COMPARING SYNTACTIC PERSISTENCE

IN WRITTEN AND SPOKEN MONOLOGUE

A thesis submitted in partial fulfilment of the requirements for the

Degree

of Master of Linguistics

in the Department of Linguistics

by Jennifer Middendorf

University of Canterbury

2017

Contents

Acknowledgements

First and foremost, my thanks to my supervisors, Heidi Quinn and Lynn Clark, who helped me turn my vague and inarticulate initial idea of "something about syntax maybe using the QuakeBox corpus?" into an actual topic, and guided me through every step of the research process. They were the best supervisory team I could imagine, providing unending support, encouragement, and confidence in me, even when my confidence in myself was flagging. Since the first undergraduate paper I took with Heidi, her encouragement and enthusiasm has pushed me to continue my academic journey further than I ever thought I would take it. As my Senior Supervisor, she was always available, whether to read a draft or to solve a last-minute administrative hiccough. Lynn has brought my understanding of statistics into the 21st century, introducing me to the wonders of mixed effects modelling (as well as, in her variation and theory class, convincing me of the exciting potential of corpus linguistics). Thank you both for your faith in me – it's been one of the hardest years of my life, but also the most rewarding.

I also want to acknowledge the support of my colleagues in the UC Arts Digital Lab. Lucy-Jane Walsh not only built the *Press* database which was such an invaluable source of data for this thesis, but also wrote queries for me, and helped me debug code. Lab directors Paul Millar and Christopher Thomson generously allowed me to take time away from my other duties to work on my research. Thank you too to Rosalee Jenkin for picking up the slack those times when my mind was less on web development or digital archiving and more on syntactic priming. I am also grateful to the College of Arts for financial support for both this thesis, and my undergraduate degree, through the Staff Fee Waiver programme.

Thank you also to Liam Walsh and Robert Fromont of the New Zealand Institute of Language, Brain and Behaviour. Liam taught me the mysteries of regular expressions, and patiently guided me through extracting the data I needed from QuakeBox. Robert was unfazed by my rather strange request for a character count in a time-aligned spoken corpus, and tweaked the LaBB-CAT tool so that I was able to measure textual distances.

The Linguistics Thesis group provided solidarity, inspiration, a place to ask stupid questions, and (most importantly) a regular supply of snacks. As well as the sharing of practical advice on everything from data modelling to presenting at a conference, being able to hear about the struggles and successes of other thesis students was invaluable. Thank you especially to Darcy Rose, who spent several hours with me as I struggled to make sense of graphing in R.

Sincere thanks must go to my friends and family, who supported and encouraged me through this long journey, listened patiently to my longwinded, over-excited descriptions of genitive structures, and were understanding when, juggling work and study, I couldn’t make time for them. Especial thanks to Megan for the emergency yum cha, to Lesley for refraining from tempting me with road trips, and to Gavin, Raina and Robert for games evenings at short notice when I needed a sanity break.

And finally, because I always promised I would:

To my mother, without whom.

Abstract

Syntactic persistence, the tendency for speakers to repeat recently-used syntactic structures, has been well demonstrated in dialogue and in single-sentence monologue primed by reading aloud pre-prepared material. Models advanced to explain syntactic persistence assume that priming will also occur in extended monologue, but there is no clear evidence that this is so.

This thesis examines within-speaker syntactic persistence of the genitive alternation in spoken and written monologue from the QuakeBox corpus and the *Press* database, two New Zealand corpora selected for their close match of time period, geographic location, and topic. Two research questions are considered: is priming present in extended monologue, and does priming differ between speech and writing?

In order to address these questions, I use binomial mixed-effect models to find the relative contribution of factors predicted to affect genitive choice and priming, and compare the relative impact of these factors, and the overall effect of priming, on the two corpora.

The findings of my research indicate that syntactic priming is present in extended monologue, and that this priming occurs more frequently in speech than in writing. My results also support observations in the existing literature that genitive choice is affected by animacy, the presence of a sibilant sound, and the semantic relationship between possessor and possessum.

While this study was not able to offer conclusive insights into the differences between α- and β-priming, and the issue of priming in nested structures, my findings indicate that these would be promising areas for further research.

1 Introduction

There are two main ways of expressing a possessive relationship within a noun phrase in English, either by inflection of the possessor (1), or via a prepositional phrase (2). These will be referred to as the s-genitive and the of-genitive respectively (cf. Rosenbach, 2002; Szmrecsanyi, 2006).

(1) I'll take her down to **a friend's place**

(QuakeBox NB893_KerryMcCammon_teacher)[1]

(2) these people were **victims of a very very extreme event**

(QuakeBox BR506_BlairAnderson)

Although the s-genitive and the of-genitive are often used to express the same semantic relationship, the two structures are not in free distribution. Rosenbach (2002, pp. 27-28) distinguishes "categorical contexts", where only one construction is available to express the proposition, from "choice contexts", where either construction is acceptable. For example, a quantitative such as (3a) can only be expressed using an of-genitive, and its s-genitive alternative (3b) is ungrammatical. Similarly, an s-genitive used as a classifier, such as (4a), cannot be expressed with an of-genitive in (4b), but instead alternates with the *for* prepositional phrase in (4c).

(3)

a. many of the teams

(Martin Van Beynen, 'EQC staff selected without interviews', *Press*, 9 December 2011)

b. * the teams' many

[1] All QuakeBox participants quoted in this thesis have consented to their transcripts being made publically available, and attributed to the name or pseudonym given. Full transcripts are available at https://quakestudies.canterbury.ac.nz/store/collection/235.

(4)

a. A Canterbury children's charity

(Georgina Stylianou, 'Record levels of need among children', *Press*, 1 September 2012)

b. ? A Canterbury charity of children

c. A Canterbury charity **for** children

Within choice contexts, a number of factors have been reported to affect the likelihood of a speaker choosing one construction over the other. For example, Altenberg (1982) found that s-genitives are more likely to be used with animate possessors and pre-modified possessums, whereas of-genitives are more likely to be used with inanimate possessors, post-modified possessors, and possessors which end in a sibilant sound.

Altenberg (1982, p. 288) also identified what he calls "structural parallelism", the recurrence of a previously used syntactic structure, as a factor influencing genitive alternation. He considers this parallelism to be a conscious stylistic choice, used, for example, to enhance rhetorical devices. Later research (e.g. Bock, 1986) took a psycholinguistic approach to the phenomenon, attributing the recurrence to syntactic priming - the activation of a syntactic structure persisting over time and making that structure more easily accessible for subsequent use. So, for example, when a speaker produces an utterance such as (5), *the nature of the damage* activates the of-genitive structure, making it more likely that when the speaker next wishes to express a possessive the same structure will be used, producing *the extent of it* rather than *its extent*.

(5) I suspect from **the nature of the damage** and **the extent of it** nothing will be the same again (QuakeBox BR506_BlairAnderson)

This thesis examines within-speaker syntactic persistence of the s-genitive and the of-genitive in spoken and written monologue from the QuakeBox corpus and the *Press* database, two New Zealand corpora selected for their close match of time period, geographic location, and topic. The findings of my research indicate that syntactic priming is present in extended monologue, and that this priming occurs more frequently in speech than in writing. The research results also provide further evidence that genitive choice is affected by animacy, the presence of a sibilant sound, and the semantic relationship between possessor and possessum.

Chapter 2 of this thesis discusses previous research on syntactic priming, and factors which have been found to influence genitive choice and the likelihood of syntactic priming. It also briefly discusses several theories which have been advanced to explain syntactic priming. Chapter 3 describes the methodology of a corpus study used to investigate and compare syntactic priming in spoken monologue and writing. The results of this study are presented and discussed in Chapter 4, and Chapter 5 discusses the broader implications of the results, and suggests areas for future research.

2 Literature Review

2.1 Priming and genitive choice

That syntactic structures can be primed has been well demonstrated in the experimental literature. Bock (1986) showed that participants described pictures using the same syntactic structure as they had read in a previous task, and crucially, that this priming is present even when the priming sentence and description used different words (for example, a prepositional dative using the preposition *to* would prime a prepositional dative using *for*), thus showing that the persistence was not merely due to lexical priming.

In order to show that it is the syntactic structure being primed, not the underlying semantic concepts, Bock and Loebell (1990) compared persistence of datives following priming sentences that contained either a prepositional dative (such as (6a)) or a prepositional locative (6b) – i.e. sentences which share a syntactic structure, but have different semantic roles. They found that both priming sentences equally increased the likelihood of the participant using a prepositional dative, meaning that it is indeed the syntactic structure that is primed.

(6)

a. The wealthy widow gave an old Mercedes to the church.

b. The wealthy widow drove an old Mercedes to the church.

(Bock & Loebell, 1990, p. 11)

In an experiment where the participant and a confederate (disguised as a second participant) were asked to describe pictures to each other, Branigan, Pickering, and Cleland (2000) confirmed that syntactic priming is distinct from both lexical and semantic priming, and that it is present in dialogue. When the confederate, who was following a script, used either a prepositional or double-object dative, the probability of the participant using the same structure was increased, showing that syntactic priming can operate between comprehension and production.

In a study of dative alternation in the British section of the International Corpus of English, Gries (2005) found a similar degree of priming to that found in experimental studies, demonstrating that corpus studies are a valid alternative to experimental studies of persistence, and moreover, he suggests, allow a much wider range of structures to be studied than would be possible with experimental studies alone. Examining five different alternating structures in the Corpus of Spoken American English and the Freiburg Corpus of English Dialects, Szmrecsanyi (2006) found a significant priming effect for each, although he notes the effect for all but one of the structures is weaker than that found experimentally for the dative by Bock (1986). Szmrecsanyi attributes this difference to the inability to account for all of the confounding factors that may be present, and concludes that it does not detract from the validity of corpus study for investigating persistence.

