-english study of (dis) fluency. *Languages in Contrast*, *17*(1), 69–95.

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, *20*(4), 659–678.

Cutler, A. (2005). Lexical stress. In *The Handbook of Speech Perception* (pp. 264–289). Oxford: Blackwell.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive psychology*, *42*(4), 317–367.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in french. *Journal of Memory and Language*, *42*(4), 465–480.

Dall, R., Wester, M., & Corley, M. (2014). The effect of filled pauses and speaking rate on speech comprehension in natural, vocoded and synthetic speech. *Fifteenth Annual Conference of the International Speech Communication Association*.

Davies, A. (2003). *The native speaker: Myth and reality* (Vol. 38). Multilingual matters.

Dawes, R. M. (2008). The robust beauty of improper linear models in decision making. In *Rationality and social responsibility* (pp. 321–344). Psychology Press.

De Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences*, *9*(3), 33.

De Cat, C., Klepousniotou, E., & Baayen, R. H. (2015). Representational deficit or processing effect? an electrophysiological study of noun-noun compound processing by very advanced l2 speakers of english. *Frontiers in psychology*, *6*, 77.

De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 113–132.

De Leeuw, E. (2007). Hesitation markers in english, german, and dutch. *Journal of Germanic Linguistics*, *19*(2), 85–114.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology*, *54*(2), 128–133.

Degand, L., & Gilquin, G. (2013). The clustering of 'fluencemes' in french and english. *7th international contrastive linguistics conference (ICLC 7)-3rd conference on using Corpora in contrastive and translation studies (UCCTS 3)*.

Degen, J., & Goodman, N. (2014). Lost your marbles? the puzzle of dependent measures in experimental pragmatics. *Proceedings of the annual meeting of the cognitive science society*, *36*(36).

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive science*, *39*(4), 667–710.

Degen, J., & Tanenhaus, M. K. (2019). Constraint-based pragmatic processing. In C. Cummins & N. Katsos (Eds.), *The oxford handbook of experimental semantics and pragmatics.* (pp. 21–38). Oxford University Press.

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.

Dell, G. S., & Chang, F. (2014). The p-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2017). Is there a replication crisis? perhaps. is this an example? no: A commentary on Ito, Martin, and Nieuwland (2016). *Language, Cognition and Neuroscience*, *32*(8), 966–973.

DePaulo, B. M., Blank, A. L., Swaim, G. W., & Hairfield, J. G. (1992). Expressiveness and expressive control. *Personality and Social Psychology Bulletin*, *18*(3), 276–285.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74.

DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception 1. *Journal of Applied Social Psychology*, *16*(3), 249–267.

DePaulo, B. M., Wetzel, C., Weylin Sternglanz, R., & Wilson, M. J. W. (2003). Verbal and nonverbal dynamics of privacy, secrecy, and deceit. *Journal of Social Issues*, *59*(2), 391–410.

Dewaele, J. M., & Pavlenko, A. (2003). Productivity and lexical diversity in native and non-native speech: A study of cross-cultural effects. In V. Cook (Ed.), *Effects of the second language on the first* (p. 120). Multilingual Matters.

Dewaele, J.-M., & Furnham, A. (1999). Extraversion: The unloved variable in applied linguistic research. *Language Learning*, *49*(3), 509–544.

Diachek, E., & Brown-Schmidt, S. (2022). The effect of disfluency on memory for what was said. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, *66*(5), 843–863.

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition*, *20*(5), 917–930.

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2019). Prediction and integration of semantics during L2 and L1 listening. *Language, Cognition and Neuroscience*, *34*(7), 881–900.

Donnelly, S., & Verkuilen, J. (2017). Empirical logit analysis is not logistic regression. *Journal of Memory and Language*, *94*, 28–42.

Dragojevic, M., Giles, H., Beck, A.-C., & Tatum, N. T. (2017). The fluency principle: Why foreign accent strength negatively biases language attitudes. *Communication Monographs*, *84*(3), 385–405.

Duez, D. (1985). Perception of silent pauses in continuous speech. *Language and speech*, *28*(4), 377–389.

Dupuy, L., Stateva, P., Andreetta, S., Cheylus, A., Déprez, V., Van der Henst, J.-B., Jayez, J., Stepanov, A., & Reboul, A. (2019). Pragmatic abilities in bilinguals: The case of scalar implicatures. *Linguistic Approaches to Bilingualism*, *9*(2), 314–340.

Dussias, P. E., Kroff, J. R. V., Tamargo, R. E. G., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 spanish. *Studies in Second Language Acquisition*, *35*(2), 353–387.

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in spanish–english bilinguals. *Bilingualism: Language and Cognition*, *10*(1), 101–116.

Duyck, W. (2005). Translation and associative priming with cross-lingual pseudohomophones: Evidence for nonselective phonological activation in bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1340.

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 663.

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition.

Eklund, R., & Ingvar, M. (2016). Supplementary motor area activation in disfluency perception: An fmri study of listener neural responses to spontaneously produced unfilled and filled pauses. *Understanding Speech Processing in Humans and Machines, September 8-12, 2016, The Hyatt Regency, San Francisco, California, USA*, 1378–1381.

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

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, *32*(1), 88–106.

Elliott, E., & Leach, A.-M. (2016). You must be lying because i don't understand you: Language proficiency and lie detection. *Journal of Experimental Psychology: Applied*, *22*(4), 488–499.

Elman, J. L., Hare, M., & McRae, K. (2004). Cues, constraints, and competition in sentence processing. In M. Tomasello & D. I. Slobin (Eds.), *Beyond naturenurture: Essays in honor of elizabeth bates* (pp. 111–138). Erlbaum.

Evans, J. R., & Michael, S. W. (2013). Detecting deception in non-native english speakers. *Applied Cognitive Psychology*, *28*(2), 226–237.

Evans, J. R., Michael, S. W., Meissner, C. A., & Brandon, S. E. (2013). Validating a new assessment method for deception detection: Introducing a psychologically based credibility assessment tool. *Journal of Applied Research in Memory and Cognition*, *2*(1), 33–41.

Evans, J. R., Pimentel, P. S., Pena, M. M., & Michael, S. W. (2017). The ability to detect false statements as a function of the type of statement and the language proficiency of the statement provider. *Psychology, Public Policy, and Law*, *23*(3), 290–300.

Fairchild, S., Mathis, A., & Papafragou, A. (2020). Pragmatics and social meaning: Understanding under-informativeness in native and non-native speakers. *Cognition*, *200*, 104171.

Fairchild, S., & Papafragou, A. (2018). Sins of omission are more likely to be forgiven in non-native speakers. *Cognition*, *181*, 80–92.

Fairchild, S., & Papafragou, A. (2021). The role of executive function and theory of mind in pragmatic computations. *Cognitive Science*, *45*(2), e12938.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505.

Feng, J., Cho, S., & Luk, G. (2023). Assessing theory of mind in bilinguals: A scoping review on tasks and study designs. *Bilingualism: Language and Cognition*, 1–15.

Ferreira, F., & Chantavarin, S. (2018). Integration and prediction in language processing: A synthesis of old and new. *Current Directions in Psychological Science*, *27*(6), 443–448.

Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language*, *69*(3), 165–182.

Ferreira, F., Lau, E. F., & Bailey, K. G. (2004). Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science*, *28*(5), 721–749.

Ferreira, F., & Yang, Z. (2019). The problem of comprehension in psycholinguistics. *Discourse Processes*, *56*(7), 485–495.

Finlayson, I. R., & Corley, M. (2012). Disfluency in dialogue: An intentional signal from the speaker? *Psychonomic Bulletin & Review*, *19*, 921–928.

Fitz, H., & Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*, *111*, 15–52.

Foster, K. (1976). Accessing the mental lexicon. *New approaches to language mechanisms*, 257–287.

Foucart, A., Costa, A., Morıs-Fernández, L., & Hartsuiker, R. J. (2020). Foreignness or processing fluency? On understanding the negative bias toward foreign-accented speakers. *Language Learning*, *70*(4), 974–1016.

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*(1), 226–248.

Foucart, A., Garcia, X., Ayguasanosa, M., Thierry, G., Martin, C., & Costa, A. (2015). Does the speaker matter? Online processing of semantic and pragmatic information in L2 speech comprehension. *Neuropsychologia*, *75*, 291–303.

Foucart, A., & Hartsuiker, R. J. (2021). Are foreign-accented speakers that 'incredible'? The impact of the speaker's indexical properties on sentence processing. *Neuropsychologia*, *158*, 107902.

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461.

Foucart, A., Romero-Rivas, C., Gort, B. L., & Costa, A. (2016). Discourse comprehension in L2: Making sense of what is not explicitly said. *Brain and Language*, *163*, 32–41.

Foucart, A., Ruiz-Tada, E., & Costa, A. (2016). Anticipation processes in L2 speech comprehension: Evidence from ERPs and lexical recognition task. *Bilingualism: Language and Cognition*, *19*(1), 213–219.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34*(6), 709–738.

Fox Tree, J. E. (2001). Listeners uses of um and uh in speech comprehension. *Memory & Cognition*, *29*(2), 320–326.

Fox Tree, J. E. (2002). Interpreting pauses and ums at turn exchanges. *Discourse processes*, *34*(1), 37–55.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, *62*(2), 151–167.

Fox Tree, J. E., & Schrock, J. C. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language*, *40*(2), 280–295.

Frank, M. G., & Feeley, T. H. (2003). To catch a liar: Challenges for research in lie detection training. *Journal of Applied Communication Research*, *31*(1), 58–75.

Franke, M., De Jager, T., & Van Rooij, R. (2012). Relevance in cooperation and conflict. *Journal of Logic and Computation*, *22*(1), 23–54.

Fraundorf, S. H., & Watson, D. G. (2011). The disfluent discourse: Effects of filled pauses on recall. *Journal of memory and language*, *65*(2), 161–175.

Freedle, R., & Kostin, I. (1999). Does the text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language testing*, *16*(1), 2–32.

Futrell, R., & Gibson, E. (2017). L2 processing as noisy channel language comprehension. *Bilingualism: Language and Cognition*, *20*(4), 683–684.

Gershman, S., & Goodman, N. (2014). Amortized inference in probabilistic reasoning. *Proceedings of the annual meeting of the cognitive science society*, *36*(36).

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056.

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't underestimate the benefits of being misunderstood. *Psychological Science*, *28*(6), 703–712.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological review*, *103*(4), 650.

Goldman-Eisler, F. (1958). The predictability of words in context and the length of pauses in speech. *Language and Speech*, *1*(3), 226–231.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, *5*(1), 173–184.

Gorman, K., Olson, L., Hill, A. P., Lunsford, R., Heeman, P. A., & van Santen, J. P. (2016). Uh and um in children with autism spectrum disorders or language impairment. *Autism Research*, *9*(8), 854–865.

Gosselin, L., Martin, C. D., Martın, A. G., & Caffarra, S. (2022). When a nonnative accent lets you spot all the errors: Examining the syntactic interlanguage benefit. *Journal of Cognitive Neuroscience*, *34*(9), 1650–1669.

Gósy, M., Bóna, J., Beke, A., & Horváth, V. (2014). Phonetic characteristics of filled pauses: The effects of speakers' age. *Proceedings of the 10th International Seminar on Speech Production, ISSP 2014*, 150–153.

Gotzner, N., & Spalek, K. (2017). The connection between focus and implicatures: Investigating alternative activation under working memory load. *Linguistic and psycholinguistic approaches on implicatures and presuppositions*, 175–198.

Goupil, L., Ponsot, E., Richardson, D., Reyes, G., & Aucouturier, J.-J. (2021). Listeners' perceptions of the certainty and honesty of a speaker are associated with a common prosodic signature. *Nature communications*, *12*(1), 861.

Gregersen, T. S. (2005). Nonverbal cues: Clues to the detection of foreign language anxiety. *Foreign Language Annals*, *38*(3), 388–400.

Grey, S., Cosgrove, A. L., & van Hell, J. G. (2020). Faces with foreign accents: An event-related potential study of accented sentence comprehension. *Neuropsychologia*, *147*, 107575.

Grey, S., Schubel, L. C., McQueen, J. M., & Van Hell, J. G. (2019). Processing foreign-accented speech in a second language: Evidence from ERPs during sentence comprehension in bilinguals. *Bilingualism: Language and Cognition*, *22*(5), 912–929.

Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics*, *42*, 93–108.

Grice, H. P. (1975). Logic and conversation. In P. Coleman & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (pp. 41–58). New York: Academic Press.

Griffin, Z. M., & Huitema, J. (1999). Beckman spoken picture naming norms. *Unpublished Data Available for Download at Www. Langprod. Cogsci. Uiuc. Edu/norms.*

Griffiths, R. (1991). The paradox of comprehensible input: Hesitation phenomena in l2 teacher talk. *JALT Journal*, *13*(1), 23–41.

Grodner, D., & Sedivy, J. C. (2011). 10 the effect of speaker-specific information on pragmatic inferences. In *The processing and acquisition of reference* (pp. 239–272). MIT Press.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42–55.

Grondelaers, S., Speelman, D., Drieghe, D., Brysbaert, M., & Geeraerts, D. (2009). Introducing a new entity into discourse: Comprehension and production evidence for the status of dutch er "there" as a higher-level expectancy monitor. *Acta Psychologica*, *130*(2), 153–160.

Grosjean, F., & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français: Vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica*, *31*(3-4), 144–184.

Grundy, J. G. (2020). The effects of bilingualism on executive functions: An updated quantitative analysis. *Journal of Cultural Cognitive Science*, *4*(2), 177–199.

Hala, S., Chandler, M., & Fritz, A. S. (1991). Fledgling theories of mind: Deception as a marker of three-year-olds' understanding of false belief. *Child Development*, *62*(1), 83–97.

Hanuíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887.

Hanulíková, A., & Weber, A. (2011). Sink positive: Linguistic experience with th substitutions influences nonnative word recognition. *Attention, Perception, &amp Psychophysics*, *74*(3), 613–629.

Hanus, D. (2016). Causal reasoning versus associative learning: A useful dichotomy or a strawman battle in comparative psychology? *Journal of Comparative Psychology*, *130*(3), 241.

Hanzlíková, D., & Skarnitzl, R. (2017). Credibility of native and non-native speakers of english revisited: Do non-native listeners feel the same? *Research in Language*, *15*(3), 285–298.

Hartsuiker, R. J., & Kolk, H. H. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, *42*(2), 113–157.

Hartsuiker, R. J., & Notebaert, L. (2009). Lexical access problems lead to disfluencies in speech. *Experimental Psychology*, *57*, 169–77.

Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language*, *88*, 70–86.

Heller, D., Arnold, J. E., Klein, N., & Tanenhaus, M. K. (2015). Inferring difficulty: Flexibility in the real-time processing of disfluency. *Language and Speech*, *58*(2), 190–203.

Henningsen, D. D., Valde, K. S., & Davies, E. (2005). Exploring the effect of verbal and nonverbal cues on perceptions of deception. *Communication Quarterly*, *53*(3), 359–375.

Hirose, Y., & Mazuka, R. (2015). Predictive processing of novel compounds: Evidence from japanese. *Cognition*, *136*, 350–358.

Hopp, H. (2018). The bilingual mental lexicon in L2 sentence processing. *Second Language*, *17*, 5–27.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56.

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching*, *53*(3), 277–306.

Hopp, H. (2016). The timing of lexical and syntactic processes in second language sentence comprehension. *Applied Psycholinguistics*, *37*(5), 1253–1280.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology*, *58*(3), 376–415.

Hübscher, I., Esteve-Gibert, N., Igualada, A., & Prieto, P. (2017). Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Language*, *37*(1), 24–41.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135.

Huettig, F., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition*, *224*, 105050.

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? probably not. *Language, Cognition and Neuroscience*, *31*(1), 19–31.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171.

Hughes, V., Foulkes, P., & Wood, S. (2016). Strength of forensic voice comparison evidence from the acoustics of filled pauses. *International Journal of Speech, Language and the Law*, 99–132.

Huizeling, E., Peeters, D., & Hagoort, P. (2022). Prediction of upcoming speech under fluent and disfluent conditions: Eye tracking evidence from immersive virtual reality. *Language, Cognition and Neuroscience*, *37*(4), 481–508.

Ip, M. H. K., & Papafragou, A. (2022). Integrating non-native speaker identity in semantic and pragmatic processing. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44).

Ito, A., Corley, M., & Pickering, M. J. (2018). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition*, *21*(2), 251–264.

Ito, A., & Knoeferle, P. (2022). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, 1–33.

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). Why the a/an prediction effect may be hard to replicate: A rebuttal to delong, urbach, and kutas (2017). *Language, Cognition and Neuroscience*, *32*(8), 974–983.

Ito, A., & Pickering, M. (2021). Automaticity and prediction in non-native language comprehension. In E. Kaan & T. Grüter (Eds.), *Prediction in second-language processing and learning* (pp. 26–46). John Benjamins Publishing Company.

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of english: A visual world eye-tracking study. *Journal of Memory and Language*, *98*, 1–11.

Ivanova, I., Branigan, H. P., McLean, J., Costa, A., & Pickering, M. J. (2021). Lexical alignment to non-native speakers. *Dialogue and Discourse*, *12*(2), 145–173.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*(2), 277–288.

Jaeger, F. T., Furth, K., & Hilliard, C. (2012). Phonological overlap affects lexical selection during sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1439.

Jaeger, T. F. (2008). Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Jiang, X., Gossack-Keenan, K., & Pell, M. D. (2019). To believe or not to believe? How voice and accent information in speech alter listener impressions of trust. *Quarterly Journal of Experimental Psychology*, *73*(1), 55–79.

Jiang, X., & Pell, M. D. (2017). The sound of confidence and doubt. *Speech Communication*, *88*, 106–126.

Johnston, R. A., Dent, K., Humphreys, G. W., & Barry, C. (2010). British-english norms and naming times for a set of 539 pictures: The role of age of acquisition. *Behavior Research Methods*, *42*(2), 461–469.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*(2), 257–282.

Kamide, Y., Scheepers, C., & Altmann, G. T. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from german and english. *Journal of Psycholinguistic Research*, *32*, 37–55.

Kasl, S. V., & Mahl, G. F. (1965). Relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology*, *1*(5), 425.

Kendall, T. (2013). *Speech rate, pause and sociolinguistic variation: Studies in corpus sociophonetics*. Springer.

Kidd, C., White, K. S., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science*, *14*(4), 925–934.

King, J. P., Loy, J. E., & Corley, M. (2018). Contextual effects on online pragmatic inferences of deception. *Discourse Processes*, *55*(2), 123–135.

Kirjavainen, M., Crible, L., & Beeching, K. (2022). Can filled pauses be represented as linguistic items? investigating the effect of exposure on the perception and production of um. *Language and Speech*, *65*(2), 263–289.

Kjellmer, G. (2003). Hesitation. in defence of er and erm. *English Studies*, *84*(2), 170–198.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127.

Kochari, A. R., & Flecken, M. (2019). Lexical prediction in language comprehension: A replication study of grammatical gender effects in dutch. *Language, Cognition and Neuroscience*, *34*(2), 239–253.

Kosmala, L., & Crible, L. (2022). The dual status of filled pauses: Evidence from genre, proficiency and co-occurrence. *Language and Speech*, *65*(1), 216–239.

Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech*, *48*(1), 29–53.

Kraut, R. (1980). Humans as lie detectors. *Journal of Communication*, *30*(4), 209–218.

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

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. (2014a). Rapid adaptation in online pragmatic interpretation of contrastive prosody. *Proceedings of the annual meeting of the cognitive science society*, *36*(36).

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014b). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, *133*(2), 335–342.

Kurumada, C., Brown, M., & Tanenhaus, M. (2012). Pragmatic interpretation of contrastive prosody: It looks like speech adaptation. *Proceedings of the annual meeting of the cognitive science society*, *34*(34).

Kurumada, C., & Clark, E. V. (2017). Pragmatic inferences in context: Learning to interpret contrastive prosody. *Journal of Child Language*, *44*(4), 850–880.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford University Press.

Lake, J. K. (2010). *The uses of conversational speech in measuring language performance and predicting behavioural and emotional problems* (Doctoral dissertation).

Lau, E. F., & Ferreira, F. (2005). Lingering effects of disfluent material on comprehension of garden path sentences. *Language and Cognitive Processes*, *20*(5), 633–666.

Leach, A.-M., Silva, C. S. D., Connors, C. J., Vrantsidis, M. R. T., Meissner, C. A., & Kassin, S. M. (2019). Looks like a liar? beliefs about native and non-native speakers' deception. *Applied Cognitive Psychology*, *34*(2), 387–396.

Leach, A.-M., Snellings, R. L., & Gazaille, M. (2017). Observers' language proficiencies and the detection of non-native speakers' deception. *Applied Cognitive Psychology*, *31*(2), 247–257.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of english. *Behavior research methods*, *44*, 325–343.

Leonard, C. M., Järvikivi, J., Porretta, V., & Langevin, M. (2016). Processing of stuttered speech by fluent listeners. *Speech Prosody*, 1216–1220.

Lev-Ari, S. (2015). Comprehending non-native speakers: Theory and evidence for adjustment in manner of processing. *Frontiers in Psychology*, *5*, 1546.

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096.

Lev-Ari, S., van Heugten, M., & Peperkamp, S. (2017). Relative difficulty of understanding foreign accents as a marker of proficiency. *Cognitive science*, *41*(4), 1106–1118.

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

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

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.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464.

Li, W., Rohde, H., & Corley, M. (2022). Veritable untruths: Autistic traits and the processing of deception. *Journal of Autism and Developmental Disorders*, 1–10.

Libersky, E., Neveu, A., & Kaushanskaya, M. (2022). One fish, uh, two fish: Effects of fluency and bilingualism on adults' novel word learning. *Psychonomic Bulletin & Review*, 1–11.

Lickley, R. J., & Bard, E. G. (1998). When can listeners detect disfluency in spontaneous speech? *Language and speech*, *41*(2), 203–226.

Lin, Y. (2016). Processing of scalar inferences by mandarin learners of english: An online measure. *PLoS One*, *11*(1), e0145494.

Lindemann, S. (2005). Who speaks "broken english"? us undergraduates' perceptions of non-native english 1. *International Journal of Applied Linguistics*, *15*(2), 187–212.

Lindsay, J., & O'Connell, D. C. (1995). How do transcribers deal with audio recordings of spoken discourse? *Journal of Psycholinguistic Research*, *24*, 101–115.

Long, D. L., & Lea, R. B. (2005). Have we been searching for meaning in all the wrong places? Defining the "search after meaning" principle in comprehension. *Discourse processes*, *39*(2-3), 279–298.

Lorenzoni, A., Pagliarini, E., Vespignani, F., & Navarrete, E. (2022). Pragmatic and knowledge range lenience towards foreigners. *Acta Psychologica*, *226*, 103572.

Loy, J. E., Rohde, H., & Corley, M. (2017). Effects of disfluency in online interpretation of deception. *Cognitive Science*, *41*, 1434–1456.

Loy, J. E., Rohde, H., & Corley, M. (2018). Cues to lying may be deceptive: Speaker and listener behaviour in an interactive game of deception. *Journal of Cognition*, *1*(1).

Loy, J. E., Rohde, H., & Corley, M. (2019). Real-time social reasoning: The effect of disfluency on the meaning of some. *Journal of Cultural Cognitive Science*, *3*(2), 159–173.