Some of these confounding factors influencing the strength of syntactic persistence are well-described in the literature. For example, priming is stronger in dative structures when both prime and target structure use the same verb (Branigan et al., 2000). Hartsuiker, Bernolet, Schoonbaert, Speybroeck, and Vanderelst (2008) refer to this effect as the "lexical boost" (p.215). The effect appears to mainly be dependent on the lemmas being identical (Pickering & Branigan, 1998), although identical morphology may also have a small effect (Gries, 2005; Szmrecsanyi, 2006). Cleland and Pickering (2003) showed a similar effect for nouns, with priming of relative clauses being more likely if the head of the noun phrase was repeated, or if the heads of the prime and target were semantically closely related.

Priming is more likely to occur when the target construction closely follows the prime, with strength of priming correlating to the log of the distance between prime and target (Gries, 2005; Szmrecsanyi, 2006). Hartsuiker et al. (2008) found that lexical boost decays rapidly, but also found evidence for syntactic priming being long lasting.

Szmrecsanyi (2006) also found that syntactic priming is weaker in more formal registers, and that priming is less likely when the linguistic environment is more complex, although this effect was not significant in all corpora he examined.

Szmrecsanyi (2006, p. 2) distinguishes between α-persistence, where use of a specific structure in a choice context makes it more likely that that exact structure will be used again in a choice context (as in (7)), and β-persistence, where a structure that is in a categorical context nevertheless primes a similar structure in a choice context. For example, in (8), although the quantitative *a lot of really bad things* is categorical, β-persistence of the of-genitive structure would increase the likelihood that the subsequent possessive would also be expressed by the of-genitive *the chimney of our house*, rather than the s-genitive *our house's chimney*.

(7) He says the unit will combine **the council's plan** with **Cera's ability** to make things happen.

(Sam Sachdeva, 'Isaacs - starter motor for rebuild', *Press*, 21 April 2012)

(8) you were expecting **a lot of really bad things** to happen ... **the chimney of our house** had fallen (QuakeBox EG864_SimonNewcombe)

Although Szmrecsanyi found a much stronger effect for α-persistence, he did find some β-persistence effect for the genitive, especially in more formal registers, and in fact found that any use of *of* could lexically prime the of-genitive, although this effect decays rapidly with textual distance.

Syntactic priming is of course not the only factor affecting genitive choice. In a study of the dative alternation, Bresnan, Cueni, Nikitina, and Baayen (2007) found that only 6% of the variation was due to priming, with the remainder being accounted for by other factors, so it would be expected that the situation would be similar for the genitive.

Animacy has a strong effect on genitive choice. The s-genitive is more likely to occur with animate possessors (such as humans, animals or human collectives) than with inanimate possessors (Altenberg, 1982; Grafmiller, 2014; Quirk, Greenbaum, Leech, & Svartvik, 1985; Rosenbach, 2005; Shih, Grafmiller, Futrell, & Bresnan, 2015; Wolk, Bresnan, Rosenbach, & Szmrecsanyi, 2013). The animacy of the possessum, however, appears to have no effect (Gries & Stefanowitsch, 2004). The effect of animacy appears to be particularly strong in New Zealand English. Hundt and Szmrecsanyi (2012) found that genitive choice in early New Zealand English is more strongly influenced by animacy than British English of the same period, while Bresnan and Hay (2008) found a similar effect for the dative alternation, with animacy affecting New Zealand English more than American English. The effect of animacy has been found to be stronger in speech than in writing (Grafmiller, 2014), and weaker in journalistic writing than in other genres (Dahl, 1971; Grafmiller, 2014).

As would be expected from the principle of end-weight, which generally attempts to put longer constituents at the end of sentences, long possessor phrases are more likely to be found with the of-genitive, and long possessum phrases with the s-genitive (Altenberg, 1982; Quirk et al., 1985; Rosenbach, 2005; Shih et al., 2015; Szmrecsanyi, 2006). Grafmiller (2014) found that an increase in the ratio of possessor to possessum length makes the s-genitive less likely. However, Wolk et al. (2013) found that the end-weight principle does not apply when both constituents are very short, and Szmrecsanyi (2006) found a tendency for both long possessors and long possessums to prefer the of-genitive.

According to Quirk et al. (1985), the principle of end-focus, which places more pertinent information at the end of a sentence, means that the of-genitive is more likely to be used when the possessor is new information. However, Szmrecsanyi (2006) found no main effect for givenness, and similarly no main effect was found by Hinrichs and Szmrecsanyi (2007), who suggested that the apparent effect is likely to be an epiphenomenon of other factors.

The semantic relationship between possessor and possessum has an effect on genitive choice, although there is debate as to how precisely the differences between relationships should be classified. Quirk et al. (1985) note that genitives that express objective relationships, such as in (9), where the relationship can be rephrased as a verb phrase with the possessor in the object position (as opposed to subjective relationships, where the possessor can be thought of as the subject), are more likely to be expressed using an of-genitive.

(9)

a. the imprisonment of the murderer

b. (Someone) imprisoned the murderer. (Quirk et al., 1985, p. 1278)

Stefanowitsch (2003) argues that the s-genitive generally expresses a possessor-possessee relationship, with the of-genitive expressing a part-whole relationship, while Rosenbach (2002, 2003) distinguishes between prototypical relationships (those expressing legal possession, kinship, body parts, and part-whole relations) and non-prototypical relations, and reports finding that prototypical relations are more likely to be expressed using the s-genitive.

An s-genitive is much less likely to be used when the possessor ends with a sibilant phoneme (Altenberg, 1982; Grafmiller, 2014; Shih et al., 2015; Szmrecsanyi, 2006). Shih et al. (2015) also found a small effect for rhythm, with speakers showing a preference, all else being equal, for a word order that enables a strong-weak pattern of syllables. Grafmiller (2014), however, failed to find a significant main effect for rhythm, suggesting that the effect of rhythm may be too weak to be readily detected.

Sociolinguistic factors appear to have some effect on genitive choice. Shih et al. (2015) found that of-genitives are favoured by older speakers, but found no significant correlation for gender. Jucker (1993) showed a correlation between level of formality and likelihood of the of-genitive, finding that, for example, s-genitives are used more often in the sports section of newspapers than in the news section. Szmrecsanyi (2010) showed that while variety and medium do not affect the direction of factors influencing genitive choice, they do affect the strength of those factors.

As would be expected, many interaction effects have been found between these various factors. To give just one example, Szmrecsanyi (2006) found a strong interaction effect between weight, animacy, and givenness.

2.2 Models of syntactic priming

A number of models have been proposed to explain syntactic priming. For example, Pickering and Branigan (1998) suggest that lexical entries are stored in a network of nodes, with each lexical item being connected to nodes representing its syntactic properties and features, and the syntactic structures it may appear in. When any of these nodes is activated, the activation spreads to the connected nodes, and, until the activation fully decays, makes those nodes more likely to be selected again. Cleland and Pickering (2003) add semantic information to this network.

Bock and Griffin (2000) found that priming effects are able to persist for a longer period of time than can be accounted for by an activation model, so propose that priming involves implicit learning mechanisms. When a speaker hears a particular syntactic structure used to convey a message, the association between that structure and the type of messages it can convey is strengthened. When the speaker wants to convey a similar message, they will therefore be more likely to use that structure. Bock and Griffin propose that implicit learning accounts for long-term priming, while short-term priming can be accounted for by activation. Chang, Dell, and Bock (2006) created a computer simulation of a model based on implicit learning, and were able to produce priming-like effects, but failed to reproduce the lexical boost effect.

Pickering and Garrod (2004) argue that successful dialogue relies on alignment of the participants' representations of the world. A speaker's representation of the world is unable to be accessed directly by his or her interlocutor, so in order to reach this alignment the speakers must first align their linguistic representations at a number of levels, so that they are in essence describing the world in the same way. Pickering and Garrod suggest that priming speeds up this process, by repeating structures (and thus levels of representation) that have already been used. Interconnections between the levels mean that priming at one level leads to priming at another, thus producing effects such as the lexical boost.

This idea of alignment is expanded on by models which explain priming in terms of reducing prediction errors (e.g. Jaeger & Snider, 2013; Pickering & Garrod, 2013). When a speaker comprehends an utterance by an interlocutor, a parallel process predicts the utterance, and thus allows the speaker to begin to formulate their response before the utterance has completed. In these prediction error models, the prediction is compared to the actual utterance, and the difference between the two (the prediction error) is used to adapt the speaker's prediction for the next utterance, so that prediction error is reduced. The speaker will then adapt their own production preferences in line with the model they have developed of their interlocutor's speech, in order to reduce the processing load of both comprehension and production.

Models based on learning, alignment and prediction all successfully explain why priming occurs in dialogue. However, it is less clear whether, in these models, priming would occur in extended monologue. Pickering and Garrod (2004) describe priming in monologue as being merely an epiphenomenon of priming in dialogue, arising from self-monitoring of speech.

The assumption is that priming will occur in extended monologue, but there is no clear evidence that this is so. Although Branigan (2007) suggests that priming effects will be found in all language production contexts, almost all studies of syntactic persistence in speech thus far have been based on dialogue, or on single-sentence monologue primed by reading aloud pre-prepared material (e.g. Bock, 1986; Kootstra, van Hell, & Dijkstra, 2010). Some studies have compared persistence when prime and target are uttered by the same speaker or by different speakers, with varying results. For example, Branigan et al. (2000) found a stronger persistence effect between speakers than had previously been found in single-speaker experiments, but Gries (2005) found that persistence of the dative is slightly stronger within-speaker than between speakers. Szmrecsanyi (2006) did not find any significant effect on the degree of persistence of genitive structures by prime and target being uttered by the same speaker, or within the same turn.

The primary aim of this study is therefore to discover whether priming effects are present in the extended monologues of the QuakeBox corpus. QuakeBox (Walsh et al., 2013) collected the stories of Canterbury residents following the series of devastating earthquakes which struck the region in 2010-2011. Participants told their stories to camera and the resulting recordings were transcribed, producing a corpus made up almost entirely of extended monologues.