Loy, J. E., & Smith, K. (2019). Syntactic adaptation depends on perceived linguistic knowledge: Native english speakers differentially adapt to native and non-native confederates in dialogue. *PsyArXiv*.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, *88*, 22–60.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703.

MacFarlane, H., Gorman, K., Ingham, R., Presmanes Hill, A., Papadakis, K., Kiss, G., & Van Santen, J. (2017). Quantitative analysis of disfluency in children with autism spectrum disorder or language impairment. *PLoS one*, *12*(3), e0173936.

MacGregor, L. J. (2008). *Disfluencies affect language comprehension: Evidence from event-related potentials and recognition memory* (Doctoral dissertation). The University of Edinburgh.

MacGregor, L. J., Corley, M., & Donaldson, D. I. (2009). Not all disfluencies are are equal: The effects of disfluent repetitions on language comprehension. *Brain and Language*, *111*(1), 36–45.

MacGregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: Disfluent silent pauses in speech have consequences for listeners. *Neuropsychologia*, *48*(14), 3982–3992.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous english speech. *Word*, *15*(1), 19–44.

Magnuson, J. S. (2019). Fixations in the visual world paradigm: Where, when, why? *Journal of Cultural Cognitive Science*, *3*(2), 113–139.

Mai, R., & Hoffmann, S. (2013). Accents in business communication: An integrative model and propositions for future research. *Journal of Consumer Psychology*, *24*(1), 137–158.

Mårback, S., Sjöberg, G., Schwarz, I. C., & EKlund, R. (2009). Uhm... What's going on? An EEG study on perception of filled pauses in spontaneous Swedish speech. *FONETIK 2009*, 92–95.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*(5417), 522–523.

Martin, J. G., & Strange, W. (1968). The perception of hesitation in spontaneous speech. *Perception & Psychophysics*, *3*(6), 427–438.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324.

Matin, E., Shao, K.-C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & psychophysics*, *53*, 372–380.

Mazzaggio, G., Panizza, D., & Surian, L. (2021). On the interpretation of scalar implicatures in first and second language. *Journal of Pragmatics*, *171*, 62–75.

Mazzarella, D., & Pouscoulous, N. (2021). Pragmatics and epistemic vigilance: A developmental perspective. *Mind & Language*, *36*(3), 355–376.

Mazzone, M. (2009). Pragmatics and cognition: Intentions and pattern recognition in context. *International Review of Pragmatics*, *1*(2), 321–347.

Mazzone, M. (2013). Automatic and controlled processes in pragmatics. *Perspectives on linguistic pragmatics*, 443–467.

Merlo, S., & Mansur, L. L. (2004). Descriptive discourse: Topic familiarity and disfluencies. *Journal of Communication Disorders*, *37*(6), 489–503.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, *66*(4), 545–567.

Mirman, D. (2017). *Growth curve analysis and visualization using R*. CRC press.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494.

Mishra, R. K., Olivers, C. N. L., & Huettig, F. (2013). Spoken language and the decision to move the eyes: To what extent are language-mediated eye movements automatic? *Progress in brain research*, *202*, 135–149.

Mitsugi, S. (2017). Syntactic prediction in L2 comprehension: Evidence from japanese adverbials. *Proceedings of the 41st Annual Boston University Conference on Language Development. volume*, *2*, 509–521.

Mitsugi, S., & Macwhinney, B. (2016). The use of case marking for predictive processing in second language japanese. *Bilingualism: Language and Cognition*, *19*(1), 19–35.

Mori, Y., & Pell, M. D. (2019). The look of (un) confidence: Visual markers for inferring speaker confidence in speech. *Frontiers in Communication*, *4*, 63.

Morin-Lessard, E., & Byers-Heinlein, K. (2019). Uh and euh signal novelty for monolinguals and bilinguals: Evidence from children and adults. *Journal of Child Language*, *46*(3), 522–545.

Mortensen, L., Meyer, A. S., & Humphreys, G. W. (2006). Age-related effects on speech production: A review. *Language and Cognitive Processes*, *21*(1-3), 238–290.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–97.

Nahari, G., & Ben-Shakhar, G. (2013). Primacy effect in credibility judgements: The vulnerability of verbal cues to biased interpretations. *Applied Cognitive Psychology*, *27*(2), 247–255.

Nakamura, C., Arai, M., & Mazuka, R. (2012). Immediate use of prosody and context in predicting a syntactic structure. *Cognition*, *125*(2), 317–323.

Nakamura, C., Harris, J. A., & Jun, S.-A. (2022). Integrating prosody in anticipatory language processing: How listeners adapt to unconventional prosodic cues. *Language, Cognition and Neuroscience*, *37*(5), 624–647.

Nicenboim, B., Vasishth, S., & Rösler, F. (2020). Are words pre-activated probabilistically during sentence comprehension? Evidence from new data and a Bayesian random-effects meta-analysis using publicly available data. *Neuropsychologia*, *142*, 107427.

Nieuwland, M. S. (2019). Do 'early'brain responses reveal word form prediction during language comprehension? a critical review. *Neuroscience & Biobehavioral Reviews*, *96*, 367–400.

Nooteboom, S. G. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In *Errors in linguistic performance: Slips of the tongue, ear, pen and hand/ed. by victoria a. fromkin* (pp. 87–95). Academic Press Inc.

Norton-Ford, J. D., & Hogan, D. R. (1980). Role of nonverbal behaviors in social judgments of peers' assertiveness. *Psychological Reports*, *46*(3_suppl), 1085–1086.

O'Connell, D. C., & Kowal, S. (2005). Uh and um revisited: Are they interjections for signaling delay? *Journal of Psycholinguistic Research*, *34*, 555–576.

Oomen, C. C., & Postma, A. (2001). Effects of divided attention on the production of filled pauses and repetitions. *Journal of speech, language, and hearing research : JSLHR*, *44*, 997–1004.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–241.

Orena, A., & White, K. S. (2015). I forget what that's called! children's online processing of disfluencies depends on speaker knowledge. *Child development*, *86*(6), 1701–1709.

O'Shaughnessy, D. (1992). Recognition of hesitations in spontaneous speech. *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, *1*, 521–524.

Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language*, *9*(1), 19–36.

Owens, S. J., & Graham, S. A. (2016). Thee, uhh disfluency effect in preschoolers: A cue to discourse status. *British Journal of Developmental Psychology*, *34*(3), 388–401.

Owens, S. J., Thacker, J. M., & Graham, S. A. (2018). Disfluencies signal reference to novel objects for adults but not children. *Journal of Child Language*, *45*(3), 581–609.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, *29*(4), 434–446.

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

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190.

Pickering, M. J., & Garrod, S. (2014). Neural integration of language production and comprehension. *Proceedings of the National Academy of Sciences*, *111*(43), 15291–15292.

Pinget, A. F., Bosker, H. R., Quené, H., & De Jong, N. H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, *31*(3), 349–365.

Pistono, A., & Hartsuiker, R. J. (2021). Eye-movements can help disentangle mechanisms underlying disfluency. *Language, Cognition and Neuroscience*, *36*(8), 1038–1055.

Pistono, A., & Hartsuiker, R. J. (2023). Can object identification difficulty be predicted based on disfluencies and eye-movements in connected speech? *PLoS One*, *18*(3), e0281589.

Pistono, A., Senoussi, M., & Hartsuiker, R. (2023). Disfluencies reflect a... hm... competition between response options: Evidence from a drift diffusion analysis. *PsyArXiv*.

Pivneva, I., Palmer, C., & Titone, D. (2012). Inhibitory control and L2 proficiency modulate bilingual language production: Evidence from spontaneous monologue and dialogue speech. *Frontiers in Psychology*, *3*.

Politzer-Ahles, S., & Fiorentino, R. (2013). The realization of scalar inferences: Context sensitivity without processing cost. *PLoS One*, *8*(5), e63943.

Politzer-Ahles, S., & Gwilliams, L. (2015). Involvement of prefrontal cortex in scalar implicatures: Evidence from magnetoencephalography. *Language, Cognition and Neuroscience*, *30*(7), 853–866.

Porretta, V., Buchanan, L., & Järvikivi, J. (2020). When processing costs impact predictive processing: The case of foreign-accented speech and accent experience. *Attention, Perception, & Psychophysics*, *82*(4), 1558–1565.

Porretta, V., Kyröläinen, A.-J., van Rij, J., & Järvikivi, J. (2018). Visual world paradigm data: From preprocessing to nonlinear time-course analysis. *Intelligent Decision Technologies 2017: Proceedings of the 9th KES International Conference on Intelligent Decision Technologies (KES-IDT 2017)–Part II 9*, 268–277.

Porretta, V., Tremblay, A., & Bolger, P. (2017). Got experience? pmn amplitudes to foreign-accented speech modulated by listener experience. *Journal of Neurolinguistics*, *44*, 54–67.

Posner, M. I. (1988). Structures and function of selective attention. In T. Boll & B. K. Bryant (Eds.), *Clinical neuropsychology and brain function: Research, measurement, and practice* (pp. 173–202). American Psychological Association.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.

Ram, K., & Wickham, H. (2018). *Wesanderson: A wes anderson palette generator* [R package version 0.3.6].

Rapoeye, K., Hartsuiker, R. J., & Pistono, A. (2022). Semantic interference affects speech production by increasing disfluencies, not errors. *PsyArXiv*.

Rasenberg, M., Rommers, J., & Van Bergen, G. (2020). Anticipating predictability: An erp investigation of expectation-managing discourse markers in dialogue comprehension. *Language, Cognition and Neuroscience*, *35*(1), 1–16.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1290.

Rohanian, M., Hough, J., & Purver, M. (2021). Multi-modal fusion with gating using audio, lexical and disfluency features for alzheimer's dementia recognition from spontaneous speech. *arXiv preprint arXiv:2106.09668*.

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

Romero-Rivas, C., Martin, C., & Costa, A. (2016). Foreign-accented speech modulates linguistic anticipatory processes. *Neuropsychologia*, *85*, 245–255.

Romero-Rivas, C., Corey, J. D., Garcia, X., Thierry, G., Martin, C. D., & Costa, A. (2017). World knowledge and novel information integration during L2 speech comprehension. *Bilingualism: Language and Cognition*, *20*(3), 576–587.

Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, *51*(3), 437–447.

Rose, R. L. (2017). Silent and filled pauses and speech planning in first and second language production. *TMH-QPSR*, *2017*(58), 10–13.

Rose, R. L. (1996). *The communicative value of filled pauses in spontaneous speech* (Doctoral dissertation). Univerty of Birmingham.

Roseano, P., González, M., Borràs-Comes, J., & Prieto, P. (2016). Communicating epistemic stance: How speech and gesture patterns reflect epistemicity and evidentiality. *Discourse Processes*, *53*(3), 135–174.

Rubio-Fernández, P. (2017). Why are bilinguals better than monolinguals at false-belief tasks? *Psychonomic Bulletin & Review*, *24*(3), 987–998.

Rubio-Fernández, P., & Glucksberg, S. (2012). Reasoning about other people's beliefs: Bilinguals have an advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 211.

Ryskin, R., Levy, R. P., & Fedorenko, E. (2020). Do domain-general executive resources play a role in linguistic prediction? Re-evaluation of the evidence and a path forward. *Neuropsychologia*, *136*, 107258.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Sanders, G., & de Bruin, A. (2022). Examining the difference in error detection when listening to native and non-native speakers. *Quarterly Journal of Experimental Psychology*, 174702182211355.

Sanford, A. J. S., Sanford, A. J., Molle, J., & Emmott, C. (2006). Shallow processing and attention capture in written and spoken discourse. *Discourse Processes*, *42*(2), 109–130.

Saryazdi, R., DeSantis, D., Johnson, E. K., & Chambers, C. G. (2021). The use of disfluency cues in spoken language processing: Insights from aging. *Psychology and Aging*, *36*(8), 928.

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.

Schachter, S., Rauscher, F., Christenfeld, N., & Crone, K. T. (1994). The vocabularies of academia. *Psychological Science*, *5*(1), 37–41.

Schiller, N. O., Boutonnet, B. P.-A., Kloots, M. L. S. D. H., Meelen, M., Ruijgrok, B., & Cheng, L. L.-S. (2020). (Not so) Great Expectations: Listening to foreign-accented speech reduces the brain's anticipatory processes. *Frontiers in Psychology*, *11*.

Schleef, E. (2019). The evaluation of unfilled pauses: Limits of the prestige, solidarity and dynamism dimensions. *Lingua*, *228*, 102707.

Schlenter, J. (2022). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition*, *26*(2), 253–267.

Schnadt, M. J., & Corley, M. (2006). The influence of lexical, conceptual and planning based factors on disfluency production. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *28*(28).

Schroeder, S. R. (2018). Do bilinguals have an advantage in theory of mind? a meta-analysis. *Frontiers in Communication*, *3*, 36.

Schulz, J. (2021). *Pragmatic competence and pragmatic tolerance in foreign language acquisition–revisiting the case of scalar implicatures* (Doctoral dissertation). University of Oxford.

Schüppert, A., Hilton, N. H., & Gooskens, C. (2015). Swedish is beautiful, danish is ugly? Investigating the link between language attitudes and spoken word recognition. *Linguistics*, *53*(2), 375–403.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learing. In J. F. Kroll & A. M. B. DeGroot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford University Press.

Serota, K. B., & Levine, T. R. (2014). A few prolific liars. *Journal of Language and Social Psychology*, *34*(2), 138–157.

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in america: Three studies of self-reported lies. *Human Communication Research*, *36*(1), 2–25.

Shriberg, E. (1996). Disfluencies in switchboard. *Proceedings of International Conference on Spoken Language Processing*, *96*(1), 11–14.

Shriberg, E. (2001). To 'errrr'is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, *31*(1), 153–169.

Shriberg, E. E., Bates, R., & Stolcke, A. (1997). A prosody only decision-tree model for disfluency detection. *Fifth European Conference on Speech Communication and Technology*.

Shriberg, E. E., & Lickley, R. J. (1993). Intonation of clause-internal filled pauses. *Phonetica*, *50*(3), 172–179.

Silva, C. S. D., & Leach, A.-M. (2011). Detecting deception in second-language speakers. *Legal and Criminological Psychology*, *18*(1), 115–127.

Slabakova, R. (2010). Scalar implicatures in second language acquisition. *Lingua*, *120*(10), 2444–2462.

Smirnova, K., Korotaev, N., Panikratova, Y., Lebedeva, I., Pechenova, E., & Fedorova, O. (2020). Using the rupex multichannel corpus in a pilot fmri study on speech disfluencies. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 195–203.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*(1), 25–38.

Snape, N., & Hosoi, H. (2018). Acquisition of scalar implicatures: Evidence from adult japanese L2 learners of english. *Linguistic Approaches to Bilingualism*, *8*(2), 163–192.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*(1), 103–130.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & language*, *25*(4), 359–393.

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(4), 421–446.

Starr, G., & Cho, J. (2022). QUD sensitivity in the computation of scalar implicatures in second language acquisition. *Language Acquisition*, *29*(2), 182–197.

Stenstrom, A.-B. (2014). *An introduction to spoken interaction*. Routledge.

Stone, K., Lago, S., & Schad, D. J. (2021). Divergence point analyses of visual world data: Applications to bilingual research. *Bilingualism: Language and Cognition*, *24*(5), 833–841.

Suffill, E., Kutasi, T., Pickering, M. J., & Branigan, H. P. (2021). Lexical alignment is affected by addressee but not speaker nativeness. *Bilingualism: Language and Cognition*, *24*(4), 746–757.

Sussman, R. S., & Sedivy, J. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes*, *18*(2), 143–163.

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, *30*(4), 485–496.

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, *53*(1), 81–94.

Tal, S., Grossman, E., Rohde, H., & Arnon, I. (2023). Speakers use more redundant references with language learners: Evidence for communicatively-efficient referential choice. *Journal of Memory and Language*, *128*, 104378.

Tanenhaus, M. K. (2007). Eye movements and spoken language processing. In *Eye movements: A window on mind and brain* (pp. 443–469). Elsevier.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Team, G. D. R. (2006). A world of lies. *Journal of cross-cultural psychology*, *37*(1), 60–74.

Thacker, J. M., Chambers, C. G., & Graham, S. A. (2018a). Five-year-olds' and adults' use of paralinguistic cues to overcome referential uncertainty. *Frontiers in Psychology*, *9*, 143.

Thacker, J. M., Chambers, C. G., & Graham, S. A. (2018b). When it is apt to adapt: Flexible reasoning guides children's use of talker identity and disfluency cues. *Journal of Experimental Child Psychology*, *167*, 314–327.

Tiv, M., Deodato, F., Rouillard, V., Wiebe, S., & Titone, D. (2021). Second language experience impacts first language irony comprehension among bilingual adults. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *75*(2), 126.

Tolins, J., & Fox Tree, J. E. (2016). Overhearers use addressee backchannels in dialog comprehension. *Cognitive science*, *40*(6), 1412–1434.

Tomlinson, J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language*, *69*(1), 18–35.

Toth, A., Charest, M., van Rij, J., & Järvikivi, J. (2019). Applying the visual world paradigm in the investigation of preschoolers' online reference processing in a continuous discourse.

Tottie, G. (2011). Uh and um as sociolinguistic markers in British English. *International Journal of Corpus Linguistics*, *16*(2), 173–197.

Tottie, G. (2014). On the use of uh and um in american english. *Functions of Language*, *21*(1), 6–29.

Tottie, G. (2019). From pause to word: Uh, um and er in written American English. *English Language & Linguistics*, *23*(1), 105–130.

Tovar, A., Soler, A. G., Ruiz-Idiago, J., Viladrich, C. M., Pomarol-Clotet, E., Rosselló, J., & Hinzen, W. (2020). Language disintegration in spontaneous speech in huntington's disease: A more fine-grained analysis. *Journal of Communication Disorders*, *83*, 105970.

Trude, A. M., Tremblay, A., & Brown-Schmidt, S. (2013). Limitations on adaptation to foreign accents. *Journal of Memory and Language*, *69*(3), 349–367.

Trueswell, J. C. (2008). Using eye movements as a developmental measure within psycholinguistics. *Language Acquisition and Language Disorders*, *44*, 73.

Turner, J. C., Brown, R. J., & Tajfel, H. (1979). Social comparison and group interest in ingroup favouritism. *European Journal of Social Psychology*, *9*(2), 187–204.

Van Bergen, G., & Bosker, H. R. (2018). Linguistic expectation management in online discourse processing: An investigation of dutch inderdaad'indeed'and eigenlijk'actually'. *Journal of Memory and Language*, *103*, 191–209.

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.

Van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology Section A*, *58*(2), 251–273.

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

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2022). itsadug: Interpreting time series and autocorrelated data using gamms [R package version 2.4.1].

van Rij, J., Hollebrandse, B., & Hendriks, P. (2016). Children's eye gaze reveals their use of discourse context in object pronoun resolution. *Empirical perspectives on anaphora resolution*, *563*, 267–293.

Voss, B. (1979). Hesitation phenomena as sources of perceptual errors for non-native speakers. *Language and Speech*, *22*(2), 129–144.

Vrij, A. (2004). Why professionals fail to catch liars and how they can improve. *Legal and Criminological Psychology*, *9*(2), 159–181.

Vrij, A., & Baxter, M. (1999). Accuracy and confidence in detecting truths andlies in elaborations and denials: Truth bias, lie bias and individual differences. *Expert Evidence*, *7*, 25–36.

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, *24*, 239–263.

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and human behavior*, *32*, 253–265.

Walker, E. J., Risko, E. F., & Kingstone, A. (2014). Fillers as signals: Evidence from a question–answering paradigm. *Discourse Processes*, *51*(3), 264–286.

Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech communication*, *50*(2), 81–94.

Weber, A., & Broersma, M. (2012). Spoken word recognition in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 5368–5375). Wiley-Blackwell Bognor Regis.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25.

Weber, A., Di Betta, A. M., & McQueen, J. M. (2014). Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of phonetics*, *46*, 34–51.

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*(2), B63–B72.

Weldon, M. S., & Roediger, H. L. (1987). Altering retrieval demands reverses the picture superiority effect. *Memory &amp Cognition*, *15*(4), 269–280.

Wellwood, A., Pancheva, R., Hacquard, V., & Phillips, C. (2018). The anatomy of a comparative illusion. *Journal of Semantics*, *35*(3), 543–583.

White, K. S., Nilsen, E. S., Deglint, T., & Silva, J. (2020). That's thee, uuh blicket! how does disfluency affect children's word learning? *First Language*, *40*(1), 3–20.

Whitford, V., & Titone, D. (2012). Second-language experience modulates first-and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review*, *19*, 73–80.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between l1 and l2 speakers of english. *Journal of Phonetics*, *70*, 86–116.

Witteman, M. J., Bardhan, N. P., Weber, A., & McQueen, J. M. (2015). Automaticity and stability of adaptation to a foreign-accented speaker. *Language and speech*, *58*(2), 168–189.

Woensdregt, M., & Smith, K. (2017). Pragmatics and language evolution. In *Oxford research encyclopedia of linguistics*.

Wood, S. N. (2003). Thin-plate regression splines. *Journal of the Royal Statistical Society (B)*, *65*(1), 95–114.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, *99*(467), 673–686.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *73*(1), 3–36.

Wood, S. (2012). Mgcv: Mixed gam computation vehicle with gcv/aic/reml smoothness estimation.

Wood, S. N. (2006). On confidence intervals for generalized additive models based on penalized regression splines. *Australian & New Zealand Journal of Statistics*, *48*(4), 445–464.

Wood, S. (2017). *Generalized additive models: An introduction with r* (2nd ed.). Chapman; Hall/CRC.

Wood, S., N., Pya, & S"afken, B. (2016). Smoothing parameter and model selection for general smooth models (with discussion). *Journal of the American Statistical Association*, *111*, 1548–1575.

Woodland, J., & Voyer, D. (2011). Context and intonation in the perception of sarcasm. *Metaphor and Symbol*, *26*(3), 227–239.

Wu, Y. J., & Thierry, G. (2013). Fast modulation of executive function by language context in bilinguals. *Journal of Neuroscience*, *33*(33), 13533–13537.

Yan, S., Kuperberg, G. R., & Jaeger, T. F. (2017). Prediction (or not) during language processing. a commentary on Nieuwland et al.(2017) and DeLong et al.(2005). *BioRxiv*, 143750.

Yeo, A., & Alibali, M. W. (2017). Evidence for overt visual attention to hand gestures as a function of redundancy and speech disfluency. *CogSci: The annual meeting of the cognitive science society*.

Yildirim, I., Degen, J., Tanenhaus, M. K., & Jaeger, T. F. (2016). Talker-specificity and adaptation in quantifier interpretation. *Journal of Memory and Language*, *87*, 128–143.

Yoon, S. O., & Fisher, C. (2020). Children's attribution of disfluency to different sources. *CogSci: The annual meeting of the cognitive science society*.

Zhang, J., & Wu, Y. (2022). Epistemic reasoning in pragmatic inferencing by non-native speakers: The case of scalar implicatures. *Second Language Research*, 02676583211069735.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. *14*, 1–59.

Zuckerman, M., Koestner, R., & Driver, R. (1981). Beliefs about cues associated with deception. *Journal of Nonverbal Behavior*, *6*, 105–114.