A secondary aim of this study is to compare syntactic priming in spoken and written language. Whether there is any difference in priming between speech and writing is unclear. Does the increased time available when writing reduce processing demands, so that less repetition is needed, or do writers consciously repeat structures in order to sound more natural, as is suggested by Tannen (1987)? Hartsuiker and Westenberg (2000) found no difference between speech and writing in priming of the order the auxiliary verb and past participle in Dutch, and in an experiment using simulated computer chat, Hartsuiker et al. (2008) demonstrated that syntactic priming is present in written dialogue to the same degree as in spoken dialogue. However, when Szmrecsanyi (2006) compared priming in the spoken section of the British National Corpus with a sample of its written section, he found no evidence for priming in the written data, and suggested that “persistence is what makes a basic difference between spoken and written language” (p. 203). In a later study over a range of corpora Szmrecsanyi (2010) did find priming in writing, but to a much lesser extent than in speech.

Where differences have been found, they may represent an actual difference between speech and writing, or, given that the data in spoken corpora is almost always dialogue, they may instead reflect a difference between dialogue and monologue. Additionally, corpus studies comparing speech and writing often rely on corpora collected in different time periods (e.g. Grafmiller, 2014), so their results will be affected by any changes over time.

These shortcomings can be overcome by comparing the QuakeBox monologues with a corpus of written material from the *Christchurch Press*[2]. By comparing the QuakeBox monologues with *Press* articles written during the year in which QuakeBox was collecting stories, I am able to directly compare speech and writing from the same time period, the same geographical area, and covering the same general topic: the earthquakes and their effects on the region.

This study thus considers two research questions:

- Is priming present in extended monologue?
- Does priming differ between speech and writing?

In order to address these questions, I use binomial mixed-effect models to find the relative contribution of factors predicted to affect genitive choice and priming, and compare the relative impact of these factors, and the overall effect of priming, on the QuakeBox and *Press* corpora. This methodology will be discussed in detail in the following chapter.

[2] The *Christchurch Press* is the daily newspaper for the city of Christchurch and the surrounding Canterbury region.

3 Methodology

This chapter outlines the methodology used in this study. Section 3.1 outlines the corpora used as the data sources for the study, Section 3.2 describes the choices made in preparing and coding the data, and Section 3.3 describes the statistical methods use to model the data.

3.1 Data

Data for this study were sourced from the spoken monologues of the QuakeBox corpus and from a database of *Press* articles written during the same time period as the QuakeBox monologues were being collected. These two corpora allow a direct comparison of spoken and written monologues from the same time period and geographical location, and covering the same general topic.

3.1.1 QuakeBox corpus

QuakeBox (Walsh et al., 2013) is a collection of transcribed and time-aligned recordings made in Christchurch and Lyttelton, New Zealand, following the 2010-2011 Canterbury Earthquakes. A transportable recording studio was built in a converted shipping container, and placed in various locations around the city during 2011 and 2012. Members of the public were invited to visit the QuakeBox to record their stories about their earthquake experiences.

Participants either delivered their story to the camera alone in the studio or, at their request, with an interviewer present. However, the interviewers were instructed to keep any dialogue to a minimum, and simply let the participant tell their story without interruption (Lucy-Jane Walsh, QuakeBox interviewer, p.c.). The QuakeBox corpus thus is made up almost entirely of extended monologues. There are 523 of these monologues in the corpus, with a combined duration of approximately 85 hours.

The LaBB-CAT tool (Fromont & Hay, 2008, 2012) was used to extract all occurrences of *of*, *'s* and *s'* from the QuakeBox corpus. LaBB-CAT stores transcripts in layers, each of which contains a set of annotations to the transcript (Figure 1). I used the transcript layer for these searches as it is the only layer in which the string *s'* is represented.

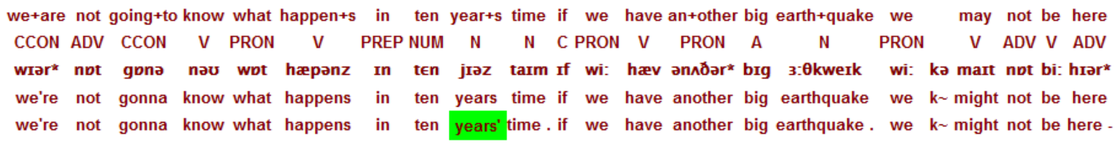


Figure 1: A screenshot illustrating some of the layers available within LaBB-CAT for a QuakeBox transcript (AP2513LJ_JulieGray). From top to bottom: morphology, syntax, phoneme, orthography and transcript.

Three searches were performed across all participants, using the search terms: **of**, **.*'s**, and **.*s'** (where **.*** is a regular expression used in LaBB-CAT to indicate a string of zero or more characters). These searches returned the target word (either *of* or a word ending in *'s* or *s'*, such as *daughter's*) and an additional twenty words on either side of the token, which were included in order to ensure the full genitive phrase was captured. In some cases where the possessor or possessum NP extended for more than 20 words, it was necessary to later manually copy the remainder of the phrase directly from the relevant QuakeBox transcript.

Personal possessive pronouns such as *my*, *your*, and *his* were not included in the search terms. Rosenbach (2014, p. 225) describes possessive pronouns as "quasi-categorical", in that although they are not strictly categorical, they occur so much more frequently with s-genitives than with of-genitives, that they are generally treated as categorical and excluded from analysis. For example, O'Connor, Maling, and Skarabela (2013, p. 100) found that 99% of the pronominal possessors in their data occurred with s-genitives. They elected to exclude possessive pronouns from their analysis, although they point out that in doing so, they may have reduced the effect of discourse status in their model. Rosenbach (2014) argues that for this reason, genitives with pronominal possessives should be included in studies of genitive alternation. However, Lowe (2016) suggests that genitives containing pronominal possessors have a different syntactic structure than those with a full NP possessor. If this is the case, then syntactic priming cannot occur between a genitive with a pronominal possessor and a genitive with a full NP possessor, strengthening the argument for excluding possessive pronouns. On balance therefore I decided not to include genitives with pronominal possessors in this study.

Each search result also returned the unique ID of the speaker; the timestamp, in hundredths of a second, of the end of the target word; and the textual distance, in orthographic characters, from the beginning of the transcript to the end of the target word. A total of 20,329 results were returned, with the number of results per speaker ranging from 1 to 270.

Textual distance was measured on the transcript layer and included spaces and pause markers, but did not include markers of non-speech sounds, such as coughs. These textual distances were used to approximate the time elapsed between instances of the genitive, so pause markers were included in the character count to give a better approximation to the actual time elapsed.

Textual distance measures were not available at the time of the initial data extraction, but were added later by repeating the original searches on the transcript layer. Because coding had already been completed for much of the original dataset, and a number of non-genitive tokens removed, a Microsoft Excel Visual Basic script (Appendix A) was used to align tokens from the new set of search results to those in the working dataset by matching speaker id and timestamp. The textual distances from the new search results were then appended to the working data.

Textual distance was used rather than temporal distance in order to allow meaningful comparison between the *Press* and the QuakeBox data. Because there is no way to directly measure how much time elapsed between the writing of genitives in the *Press* data, textual distance between the token and the preceding genitive was used as an approximation. Wolk, Bresnan, Rosenbach, and Szmrecsanyi (2013) showed that word count, character count (including spaces) and syllable count correlate strongly, meaning that any of these can be used as equally valid measure. Character count was chosen as the measure for textual distance, as it could be most easily obtained computationally.

In order to confirm that textual distance can be used as a valid substitute measure for temporal distance, textual distance and temporal distance across the full set of QuakeBox results were compared. The two measures are highly correlated, with an R^2 value of 0.9544 (Figure 2).

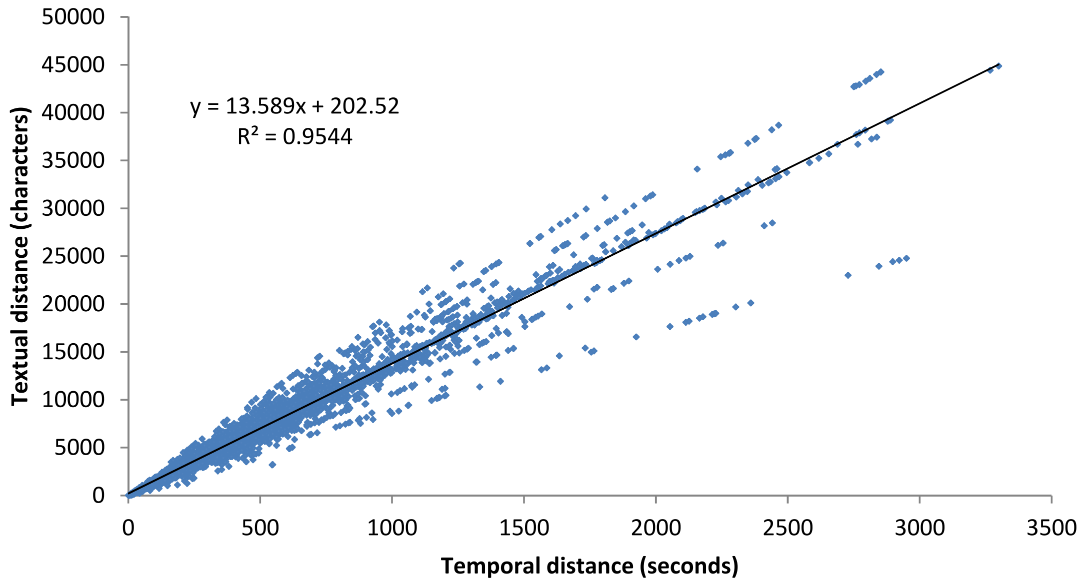


Figure 2: Correlation of textual distance and temporal distance across 4715 QuakeBox tokens.

For some individual speakers with a large number of tokens, temporal distance and textual distance are even more highly correlated. For example, BR506_BlairAnderson (Figure 3) and EG529_EdelWalker (Figure 4), with 92 and 57 tokens respectively, both show a clear linear correlation between temporal distance and textual distance, with R^2 values for both being over 0.99.