Appendices

Appendix A

Supplementary materials Experiments 1 and 2

A.1 Experiment 2 participants' linguistic profile

In Experiment 2, participants completed an adapted version of the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian et al., 2007), with additional questions on language use (de Bruin, 2019). Here we report participants' reported measures for English.

Table A.1

Participants' answers to the Language Experience and Proficiency Questionnaire (LEAP-Q). Self-reported ages of acquisition, fluency, communicative use and years spent learning English by speaker condition. Participants the how they acquired English, as well as its relative position out of all the languages they spoke.

	Native Speaker	Non-native Speaker
Age of Acquisition	7.48 (2.78)	6.89 (2.74)
Age of Fluency	15 (4.18)	16.5 (4.83)
Age of Communicative Use	14.3 (4.22)	15.3 (4.47)
Years Spent learning English (in years)	13.9 (5.3)	13.1 (4.71)
Mode of Acquisition	Formal instruction = 17 Interaction = 3 Mixture = 28	Formal instruction = 9 Mixture = 39
Order of English	Second language: 38 Third language: 9 Forth language: 1	Second language: 39 Third language: 9

Table A.2

Self-reported reported percentages of use of English (relative to the other languages spoken) in childhood, adolescence, and currently. The bottom panel depicts self-reported use of English (on a 10-point scale, 0: never, 10: always) across different contexts currently. Code-switch refers to participants' reporting of switching between English and any other language they speak.

	Native Speaker	Non-native Speaker
Percentage of use in childhood	7.01 (10.4)	5.39 (9.42)
Percentage of use in adolescence	25.1 (17.4)	19.3 (13.6)
Percentage of use currently	57.9 (21.5)	56.4 (18.7)
Interactions with family	0.63 (1.62)	0.73 (1.94)
Interactions with friends	7.15 (3.16)	7.02 (2.71)
Media	7.42 (2.37)	6.79 (3.17)
Music	7.67 (2.32)	7.04 (2.89)
Reading	8.65 (1.83)	8.52 (2.12)
Studying	9.97 (0.83)	9.06 (2.27)
Class	8.29 (2.42)	8.67 (2.32)
Flatmates	6.04 (4.24)	6.79 (4.09)
Code-switch	Yes = 42	Yes = 45
	No = 5	No = 2
	Rarely = 1	Rarely = 1

A.2 Visualization of eye-movements

A.2.1 Experiment 1: Visualization of raw probabilities in the integration time window

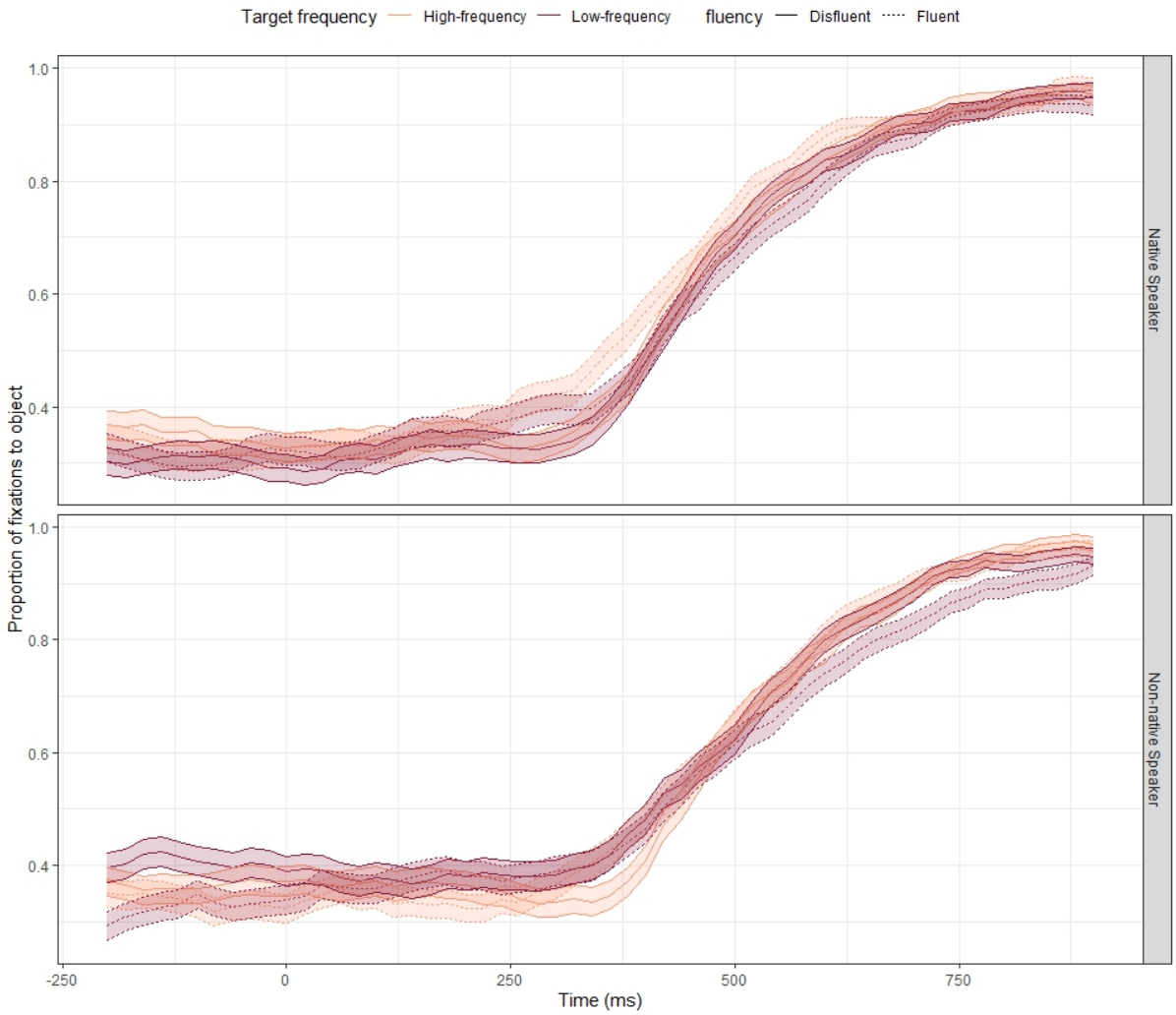

Figure A.1. Mean proportion of fixations on target (red: low-frequency, orange: high-frequency), by manner of delivery (dashed line: fluent, solid line: disfluent) and speaker's linguistic background (native/non-native), over time centred at noun onset. Proportions were calculated out of the sum of fixations towards either object for each 20-ms time bin from target onset to 760 ms post-target onset. Shaded areas represented ± 1 standard error of the mean.

A.2.2 Experiment 2: Visualization of raw probabilities in the prediction time window

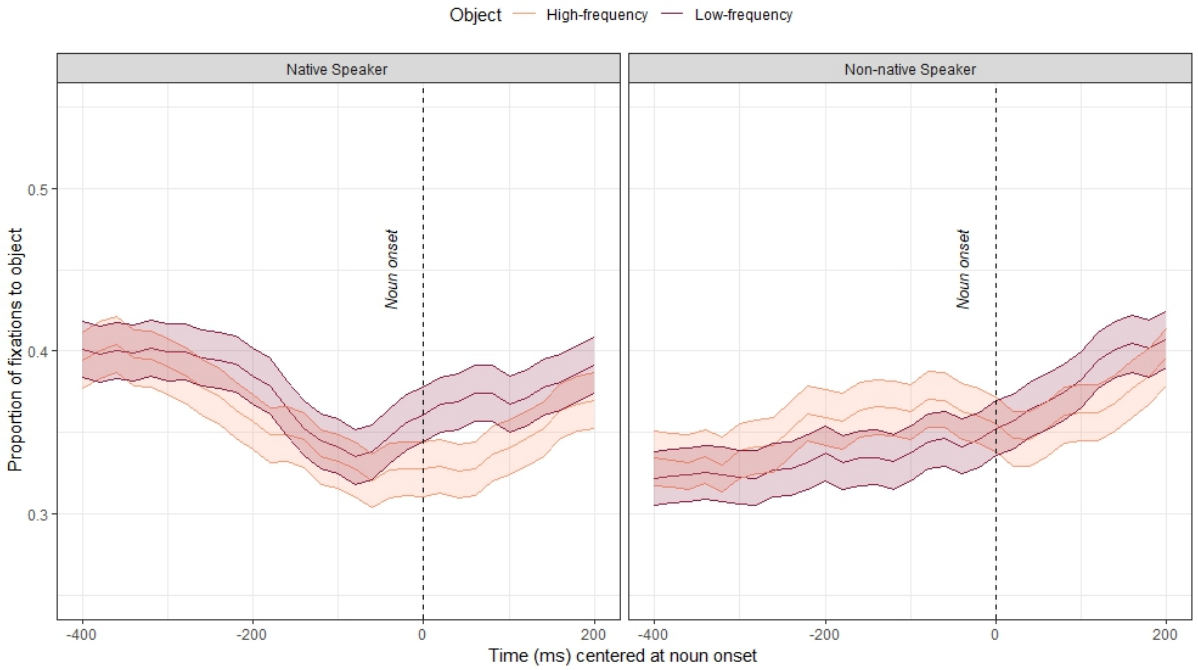

Figure A.2. Mean proportion of fixations to high- (orange line) and low-frequency (red line) items in fluent utterances by speaker's linguistic background (native/non-native) over time. Proportions were calculated out of the sum of fixations towards either object for each 20-ms time bin. Shaded areas represent ± 1 standard error of the mean.

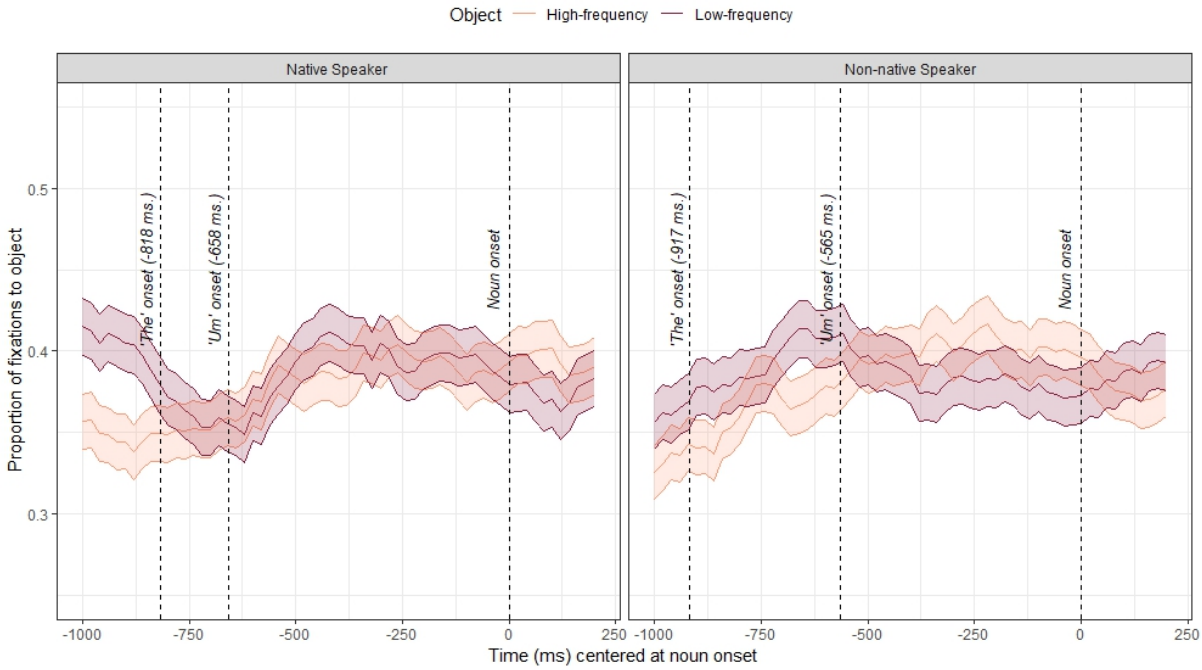

Figure A.3. Mean proportion of fixations to high- (orange line) and low-frequency (red line) items in disfluent utterances by speaker's linguistic background (native/non-native) over time. Proportions were calculated out of the sum of fixations towards either object

A.2.3 Experiment 2: Visualization of raw probabilities in the integration time window

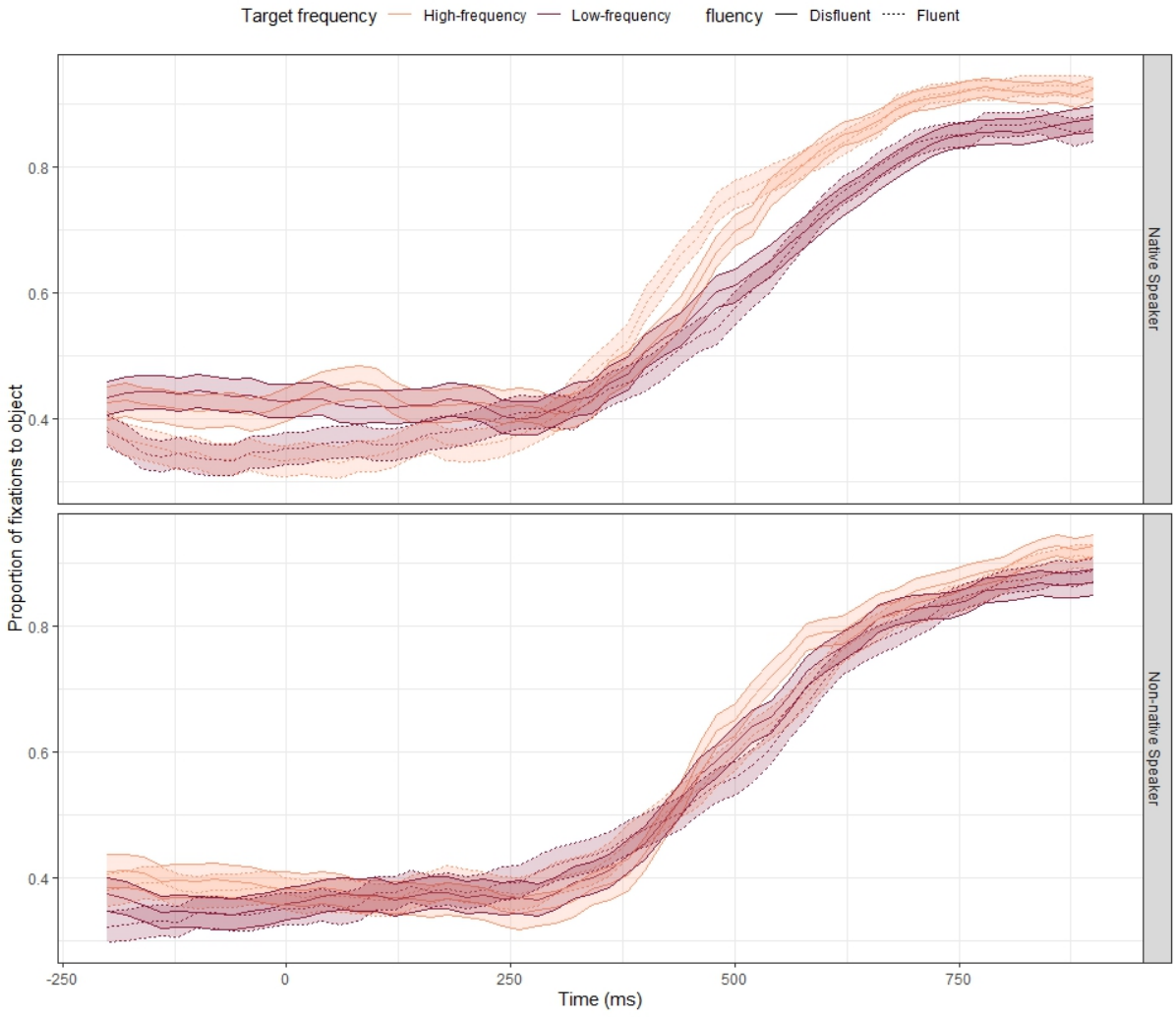

Figure A.4. Mean proportion of fixations on target (red: low-frequency, orange: high-frequency), by manner of delivery (dashed lines: fluent, solid lines: disfluent) and speaker's linguistic background (native/non-native), over time centred at noun onset. Proportions were calculated out of the sum of fixations towards either object for each 20-ms time bin from target onset to 760 ms post-target onset. Shaded areas represented ± 1 standard error of the mean.

A.3 Comparison between populations

A.3.1 Eye movements: Prediction time window

Table A.3 depicts the estimated parameters for two models (one per manner of delivery) exploring whether there were differences between participants in Experiment 1 and participants in Experiment 2 prior to target onset. The models' specifications are described in 4.4.1.

Table A.3

Estimated parameters of two mixed effects logistic regression models (one per manner of delivery) on the for the low-frequency target advantage, with fixed effects of speaker's linguistic background (native coded as 0, non-native as 1), experiment (Experiment 1 coded as 0, Experiment 2 as 1), a linear and a quadratic time components, and their interactions. Time window of analysis spanned from 'the' onset to 200 ms post-target onset.

	Estimate	Std.Error	t-value
Fluent utterances			
Intercept	-0.05	0.08	-0.53
Linear Time	0.02	0.09	0.23
Quadratic Time	-0.00006	0.0003	-0.22
Non-native Speaker	0.04	0.11	0.37
Non-native Speaker * Linear Time	0.003	0.12	0.02
Non-native Speaker * Quadratic Time	-0.000008	0.0003	-0.02
Experiment 2	0.11	0.08	1.32
Experiment 2 * Linear Time	-0.03	0.13	-0.23
Experiment 2 * Quadratic Time	0.00007	0.0004	0.20
Non-native Speaker * Experiment 2	-0.18	0.11	-1.64
Non-native Speaker * Experiment 2 * Linear Time	0.24	0.17	1.39
Non-native Speaker * Experiment 2 * Quadratic Time	-0.0006	0.0005	-1.16
Disfluent utterances			
Intercept	0.11	0.06	1.87
Linear Time	-0.12	0.02	-6.31
Quadratic Time	0.00009	0.00002	5.26
Non-native Speaker	-0.08	0.08	-0.98
Non-native Speaker * Linear Time	0.11	0.03	4.22
Non-native Speaker * Quadratic Time	-0.00007	0.00002	-3.09
Experiment 2	-0.09	0.06	-1.52
Experiment 2 * Linear Time	0.12	0.03	4.68
Experiment 2 * Quadratic Time	-0.0001	0.00003	-4.42
Non-native Speaker * Experiment 2	0.14	0.08	1.73
Non-native Speaker * Experiment 2 * Linear Time	-0.18	0.04	-5.12
Non-native Speaker * Experiment 2 * Quadratic Time	0.0001	0.00003	4.36

A.3.2 Eye movements: Word recognition

Table A.4 depicts the estimated parameters for two models (one per manner of delivery) exploring whether there were differences between participants in Experiment 1 and participants in Experiment 2 prior to target onset. The models' specifications are described in 4.4.2.

Table A.4

Estimated parameters of a mixed-effects model for target preference measured as empirical logits in a time window of analysis from target onset to 760 ms post-target onset, with speaker's linguistic background (native coded as 0, non-native coded as 1), target frequency (high-frequency coded as 0, low-frequency as 1), manner of delivery (fluent coded as 0, disfluent as 1), experiment (Experiment 1 coded as 0, Experiment 2 coded as 1), a linear and a quadratic time component, and their interactions.

	Estimate	Std.Error	t-value
Intercept	-0.14	0.07	-2.12
Linear Time	0.17	0.04	3.91
Quadratic Time	0.0005	0.0001	8.94
Low-frequency	0.02	0.07	0.30
Low-frequency * Linear Time	-0.20	0.06	-3.35
Low-frequency * Quadratic Time	0.0002	0.0001	3.12
Disfluent	-0.03	0.07	-0.36
Disfluent * Linear Time	-0.12	0.06	-1.98
Disfluent * Quadratic Time	0.0002	0.0001	2.65
Low-frequency * Disfluent	-0.12	0.07	-1.79
Low-frequency * Disfluent * Linear Time	0.29	0.08	3.40
Low-frequency * Disfluent * Quadratic Time	-0.0003	0.0001	-3.20
Non-native Speaker	0.15	0.08	1.85
Non-native Speaker * Linear Time	-0.37	0.06	-6.10
Non-native Speaker * Quadratic Time	0.0004	0.0001	5.81
Low-frequency * Non-native Speaker	0.02	0.07	0.27
Low-frequency * Non-native Speaker * Linear Time	0.20	0.08	2.41
Low-frequency * Non-native Speaker * Quadratic Time	-0.0004	0.0001	-3.38
Disfluent * Non-native Speaker	-0.004	0.10	-0.04
Disfluent * Non-native Speaker * Linear Time	0.04	0.08	0.53
Disfluent * Non-native Speaker * Quadratic Time	-0.0001	0.0001	-0.99
Low-frequency * Disfluent * Non-native Speaker	0.20	0.10	2.04
Low-frequency * Disfluent * Non-native Speaker * Linear Time	-0.17	0.12	-1.43
Low-frequency * Disfluent * Non-native Speaker * Quadratic Time	0.0002	0.0002	1.49
Experiment 2	-0.08	0.06	-1.24
Linear Time * Experiment 2	-0.06	0.06	-0.99
Quadratic Time * Experiment 2	0.0001	0.0001	1.28
Low-frequency * Experiment 2	0.33	0.07	4.75
Low-frequency * Experiment 2 * Linear Time	-0.21	0.08	-2.53
Low-frequency * Experiment 2 * Quadratic Time	0.0001	0.0001	0.68
Disfluent * Experiment 2	0.42	0.07	6.05
Disfluent * Experiment 2 * Linear Time	-0.20	0.08	-2.32
Disfluent * Experiment 2 * Quadratic Time	0.0001	0.0001	0.77
Low-frequency * Disfluent * Experiment 2	-0.22	0.10	-2.23
Low-frequency * Disfluent * Experiment 2 * Linear Time	0.14	0.12	1.15
Low-frequency * Disfluent * Experiment 2 * Quadratic Time	-0.0001	0.0002	-0.73
Non-native Speaker * Experiment 2	0.12	0.09	1.39
Non-native Speaker * Experiment 2 * Linear Time	0.01	0.08	0.09
Non-native Speaker * Experiment 2 * Quadratic Time	-0.0002	0.0001	-1.45
Low-frequency * Non-native Speaker * Experiment 2	-0.37	0.10	-3.82
Low-frequency * Non-native Speaker * Experiment 2 * Linear Time	0.22	0.12	1.86
Low-frequency * Non-native Speaker * Experiment 2 * Quadratic Time	-0.00002	0.0002	-0.13
Disfluent * Non-native Speaker * Experiment 2	-0.47	0.10	-4.80
Disfluent * Non-native Speaker * Experiment 2 * Linear Time	0.36	0.12	3.04
Disfluent * Non-native Speaker * Experiment 2 * Quadratic Time	-0.0002	0.0002	-1.57
Low-frequency * Disfluent * Non-native Speaker * Experiment 2	0.12	0.14	0.84
Low-frequency * Disfluent * Non-native Speaker * Experiment 2 * Linear Time	-0.16	0.17	-0.93
Low-frequency * Disfluent * Non-native Speaker * Experiment 2 * Quadratic Time	0.0001	0.0002	0.57

Appendix B

Supplementary materials Experiments 3 and 4

B.1 Experiment 3: Visualization of the referent disadvantage

In Experiment 3, we modelled the referent disadvantage via empirical logits. The figured below depict this pattern for fluent and disfluent utterances.