Sociolinguistic data, although available for most speakers in the QuakeBox corpus, was not extracted, because no equivalent data was available for the writers of the *Press*. Therefore, although sociolinguistic factors are likely to influence both priming and genitive choice, they were excluded from the analysis in order to keep comparisons between the two corpora as straightforward as possible. For similar reasons, speakers of varieties of English other than New Zealand English (NZE) were retained in the QuakeBox data, as it was not possible to exclude non-NZE-speaking writers from the *Press* data.

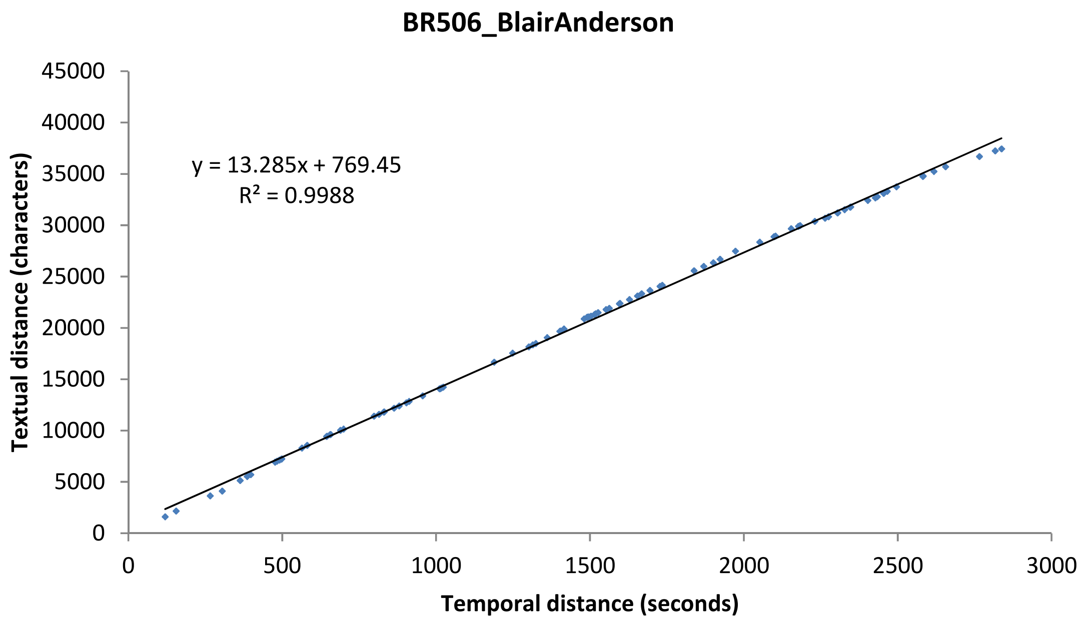


Figure 3: Correlation of textual distance and temporal distance across 92 tokens for a single QuakeBox speaker, BR506_BlairAnderson.

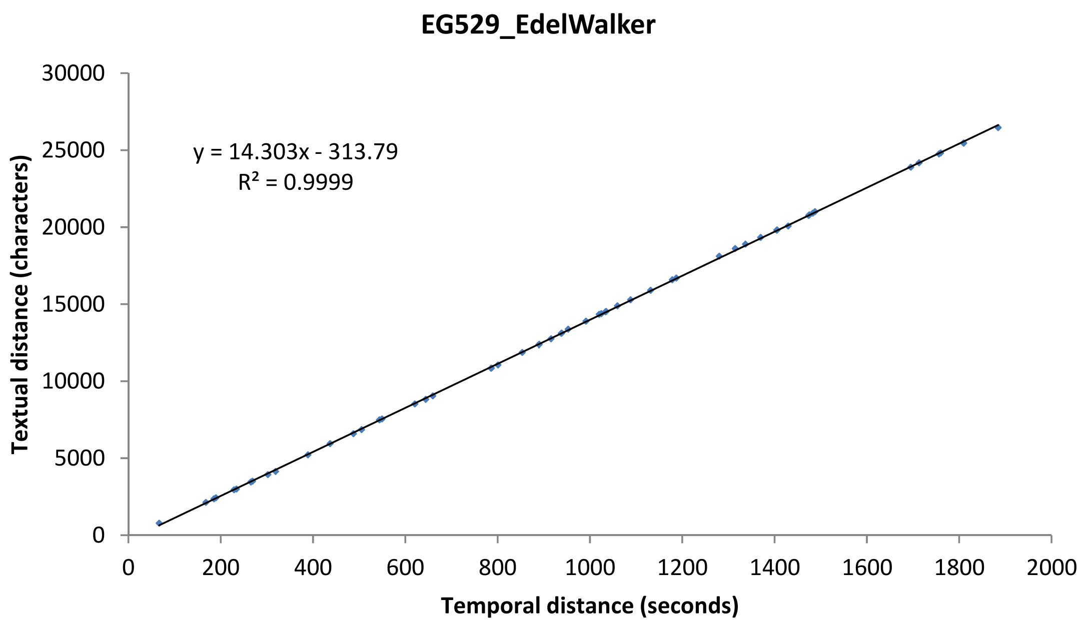


Figure 4: Correlation of textual distance and temporal distance across 57 tokens for a single QuakeBox speaker, EG529_EdelWalker.

3.1.2 *Press* database

The UC CEISMIC Canterbury Earthquakes Digital Archive (Smithies, Millar, & Thomson, 2015) is a federated archive of material relating to the Canterbury earthquakes of 2010 and 2011. Within the archive, the Fairfax Media Collection [3] contains digital copies of Christchurch's daily newspaper, the *Christchurch Press,* published between 4 September 2010 (the date of the first large earthquake to strike Christchurch) and the end of 2015. The topic of the earthquakes, and the subsequent rebuild and recovery of the city, is widely reported on in the *Press* throughout this period.

Each edition of the newspaper is available in both PDF format, and as a plain text file containing all of the articles published in that edition. These text files have been converted by Lucy-Jane Walsh of the UC Arts Digital Lab into an as yet unpublished database which cross-references each article by date, by-line, and newspaper section (allowing, for example, international news to be filtered out), enabling complex searches and analysis to be carried out across the entire collection. The full database contains 1,408,825 articles.

A MySQL search was used to find all occurrences of *of*, *'s* and *s'* in the *Press* database. As discussed above, pronominal possessors were not included in the search terms. The search was restricted to articles written between December 2011 and December 2012, the time period in which QuakeBox was collecting stories. The search was further filtered to only include articles which also included the words *earthquake* or *quake*, in order to restrict the articles to those dealing with broadly the same topic as the QuakeBox interviews.

In order to restrict articles to those written in New Zealand by a single author, the search was further filtered by newspaper section. Sections likely to include a large number of articles either fully or partially sourced from overseas newspapers, such as international news and many of the features sections, were excluded, as were sections likely to contain articles that combine writing by multiple authors, such as reviews and event listings, or content presented in a non-prose style, such as recipes, quizzes, and tables of results. Since the database did not easily allow the separation of commentary articles from results tables in the sports and racing sections, these entire sections were excluded. Thus only articles labelled by Fairfax Media as

[3] https://quakestudies.canterbury.ac.nz/store/collection/562

‘News National’, ‘Features Opinion’, or ‘Features Editorial’ were included in the search. This filtering restricted the search to 3181 articles, totalling 1,458,856 words.

Only the body of articles were searched. Headlines, lead text, and image captions were excluded as these are normally added after the article is written, and often by a different author.

The search returned 100 characters (including spaces) on either side of the target word, so that the entire genitive phrase could be captured. As with QuakeBox, in some cases where the possessor or possessum was especially long, a manual search was later needed to retrieve the remainder of the genitive phrase from the text files.

Each search result also included a unique identifier, the author of the article, and the textual distance, in orthographic characters, from the beginning of the transcript to the end of the target word. Spaces and punctuation were included in the character count, under the assumption that punctuation marks in writing will approximately correspond to pause markers in transcription of spoken language.

The database contains multiple articles written by each author (with the exception of a handful of guest contributors who wrote a single article each). There is no standardisation of the way author names are recorded in the *Press* text files (for example, the journalist Paul Gorman appears as ‘Paul Gorman’, ‘GORMAN Paul’ and ‘Paul Gorman, science reporter’), so the results were manually cleaned to standardise the names. At the same time, the results from any articles attributed to multiple authors, or with no by-line, were removed.

3.2 Data preparation and coding

Both the QuakeBox and *Press* searches returned a large number of non-genitive uses of *of*, *'s* and *s'*. For example in (10), the *'s* in *this year's* is a contraction of the word *has*, not a genitive marker. In (11) *of* modifies the preposition *out* and in (12) the adjective *full*, rather than a noun phrase possessum, and in (13), *of* is part of the discourse marker *of course*. The results were checked manually, and all such non-genitive tokens removed from the dataset.

(10) **this year's** been a lot better (QuakeBox AP2502_ChantalKennedy)

(11) the central- city courthouse has been **out of** action since the February quake (David Clarkson, 'Court building out of action', *Press*, 1 December 2011)

(12) Lucy's bedroom was so **full of bricks** (QuakeBox LY956_LizaRossie_teacher)

(13) they couldn't get through to us so **of course** they thought the worst (QuakeBox WF2536_John)

In cases of ambiguity such as (14), where *kind of* could be read either as a genitive or as a discourse marker, it was possible to return to the original audio recording for the QuakeBox data and listen to the speaker's intonation in order to determine the intended meaning. For the *Press* data, where this was not possible, the context usually supplied sufficient information to disambiguate the meaning. In those very few cases where neither method was able to fully clarify what had been intended, I picked the interpretation that seemed more likely.

(14) it was just a surreal kind of feeling (QuakeBox EG124_JulieO'Rourke)

Only genitives with an of-genitive or s-genitive structure were retained. Tokens containing double genitives (15) were removed from the dataset. Rosenbach (2005) argues that because the double genitive does not have the same structure or properties as the s-genitive, it should be excluded when examining genitive variation. For the same reason, I removed any genitives where the of-phrase was extraposed (16); instances where the of-phrase appears at the beginning of a relative clause, separated from the possessum (17); and (common in the *Press* data, where an interviewee's age is often apposited to their name) those with an appositive between the NP and the of-phrase (18).

(15) **A friend of his**, who was a barman, was also taken on as an assessor

(Martin van Beynen, 'EQC staff selected without interviews', *Press*, 9 December 2011)

(16) we started to build **a community** there **of the three parishes**

(QuakeBox EG760AW_Pauline)

(17) **the Brooklands volunteer fire brigade of** which at this stage I am now **a founder life member**

(QuakeBox BR946_RodgerCurragh)

(18) **Bruce Marsden**, 70, **of Aranui**, also believed the suburbs needed to be sorted out first.

(Francesca Lee, 'Fixing homes more of a priority than city plan', *Press*, 1 August 2012)

In cases such as (19a) (also common in the *Press*), Quirk et al. (1985) point out that *out of* can be substituted for *of* without any change to the meaning. This suggests that the structure differs from that of other quantitative genitives, such as (20), where such a substitution is not possible. Thus tokens such as (19a), where a bare number precedes the *of*, were also removed from the data.

(19)

a. 436 of the council's 2649 housing units

(Lois Cairns, 'Repairs signed off for just five quake-damaged council housing units', *Press*, 27 October 2012)

b. 436 out of the council's 2649 housing units

(20)

a. 100 hours of community work

(Anne Clarkson, 'False claim for grant miserable, judge says', *Press*, 4 February 2012)

b. *100 hours out of community work

The QuakeBox data contained many instances of false starts or stuttering, as is typical of natural language, and this often led to a genitive being repeated, resulting in multiple tokens being returned by the search. In cases such as (21), where it is only the genitive marker **of** itself which is repeated, then the utterance was treated as a single token, and only the final *of* was counted.

(21) the funeral **of of** our colleague (QuakeBox EG529_EdelWalker)

However, if anything interposed between the genitive markers, as in (22), then both tokens were retained. In a case like this, where the speaker begins to say *the back of the*, then corrects herself to *the front of the classroom*, it seems likely that the initial choice of an of-genitive structure will prime the speaker to use that same structure again in her corrected utterance, so both genitive NPs were retained as a potential prime-target pair (23a) and (23b), with (23a) having an empty possessor.

(22) he ran to **the back of the** to **the front of** the classroom

(QuakeBox EG866_Ann)

(23)

a. the back of Ø

b. the front of the classroom

As discussed above, possessive pronouns were excluded from the search terms. However, where the pronoun *it* occurred as the possessor of an of-genitive, the token was retained[4]. *It* was the only pronoun which occurred in this position in choice contexts (although the phrase *of us* was also frequently used, it only occurred in categorical contexts such as (24)). Additionally, *of it* often occurred in situations where it seemed likely to have been influenced by syntactic persistence, such as in (5) (repeated below as (25)).

(24) And yet for **nearly all of us** we did keep on

(Peter Beck, 'Making headway in time of turmoil', *Press*, 14 February 2012)

[4] This may have led to a slight skewing of the results, as *its* was not included in the original search terms, so any instances where *its* appeared in a choice condition will have been missed.

(25) I suspect from **the nature of the damage** and **the extent of it** nothing will be the same again (QuakeBox BR506_BlairAnderson)

Several QuakeBox participants' results were entirely discarded, because their recordings included singing, poetry recitation, reading from a document, or extensive dialogue with the interviewer, so could not be considered monologues. Results were retained in a small number of cases where there was some dialogue between participant and interviewer, but the interviewer did not use any genitive structures, so would not have influenced the participant's choice of genitive.

The first instance of a genitive structure in each transcript or article was excluded from the final dataset for analysis, as there is no preceding genitive to act as a prime. However, where the first genitive was followed by a genitive in a choice context, with no intervening categorical genitives, the first genitive was counted as being the immediately preceding genitive (and hence potential prime) for that second genitive. For this reason, these genitive structures (as well as other examples of genitives able to prime but unable to be targets themselves, as discussed below) were marked as such, and retained in the dataset until the coding was completed.

Almost all of the *Press* articles contain quotations within the text, such as (26). In order to minimise the potentially dialogue-like effect of these quotations, where a genitive structure appeared within a quotation (26a)[5], the genitive within the quotation was excluded entirely, and the immediately following genitive phrase (26b) was treated as if it were the first token of an entirely new article. So while (26b) would be treated as a potential prime for (26c), (26b) itself would not be included in the final dataset.

[5] I use a, b and c here to refer to the phrases marked with those superscripts within the text of the example.

(26) CHL had presented a draft contract to MidCentral that was met with silence.

"We heard nothing about that until three weeks ago, then last Monday, they said they would stop using us," English said.

"We have swapped letters expressing extreme disappointment and pleaded on **behalf of the earthquake**[a]."

The lower volume of tests[b] would have more impact than **the loss of revenue**[c], he said.

(Jo McLean-McKenzie, 'Lucrative lab contract lost in blow to DHB', *Press*, 9 December 2011)

In some cases, due to errors in the LaBB-CAT software, results from the QuakeBox search were returned with parts of the data missing (for example, having no timestamp). In these cases, if the missing data was unable to be retrieved manually, the token with the missing data was removed, and the following token treated as if it were the start of a new transcript.

After all non-genitives and problematic tokens were removed, a total of 35,390 tokens remained of the initial 67,221 search results (Table 1).

	QuakeBox	*Press*	Total
Search results	23689	43532	67221
Genitives	9720	25670	35390

Table 1: Tokens remaining in dataset after non-genitives were removed from the initial search results.

As genitives in categorical contexts do not vary, they are unable to be primed. Therefore they were excluded from the final dataset. However, in order to detect any β-persistence from these excluded tokens, where they were followed by a genitive in a choice context, I counted them as being the immediately preceding genitive for that token. So in the same way as initial genitives, categorical genitives were not immediately discarded, but instead were marked and retained until coding was complete.

In particular, I excluded genitives which expressed quantitative relationships, such as (27) and (28), of-genitives with an indefinite possessum (29), classifying s-genitives (30), elliptical genitives where there is no overt possessum (31), (32), and fixed titles and idiomatic expressions (33), (34). The idiomatic structure in (35), particularly common in the *Press* data, was also treated as categorical, as the alternate form is only possible if the elided possessor head *children* is included (36).

(27) there was **hundreds of people** in the building eventually

(QuakeBox NB2036_RodneyChambers)

(28) **the glass of water** that I normally had at the head of my bed

(QuakeBox EG865_PatPenrose)

(29) I myself was - I guess **a victim of the earthquakes** (QuakeBox WF532_Lavina)

(30) Bexley residents formally farewelled their **residents' association**

(Georgina Stylianou, 'Tears as Bexley group calls it quits', *Press*, 30 April 2012)

(31) I walked to **my sister's** [house] in Cashmere first of all

(QuakeBox LY950_StephenEstall)

(32) the rebuild of our lives and [the rebuild] **of our city**

(Peter Beck, 'Making headway in time of turmoil', *Press*, 14 February 2012)

(33) about half had been returned to **Ministry of Justice** staff

(Keith Lynch, 'Courthouse safe for jurors, says judge', *Press*, 27 January 2012)

(34) you can like win something and then **pick of the bunch** you know

(QuakeBox NB152_Lou)

(35) The unemployment beneficiary and **father of two** moved into a garage last month

(Olivia Carville, 'Abandoned homes preferred to cars for shelter', *Press*, 29 June 2012)

(36)

a. * two's father

b. two children's father

Following Szmrecsanyi (2006, p. 91), any tokens where use of the alternative form would be "ungrammatical, or very odd" were treated as categorical, and therefore excluded, but as suggested in Rosenbach (2014), I erred on the side of inclusion for marginal cases. In the case of idiomatic expressions, if a Google search for the alternate form (e.g. *the bunch's pick* for (34)) returned results from New Zealand websites, then the token was treated as non-categorical.

	QuakeBox	*Press*	Total
Search results	23689	43532	67221
Genitives	9720	25670	35390
Choice genitives	2721	11069	13790

Table 2: Choice genitives in final dataset, after removal of categorical genitives.

After marking the categorical genitives for exclusion, the remaining 13,790 choice genitives (Table 2) were coded for factors likely to have an effect on either genitive choice or priming. However, not all of the factors discussed in the literature were coded for. Rosenbach (2014) notes that strong factors can have a ceiling effect that masks the effect of weaker factors, particularly in small corpora. As the QuakeBox corpus in particular is relatively small, this was likely to occur in my data, so coding weaker factors was unlikely to reward the additional effort. Therefore I only coded for those factors known to have a strong effect.

Each token was coded for the following factors:

GENITIVE TYPE

The type of genitive, either **s** or **of**.

GENITIVE MATCH

A Boolean value, where **true** = the genitive type matches that used in the immediately preceding genitive. The preceding genitive may be in either a categorical (β-prime) or choice (α-prime) condition.

SPEAKER

A unique speaker identifier. In the case of the QuakeBox corpus, this was the speaker ID value extracted from LaBB-CAT. For the *Press* data, this was the author's name, standardised as discussed above, so that multiple articles written by the same author were coded as the same SPEAKER.

MODE

The source of the data, either **S** (spoken, i.e. the QuakeBox corpus) or **W** (written, i.e. the *Press* database).

ANIMATE

A Boolean value, where **true** = the possessor is animate. Although many schemas exist for categorising animacy, Grafmiller (2014) argues that where animacy is only of concern as a potentially confounding factor, it is sufficient to use a binary distinction. I followed his lead, coding humans, animals and human collectives (such as organisations (37)) as animate, and all other possessors as inanimate.