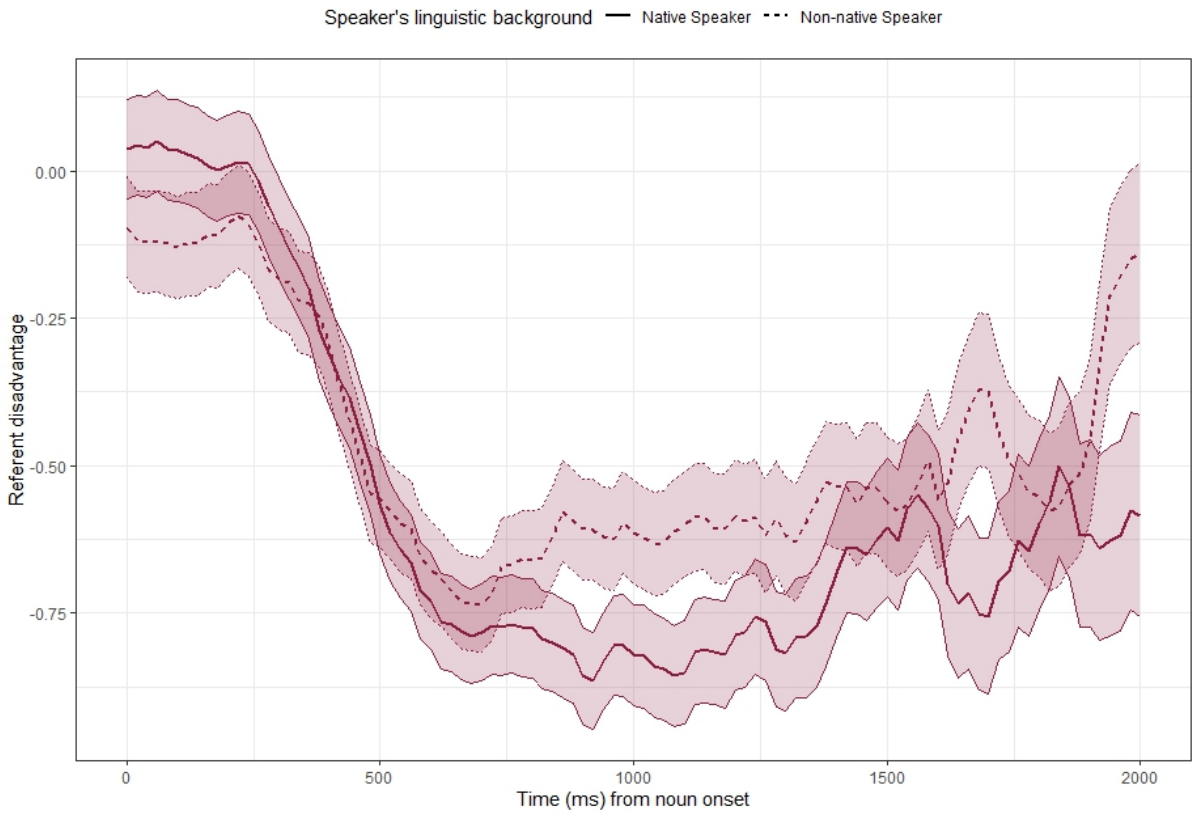

Figure B.1. Fluent utterances: Mean referent disadvantage by speaker's linguistic background (native/non-native). The referent disadvantage was calculated in empirical logits, where positive values indicate a preference to fixate on the distractor, and negative values index a preference to fixate on the referent. Shaded areas represent ± 1 standard error of the mean.

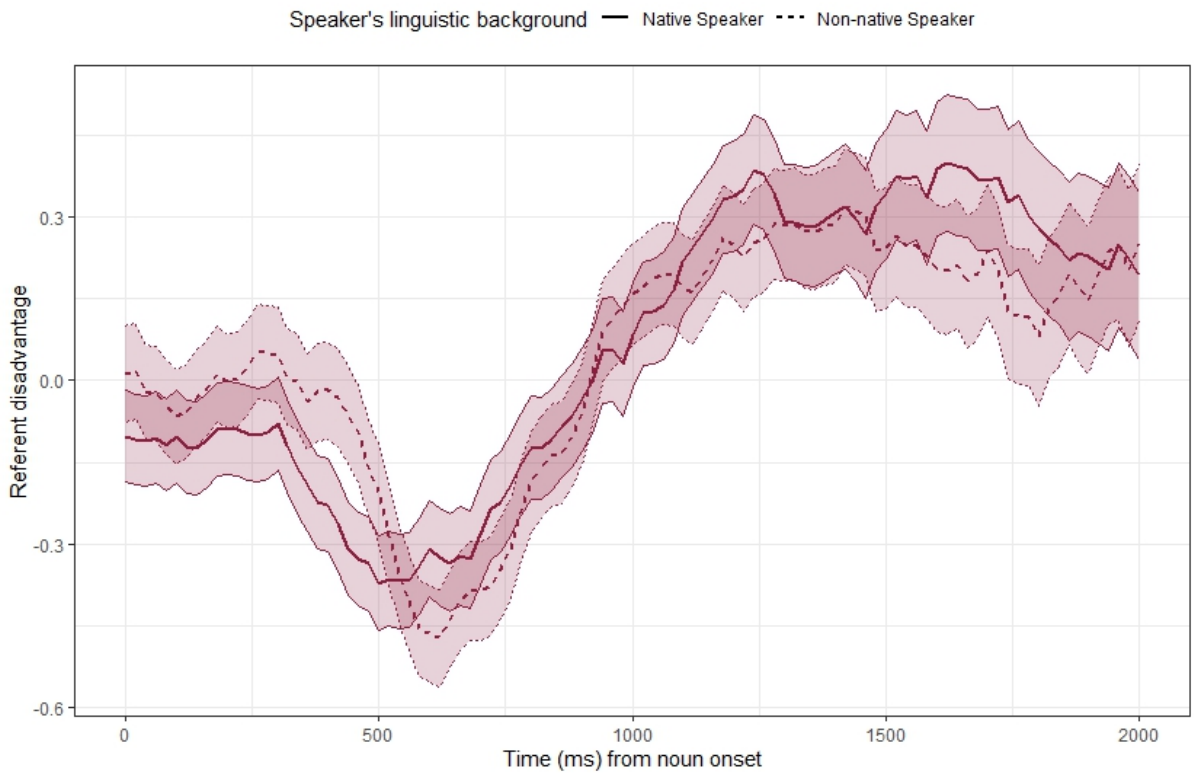

Figure B.2. Disfluent utterances: Mean referent disadvantage by speaker's linguistic background (native/non-native). The referent disadvantage was calculated in empirical logits, where positive values indicate a preference to fixate on the distractor, and negative values index a preference to fixate on the referent. Shaded areas represent ± 1 standard error of the mean.

B.2 Experiment 4 participants' linguistic profile

In Experiment 4, participants completed an adapted version of the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian et al., 2007), which included additional questions on language use (de Bruin, 2019). They also completed five questions from Bachman and Palmer (1989) wherein participants reported their abilities and difficulties to use English for communicative purposes. All the data, as well as questionnaires, can be found at https://osf.io/zfjtw/. Here we report participants' reported measures for English.

Table B.1

Self-reported ages of acquisition, fluency, communicative use and years spent learning English by speaker condition. Participants the how they acquired English, as well as its relative position out of all the languages they spoke.

	Native Speaker	Non-native Speaker
Age of Acquisition (in years)	6.43 (2.46)	7.12 (2.39)
Age of Fluency	14.36 (4.10)	14.75 (3.91)
Age of Communicative Use	13.82 (4.40)	14.69 (5.28)
Years Spent learning English	14.5 (5.28)	14.38 (4.01)
Mode of Acquisition	Formal instruction = 5 Mixture = 20	Formal instruction = 11 Mixture = 20 Other = 1
Order of English	Second language: 20 Third language: 7 Forth language: 1	Second language: 39 Third language: 9

Table B.2

Self-reported reported percentages of use of English (relative to the other languages spoken) in childhood, adolescence, and currently. The bottom panel depicts self-reported use of English (on a 10-point scale, 0: never, 10: always) across different contexts currently.

	Native Speaker	Non-native Speaker
Percentage Use Childhood	10.36 (13.63)	6.19 (7.66)
Percentage Use Adolescence	35.46 (27.8)	22.12 (20.47)
Percentage Use Currently	72.57 (17.61)	54 (19.74)
Interactions with family	0.37 (0.63)	1.38 (2.54)
Interactions with friends	8.56 (1.65)	6.94 (2.95)
Media	7.93 (2.79)	7 (2.87)
Music	8.22 (2.14)	6.94 (2.59)
Reading	8.89 (1.25)	7.75 (1.97)
Studying	8.74 (2.65)	9.12 (1.29)
Class	7.59 (3.59)	7.91 (2.36)
Flatmates	7 (3.75)	5.5 (3.89)

B.3 Experiment 4: Visualization of the referent disadvantage

In Experiment 4, we modelled the referent disadvantage via empirical logits. The figured below depict this pattern for fluent and disfluent utterances.

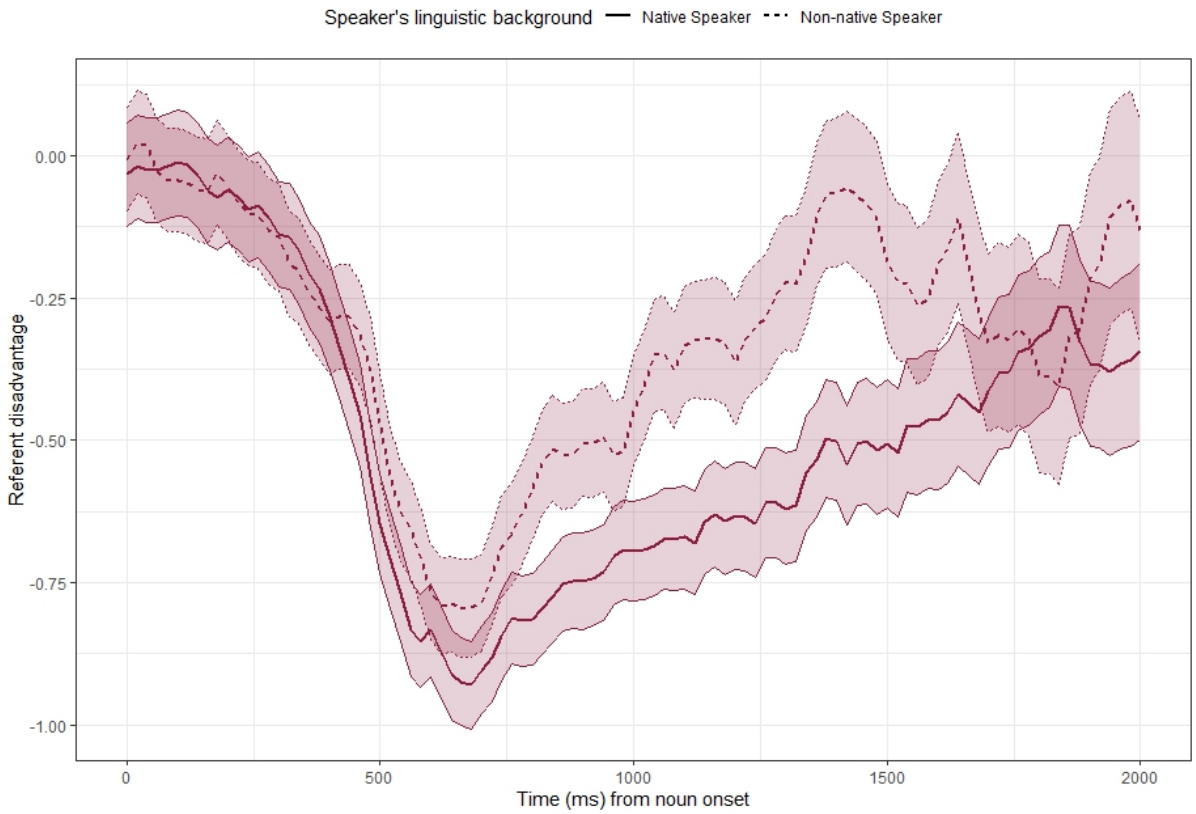

Figure B.3. Fluent utterances: Mean referent disadvantage by speaker's linguistic background (native/non-native). The referent disadvantage was calculated in empirical logits, where positive values indicate a preference to fixate on the distractor, and negative values index a preference to fixate on the referent. Shaded areas represent ± 1 standard error of the mean.