(37) ECan's governance arrangements

(Rachel Young, 'Quake no excuse – Bazley', *Press*, 8 October 2012)

Geographical entities such as *Christchurch* were coded as animate where the speaker was clearly referring to the city as a collection of people (38), and as inanimate where the speaker was referring to the city as a physical place (39). Where in doubt, geographical entities were coded as inanimate.

(38) Christchurch's housing woes

(Jo McKenzie-McLean, 'Urgent call for social housing', *Press*, 28 May 2012)

(39) the outskirts of Christchurch (QuakeBox EG885_TraceyAdams)

SIBILANT

A Boolean value, where **true** = the possessor NP ends in a sibilant phoneme. When deciding possessor sibilance for *Press* tokens, abbreviations, symbols and numbers were judged as they would normally be pronounced in New Zealand English. For example, *$2000* would normally be read as /tu θaʊzənd dɒləz/ so would be coded as sibilant.

PROTOTYPICAL

A Boolean value, where **true** = the semantic relationship between possessor and possessum is prototypical. Rosenbach (2014) argues that although semantic relationship undoubtedly exists on a gradient, it is not practical to use more complex classifications except in very large corpora, so suggests restricting classification to a binary dimension of prototypical versus non-prototypical. I followed her lead, coding legal ownership (40), kin terms (41), body parts (42), and part-whole relationships (43) as prototypical, and all other relationships as non-prototypical.

(40) my wife's French horn (QuakeBox SU2401_VicBartley)

(41) the neighbour's kids (QuakeBox NB171_AnneDavis)

(42) Dalton's finger
(Anne Clarkson, 'Man fined $750 for biker attack', *Press*, 24 July 2012)

(43) the door of our wardrobe (QuakeBox EG2024_SusanWalsh)

POSSESSUM WEIGHT and POSSESSOR WEIGHT

The number of orthographic characters, including spaces, in the possessum NP and possessor NP respectively. Before calculating the weights, any pauses[6], hesitations, false starts or repetitions in the QuakeBox data were removed (e.g. (44a) becomes (44b)). This "cleaned up" version of the NP presumably reflects what was intended by the speaker, the length of

---

[6] Where punctuation was present within a possessum or possessor NP in the *Press* data, it was retained, on the assumption that punctuation represents a deliberate choice by the author. Although some pauses in the QuakeBox data will similarly have been a deliberate choice by the speaker, there is no easy way to distinguish these deliberate pauses from unvoiced hesitations, so the decision was made to exclude all pause markers.
Pauses, hesitations and other production errors were retained in the character count for DISTANCE, however, as here what is of interest is the actual time which elapsed between genitives, not the intention of the speaker at the planning stage.

which should have a stronger influence on the genitive choice than any additional length caused by production errors. Where a speaker began by saying one phrase, but corrected themselves to another (such as in (45a), where *grandmother* is clarified to *ninety six year old grandmother*), then the first phrase uttered was used to measure weight, again on the assumption that it was what was originally intended by the speaker, so will have determined the genitive choice[7].

(44)

a. the extent of **the - ahh -** the road and the liquefaction and **. um** collapsed buildings

(QuakeBox AP2503_PeterYoung)

b. the extent of the road and the liquefaction and collapsed buildings

(45)

a. the ahh doctor's **wi~ ahh** grandmother **ninety six year old grandmother**

(QuakeBox AP2500)

b. the doctor's grandmother

When fitting the data model, both the cleaned and raw weights were tested, and the cleaned weights provided a better fit for the model, further justifying this choice.

Additionally, the determiner was removed before calculating POSSESSUM WEIGHT for of-genitives (following Rosenbach, 2005; Wolk et al., 2013), because in the equivalent s-genitive, the genitive would not appear. For example, in the s-genitive form of (44), the determiner *the* is dropped from the possessum (46), leaving a POSSESSUM WEIGHT of 6, the number of characters in *extent.*

(46) the road and the liquefaction and collapsed buildings' extent

The frequency distributions of POSSESSUM WEIGHT and POSSESSOR WEIGHT were highly skewed, with a small number of tokens having extremely high values (Figures 5 and 6). In

---

[7] The decision to "clean" the weights in this way resulted from discussion following my presentation of the early stages of this project to the New Zealand English and English in New Zealand (NZEENZ) conference, University of Canterbury, 8-9 June 2016.

order to restrict the effect of these outlier values on the model, a cut-off point was determined for each, at the ten-character-wide interval above which there were less than 10 tokens per interval. For POSSESSUM WEIGHT this cut-off was at a length of 80 characters, and for POSSESSOR WEIGHT at a length of 150 characters. The 44 tokens which fell above these cut-off points were removed from the dataset.

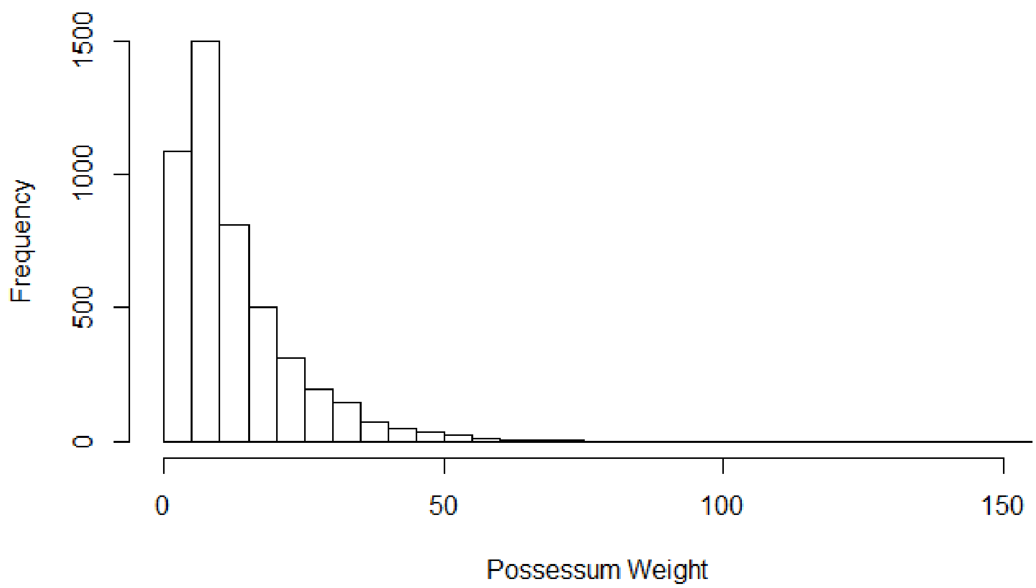


Figure 5: Frequency distribution of POSSESSUM WEIGHT across the full dataset.

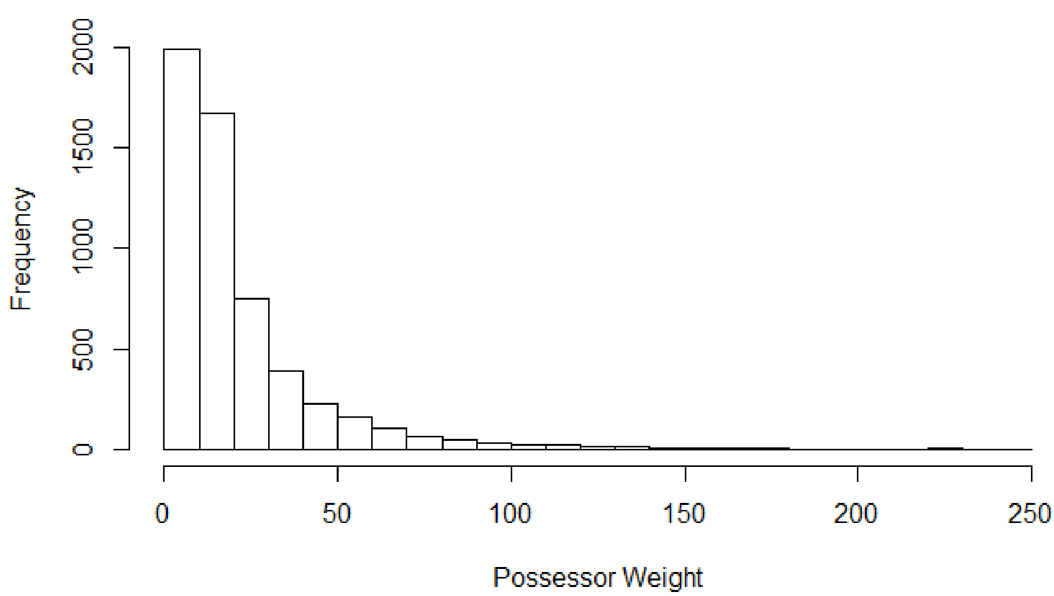


Figure 6: Frequency distribution of POSSESSOR WEIGHT across the full dataset.

WEIGHT RATIO

The ratio of POSSESSOR WEIGHT to POSSESSUM WEIGHT, calculated as POSSESSOR WEIGHT / POSSESSUM WEIGHT. This measure of the relative weights of the possessum and possessor provided a better fit for models with GENITIVE MATCH as the dependent variable.

PRIME TYPE

Whether the immediately preceding genitive is in a choice (α) or categorical (β) context[8]. In order to simplify the coding, Roman alphabet **A** and **B** were used in place of α and β.

DISTANCE

As discussed above, textual distance in characters between the end of the target genitive and the end of the immediately preceding genitive was used as a substitute for temporal distance. As the relationship between distance and priming strength has been found to be logarithmic (e.g. (Gries, 2005), (Szmrecsanyi, 2006)), the natural log of this textual distance measure was used when fitting the data model.

POSSESSUM MATCH and POSSESSOR MATCH

Boolean values, where **true** = the head of the possessum or possessor NP, respectively, matches that of the immediately preceding genitive. Cleland and Pickering (2003) showed that priming in noun phrase structures is increased when the heads of the noun phrases are identical. As the possessum is the head of the genitive structure, it would be expected that matching possessums should increase priming strength. However, Szmrecsanyi (2006) was unable to find any significant effect on priming for matching possessums. He did though find that where *possessors* match, the likelihood of s-genitive use is greatly increased. I therefore coded for both POSSESSUM MATCH and POSSESSOR MATCH.

Two NPs were considered a match if the singular form of the two head nouns matched. For example, POSSESSOR MATCH is true in (47).

[8] I am using Szmrecsanyi's (2006) terms α- and β-priming here, but my definition of β-priming is somewhat narrower than his. Szmrecsanyi considers all structures containing the preposition *of* to be potential β-primes, while I restrict β-primes to only of-genitives in a categorical context.

(47)

a. a handful of **big commercial buildings**

b. **the PGC building**'s structural shortcomings

(Ben Heather, 'Public, engineers understood "safe" in different ways', *Press*, 10 December 2011)

In an example such as (48), where the possessor NP *a group of workers* is itself a genitive, the head of the possessum of the embedded genitive (in this case, *group*) is the head of the entire possessor NP, so the following genitive would be coded POSSESSOR MATCH = true if the head of its possessor was also *group*.

(48) **a group of workers**' "negative mood states"

(Joelle Dally, 'Leafy offices cost ratepayers $48,000', *Press*, 21 November 2012)

In the case of (49), where the embedded possessor NP expresses a quantitative relationship, it could be argued that it is *land* that is the head of the NP, rather than *any*[9].

(49) the zoning of **any of the land** (QuakeBox BR2106_StephenBourke)

I did initially code separately for this alternative method of determining a match for quantitative genitive possessors or possessums. However, the number of POSSESSOR MATCH and POSSESSUM MATCH values that were changed by using the alternative method was very small (around 30 tokens in total), and the correlation between the values produced by the two methods was very high ($p < 10^{-16}$), meaning that my choice effectively made no difference to the overall results. Therefore I continued to define the head of any embedded genitive as the head of its possessum NP.

Likewise, where a possessor or possessum is not a noun phrase, a number of interpretations are possible for what should be considered as a match between heads in this context. The

[9] The logical conclusion of such an argument is that if quantitative genitives have their head in a different position to other of-genitives, then they must also have a different syntactic structure. This would mean that they cannot prime the of-genitive structure, so should have been excluded entirely from the dataset, rather than retained as potential β-primes. Discovering whether quantitative genitives are indeed able to prime other of-genitives is beyond the scope of this study, but would be an interesting area for future investigation.

decisions I made in these cases are recorded below, but all occurred in the data very infrequently, and none often enough to significantly affect the final results.

Conjoined structures such as (50) were coded as a match only where the previous possessum was also a conjoined structure, and the heads of the constituents of the two conjunctions also matched.

(50) **the government and the people** of New Zealand

(QuakeBox BR2106_StephenBourke)

Possessors with gerund heads, such as (51), were coded as a match only if the previous possessor's head was the same gerund.

(51) the cost of **restoring our facilities**

(Simone Pearson, 'Inner-city pool worth effort to save', *Press*, 3 May 2012)

Free relative possessors, such as (52), were coded as a match only if the previous possessor's head was the same relative pronoun.

(52) the value of **what dogs mean in our community**

(QuakeBox BR506_BlairAnderson)

Three datasets were produced (Table 3): the full dataset, and two subsets containing the *Press* data and the QuakeBox data respectively.

	QuakeBox	*Press*	Full dataset
Number of tokens	2719	11023	13742
Number of speakers	396	149	545

Table 3: Size of final datasets.

3.3 Statistical analysis

Binomial mixed-effects models were hand-fit to each dataset in the R software package (R Core Team, 2016), using the lme4 library (Bates, Maechler, Bolker, & Walker, 2015) and the bobyqa optimiser. Factors were tested for collinearity using the vif.mer function (Frank, 2011), and all variance inflation factors were found to be below 2, so within the acceptable range. The rcs function of the rms package (Harrell, 2016) was used in order to model POSSESSOR WEIGHT and POSSESSUM WEIGHT non-linearly, using a 3-knot restricted cubic spline. Two models were fit to each dataset, the first with GENITIVE TYPE as the dependent variable, and the second with GENITIVE MATCH as the dependent variable.

SPEAKER was treated as a random intercept in all models. Initial graphing of factors across a sample of speakers (Appendix B) suggested that by-speaker random slopes may be present for a number of factors (Barr, Levy, Scheepers, & Tily, 2013). However, because so many of the speakers provided only a small number of tokens each[10], there was insufficient data to allow models with random slopes to converge (Bates, Kliegl, Vasishth, & Baayen, 2015), so only the random intercept was included in the final models.

The model selection was guided by χ^2 likelihood tests, Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). Where AIC and BIC disagreed, AIC was used, as BIC tends to err on the side of too simple a model while AIC errs on the side of too much complexity (Dziak, Coffman, Lanza, & Li, 2012). AIC seemed more suitable since my model has potentially already been over-simplified by failing to include factors likely to have only a weak effect.

[10] This is particularly true for the QuakeBox data, with a median of 7 tokens per speaker (mean = 9.24, range = 2-55). The *Press* data has a median of 17 tokens per speaker (mean =79.67, range = 2-909).

4 Results

This chapter will present the results of this study. In Section 4.1, I discuss factors predicted to directly affect genitive choice, and in section 4.2 I discuss the effect of priming on genitive choice, and examine factors predicted to affect the priming strength.

4.1 Effects on genitive choice

This section discusses those factors which are predicted to directly affect genitive choice, namely possessor animacy, possessor-final sibilance, possessor and possessum weight, and prototypicality of the semantic relationship.

Across the full dataset, of-genitives are more common than s-genitives, with 56% of tokens being of-genitives (Table 4). This difference is more marked (χ^2 = 468, $p < 0.0001$) in the QuakeBox data, where 74.5% of the tokens are of-genitives.

	of-genitive		s-genitive		Total
	N	%	N	%	
Press	5674	51.5	5349	48.5	11023
QuakeBox	2027	74.5	692	25.5	2719
Full dataset	7701	56.0	6041	44.0	13742

Table 4: Proportion of of-genitives.

Table 5 shows the binomial mixed effects model fit over the full dataset, with GENITIVE TYPE as the dependent variable. The model predicts the likelihood of the s-genitive, so factors with a positive coefficient indicate conditions which are more likely to produce an s-genitive, and factors with a negative coefficient indicate those more likely to produce an of-genitive.

Note that in this table, and in subsequent models, two values are reported for each of POSSESSUM WEIGHT and POSSESSOR WEIGHT. These represent the first and second points of the 3-knot restricted cubic spline used to model each weight non-linearly.

	Coefficient	SE	z	Pr(>|z|)	Sig
(intercept)	-1.379	0.627	-2.20	0.03	*
ANIMATE = true	6.315	0.355	17.77	<0.0001	***
SIBILANT = true	-1.207	0.077	-15.67	<0.0001	***
PROTOTYPICAL = true	1.172	0.337	3.48	0.0005	***
POSSESSUM WEIGHT	-0.038	0.064	-0.60	0.55	
POSSESSUM WEIGHT’	0.177	0.123	1.44	0.15	
POSSESSOR WEIGHT	0.055	0.041	1.34	0.18	
POSSESSOR WEIGHT’	-0.358	0.142	-2.52	0.01	*
MODE = written	2.472	0.597	4.14	<0.0001	***
GENITIVE MATCH = true	-3.046	0.419	-7.26	<0.0001	***
log(DISTANCE)	-0.427	0.042	-10.21	<0.0001	***
POSSESSUM MATCH = true	-4.224	0.745	-5.67	<0.0001	***
POSSESSOR MATCH = true	0.717	0.127	5.64	<0.0001	***
PRIME TYPE = β	2.667	0.113	23.69	<0.0001	***
ANIMATE = true x PROTOTYPICAL = true	0.759	0.177	4.28	<0.0001	***
ANIMATE = true x POSSESSUM WEIGHT	-0.112	0.031	-3.57	0.0004	***
ANIMATE = true x POSSESSUM WEIGHT’	0.141	0.058	2.43	0.02	*
ANIMATE = true x MODE = written	-3.239	0.299	-10.82	<0.0001	***
PROTOTYPICAL = true x POSSESSUM WEIGHT	-0.124	0.042	-2.94	0.003	**
PROTOTYPICAL = true x POSSESSUM WEIGHT’	0.224	0.081	2.75	0.006	**
POSSESSUM WEIGHT x MODE = written	0.246	0.062	3.98	<0.0001	***
POSSESSUM WEIGHT’ x MODE = written	-0.422	0.119	-3.54	0.0003	***
POSSESSOR WEIGHT x MODE = written	-0.199	0.043	-4.62	<0.0001	***
POSSESSOR WEIGHT’ x MODE = written	0.397	0.146	2.73	0.006	**
GENITIVE MATCH = true x MODE = written	1.033	0.272	3.80	0.0001	***
GENITIVE MATCH = true x log(DISTANCE)	0.334	0.060	5.53	<0.0001	***
GENITIVE MATCH = true x POSSESSUM MATCH = true	1.524	0.400	3.80	0.0001	***
GENITIVE MATCH = true x PRIME TYPE = β	-5.01	0.155	-32.32	<0.0001	***
log(DISTANCE) x POSSESSUM MATCH = true	0.521	0.155	3.35	0.007	***

* significant at p <.05, ** p <.01, *** p <.001

Table 5: Binomial mixed-effects model fit to the entire dataset, with GENITIVE TYPE as the dependent variable .

As would be expected from previous studies (e.g. Altenberg, 1982; Grafmiller, 2014; Rosenbach, 2002; Stefanowitsch, 2003; Szmrecsanyi, 2006), ANIMATE, SIBILANT and PROTOTYPICAL all affect genitive choice significantly.