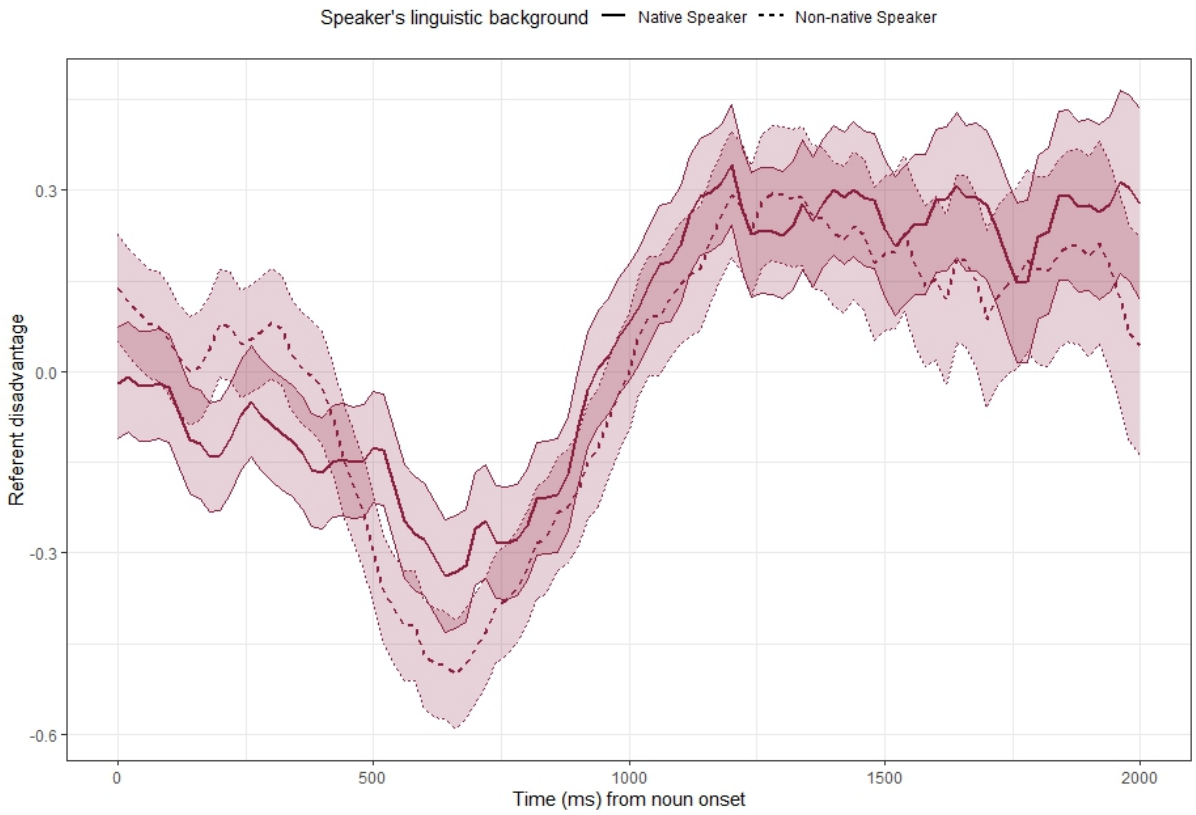

Figure B.4. Disfluent utterances: Mean referent disadvantage by speaker's linguistic background (native/non-native). The referent disadvantage was calculated in empirical logits, where positive values indicate a preference to fixate on the distractor, and negative values index a preference to fixate on the referent. Shaded areas represent ± 1 standard error of the mean.

B.4 Generalised Additive Mixed Models

As eye-tracking data is almost always non-linear (Porretta et al., 2018) and autocorrelated, gaze preference for the referent was modeled using Generalized Additive Mixed Models (GAMMs). GAMMs allow us to model non-linear trends over time, via smooth functions, i.e., a weighted sum of a set of base functions, each having a different shape (van Rij et al., 2016; Toth et al., 2019). Importantly, we wanted to model both the difference between the fluent and disfluent conditions for each speaker and assess whether these differential patterns were distinguishable from each other: That is, if our listeners

significantly differ in how they interpret fluent versus disfluent utterances from native and non-native speakers.

GAMMs allow us to model non-linear trends over time, via smooth functions, i.e., a weighted sum of a set of base functions, each having a different shape (van Rij et al., 2016; Toth et al., 2019). GAMMs present additional advantages over more traditional analyses of eye-tracking data. Firstly, GAMMs can account for the characteristic non-linearity of eye-tracking data (Porretta et al., 2018). Additionally, GAMMs can account for autocorrelations in data. Given the nature of eye movements, wherein the position of a fixation at time t depends on its position at time t-1, including this autocorrelative process in our model protects against anti-conservative p-values (De Cat et al., 2015). GAMMs also allow us to directly determine the model's random effects structure using significance testing, alleviating the need for model comparison (Wieling, 2018).

In order to understand how manner of delivery and speaker's linguistic background impact the timecourse of fixations over time, we followed the analysis of model 9b.bin in Wieling (2018). Our model included ordered factor difference smooths to account for constant differences via parametric effects and separate them from the smooth differences.

Ordered factor difference smooths allow us to investigate how the difference in empirical logits over the two displayed items (i.e., the referent and the distractor, wherein positive values represent preference over the distractor, and negative values represent preference over the referent) differs between conditions over time. These smooths do so by fitting K-1 centred smooths (where k are the levels of the factor): each smooth models the difference between the smooth for the reference level and the K-th level of a given factor (King et al. 2020). Importantly, we wanted to model both the difference between the fluent and disfluent conditions for each speaker and assess whether these differential patterns were distinguishable from each other: i.e., if our listeners significantly differ in how they interpret fluent versus disfluent utterances from native and non-native speakers. Therefore, we conducted two models (one for Experiment 3 and another for Experiment 4) to assess time differences in eye movements for each population.

Following Wieling (2018), each model included two reference levels for each speaker: one constant effect modeled by a parametric effect, and one thin-plate smooth. We included two ordered factor difference smooths (using thin-plate regressions) to separate any constant and non-linear differences. One ordered difference smooth modelled the difference between fluent-disfluent contrast for the native speaker, while the second ordered difference smooth represented the difference between the native and non-native fluent-disfluent contrast. These two ordered factor difference smooths can be summed for each condition (i.e., speaker's linguistic background and manner of delivery) to obtain the non-native fluent-disfluent contrast. For a more in-depth discussion of how the smooths in this model can be summed to obtain the smooth per each condition, we refer to Wieling (2018).

By-participant and by-item random intercepts and slopes can be modelled via shrunk factor smooths as non-linear equivalents. Our model included non-linear random smooths for participants by fluency over time, and non-linear random smooths for items by nativeness and fluency separately over time. Finally, the model included an AR(1) error model for the residuals. Models were fit with the *mgcv* package (version 1.8.33; Wood, 2004, 2017), and visualizations with the *itsadug* package (version 2.4.; van Rij et al., 2020).

B.4.1 Model results

Table B.3 reports the parametric coefficients (i.e., constant changes) and smooth terms (i.e., non-linear changes) for a Generalised Additive Mixed Model in Experiment 3. The model showed a constant linear increase in the referent disadvantage (i.e., a preference to fixate on the distractor) when the native speaker was disfluent (β = 0.15, SE = 0.04, t = 3.57, p <.001), a pattern that did not differ for those listening to a non-native speaker being disfluent (β = 0.005, SE = 0.06, t = 0.08, p-value = 0.94). The difference between fluent and disfluent for the native speaker was significant over time and non-linear (edf = 1.66, F = 7.60, p < .001). This contrast, however, did not differ from that of the non-native speaker (edf = 1.02, F = 3.47, p = 0.06).

Table B.3

Results of a Generalised Linear Mixed Model for the referent disadvantage, in a time window from target-onset to 800 ms post-target onset, for native listeners.

Parametric coefficients	*Estimate*	*Std.Error*	*t-value*	*p-value*
Intercept	-0.27	0.04	-6.46	<.001
Non-native Speaker	0.02	0.08	0.26	0.80
Native Speaker, Disfluent	0.15	0.04	3.57	<.001
Non-native Speaker, Disfluent	0.005	0.06	0.08	0.94
Smooth terms	*edf*	*Ref.df*	*F-value*	*p-value*
s(time):Native Speaker	5.90	6.67	9.50	<.001
s(time):Non-native Speaker	6.91	7.54	9.10	<.001
s(time):Native Speaker, Disfluent	1.66	1.87	7.60	<.001
s(time):Non-native Speaker, Disfluent	1.02	1.02	3.47	0.06
s(time,subject):Fluent	562.15	1150.00	1.03	<.001
s(time,subject):Disfluent	562.91	1150.00	1.04	<.001
s(time,item):Fluent	64.11	179.00	0.62	<.001
s(time,item):Disfluent	47.84	179.00	0.39	<.001
s(time,item):Native	53.98	179.00	0.79	<.001
s(time,item):Non-native	73.39	179.00	0.79	<.001

Table B.4 reports the results for non-native listeners (i.e., Experiment 4). Similar to Experiment 3, there was a constant linear increase in the referent disadvantage following disfluencies produced by a native speaker (β = 0.19, SE = 0.06, t = 2.96, p-value = 0.003), a pattern that did not differ for filled pauses produced by the non-native speaker (β = -0.04, SE = 0.09, t = -0.46, p-value = 0.64). The model likewise showed a significant non-linear difference between fluent and disfluent utterances for the native speaker (edf = 1.00, F = 5.00, p = 0.03), which did not differ for the non-native speaker (edf = 1.01, F = 1.92, p = 0.17).

Table B.4

Results of a Generalised Linear Mixed Model for the referent disadvantage, in a time window from target-onset to 800 ms post-target onset, for non-native listeners.

Parametric coefficients	*Estimate*	*Std.Error*	*t-value*	*p-value*
Intercept	-0.29	0.05	-5.46	<.001
Non-native Speaker	0.07	0.10	0.63	0.53
Native Speaker, Disfluent	0.19	0.06	2.96	0.003
Non-native Speaker, Disfluent	-0.04	0.09	-0.46	0.64
Smooth terms	*edf*	*Ref.df*	*F-value*	*p-value*
s(time):Native Speaker	4.88	5.68	5.30	<.001
s(time):Non-native Speaker	6.14	6.94	5.80	<.001
s(time):Native Speaker, Disfluent	1.00	1.01	5.00	0.03
s(time):Non-native Speaker, Disfluent	1.01	1.01	1.92	0.17
s(time,subject):Fluent	284.97	574.00	1.05	<.001
s(time,subject):Disfluent	269.90	574.00	0.95	<.001
s(time,item):Fluent	51.42	179.00	0.44	<.001
s(time,item):Disfluent	53.59	179.00	0.45	<.001
s(time,item):Native	39.03	179.00	0.43	<.001
s(time,item):sNon-native	39.74	179.00	0.29	<.001