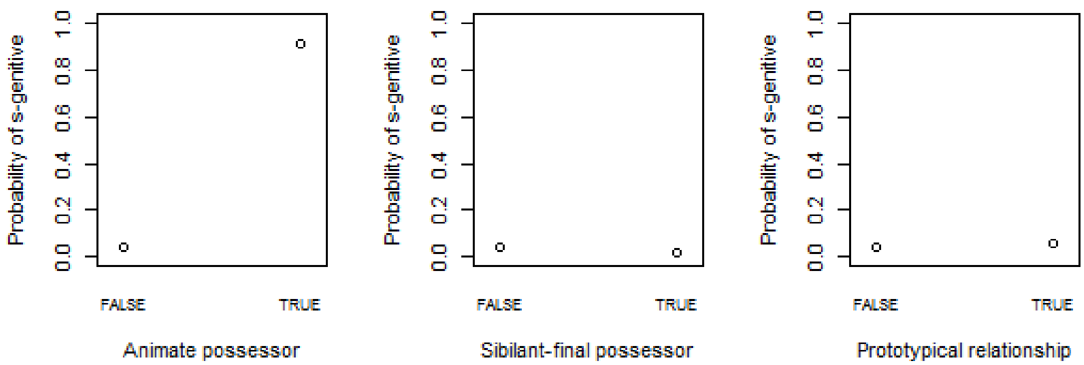


Figure 7: Main effects of animacy (left), sibilance (centre) and prototypicality (right) on genitive choice.

Animacy in particular has very strong effect (Figure 7), with an animate possessor greatly increasing the likelihood of an s-genitive. The effects of sibilance and prototypicality are much smaller, but still significant: a possessor which ends in a sibilant phoneme decreases the likelihood of an s-genitive, and a prototypical relationship is more likely to be expressed with an s-genitive.

Animacy is known to be one of the strongest factors affecting genitive choice (Rosenbach, 2005). Additionally, the effect of animacy may have been strengthened by the fact that the data was gathered in New Zealand. New Zealand English (NZE) has been shown to be particularly sensitive to animacy (Bresnan & Hay, 2008; Hundt & Szmrecsanyi, 2012), and while I made no attempt to restrict the data in this study to NZE speakers, it is highly likely that the vast majority of the QuakeBox participants and *Press* authors are indeed speakers of NZE.

There is also an interaction between ANIMATE and PROTOTYPICAL (Figure 8), the effect of which is to slightly increase the effect of animacy when the relationship is prototypical.

Figure 8: Interaction between ANIMATE *and* PROTOTYPICAL.

No main effect was found for POSSESSUM WEIGHT, and only the second point for POSSESSOR WEIGHT reached significance, although only at the $p < .05$ level. This appears to be counter to previous studies (e.g. Altenberg, 1982; Rosenbach, 2005; Shih et al., 2015; Szmrecsanyi, 2006; Wolk et al., 2013), where weight was found to be a significant factor. However, interactions were found for POSSESSUM WEIGHT with both PROTOTYPICAL and ANIMATE (Figure 9).

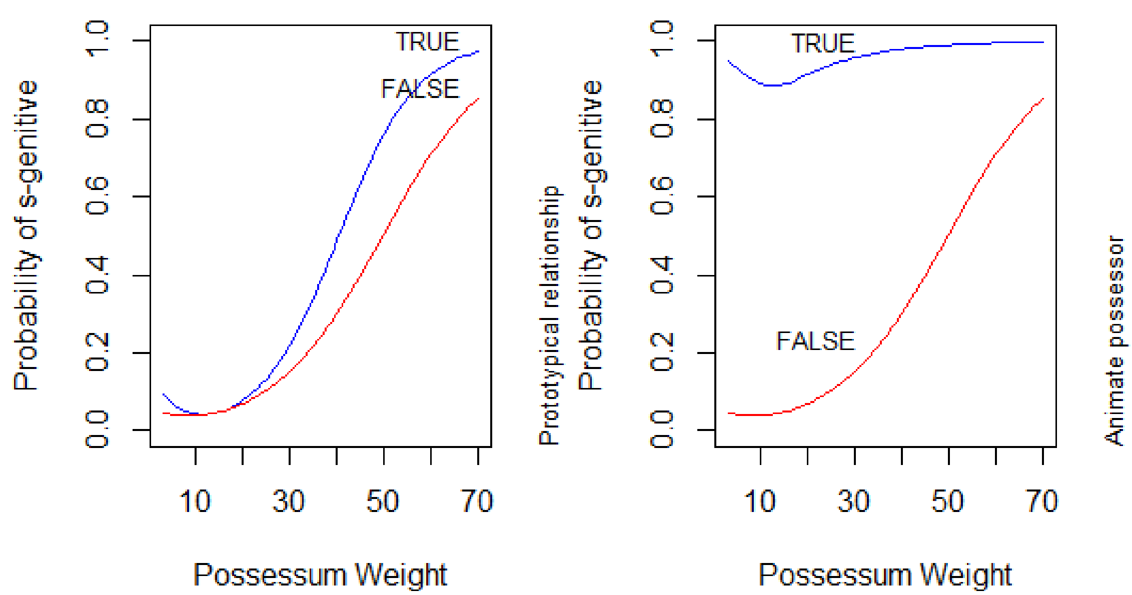


Figure 9: Interactions between POSSESSUM WEIGHT *and* PROTOTYPICAL *(left), and* POSSESSUM WEIGHT *and* ANIMATE *(right).*

When the relationship is prototypical, the effect of POSSESSUM WEIGHT is slightly increased, compared to a non-prototypical relationship. That is, an s-genitive is more likely to be used with a heavy possessum if the relationship is prototypical.

When the possessor is animate, POSSESSUM WEIGHT has little effect, and the probability of an s-genitive remains high for all weight values. This is as would be expected, given the effect size of animacy, which is so large as to effectively override the effect of weight. For inanimate possessors, however, the probability of an s-genitive is greatly increased for longer possessums, as would be expected. In all cases, the relationship is non-linear, with weight only beginning to have an effect on genitive choice when the possessum is longer than approximately 10 characters. For shorter possessums, differences in weight appear to have no effect.

This interaction between weight and animacy may explain why no main effect was found for POSSESSUM WEIGHT. Over 43% of the genitives in the dataset have animate possessors (Table 6). The strong effect of animacy overriding that of weight on this large part of the dataset may have masked the effect of POSSESSUM WEIGHT overall.

	Animate		Inanimate		Total
	N	%	N	%	
Press	5076	46.1	5947	54.0	11023
QuakeBox	853	31.4	1866	68.6	2719
Full dataset	5929	43.2	7813	56.9	13742

Table 6: Distribution of animate possessors across datasets.

The model also shows a significant main effect for MODE (Figure 10), with s-genitives more likely in the written data than the spoken. This result matches that of Grafmiller (2014), who found greater s-genitive use in a corpus of written data from the 1960s than in a corpus of spoken data from the 1990s, and thus confirms Grafmiller's assertion that the difference he found is due to an actual difference between speech and writing, and not merely reflective of changes over time.

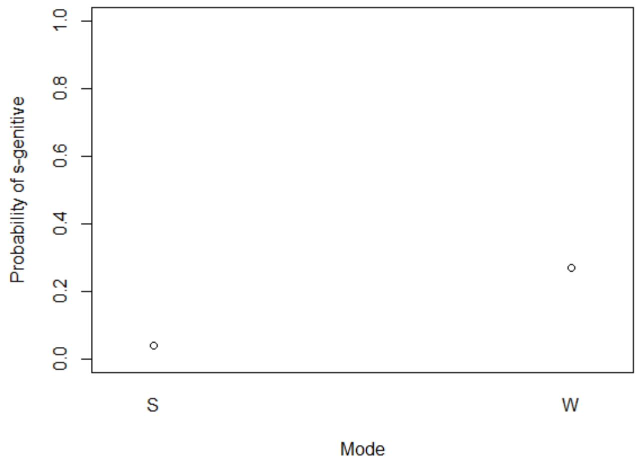


Figure 10: Probability of s-genitive across spoken (S) and written (W) data.

Two competing pressures, colloquialisation and economisation, have been proposed as driving forces behind the relatively high frequency of s-genitive use in newspaper writing. In a study of changes in written language between 1961 and 1991, Leech and Smith (2006) found an increase in the use of the s-genitive, and attributed this change to a general trend towards colloquialisation of written language, as s-genitives are found more frequently in less formal registers (e.g. Jucker, 1993). Szmrecsanyi and Hinrichs (2008) supported this view, pointing to their finding that the effect of possessor-final sibilance on genitive choice has grown over time in newspaper writing, and arguing that, as there is no reason why a phonological factor should have a strong effect in writing, this increase must reflect increasing colloquialisation and imitation of spoken style. I found no significant interaction between MODE and SIBILANT in my data, meaning that effect of sibilance in writing is as great as in speech, which would seem to support Szmrecsanyi and Hinrich's claim. But if s-genitive use is increasing in newspaper writing because it is becoming more informal and speech-like, this does not explain why there is *more* s-genitive use in the *Press* corpus than in the QuakeBox data. Colloquialisation cannot be the only factor at play here.

Another explanation for s-genitives appearing more frequently in newspaper writing is presented by Grafmiller (2014), who posits that the difference is due to journalists being more likely to use the s-genitive with inanimate possessors. Biber (2003) suggested that innovation in noun phrases in newspapers is caused by the pressure on journalists to convey information in the most economical way, and Grafmiller argues that this pressure towards economisation means journalists are more likely to choose the more compact s-genitive form even when the possessor is inanimate. This does appear, at first glance, to be supported by my data, as the effect of animacy is weaker in the written corpus (Figure 11), and Table 7 shows that while only 2% of inanimate possessors are found with an s-genitive in the QuakeBox data, the proportion of inanimate possessors with s-genitives in the *Press* data is much higher, at 28.7% ($\chi^2 = 583$, $p < 0.0001$). However, Table 7 also shows that the journalists in the *Press* data are more likely than the QuakeBox speakers to use of-genitives with animate possessors, which does not fit with the economisation hypothesis. This difference is not as large as that for s-genitive use with inanimate possessors, but it is still significant ($\chi^2 = 9.61$, $p < 0.002$).

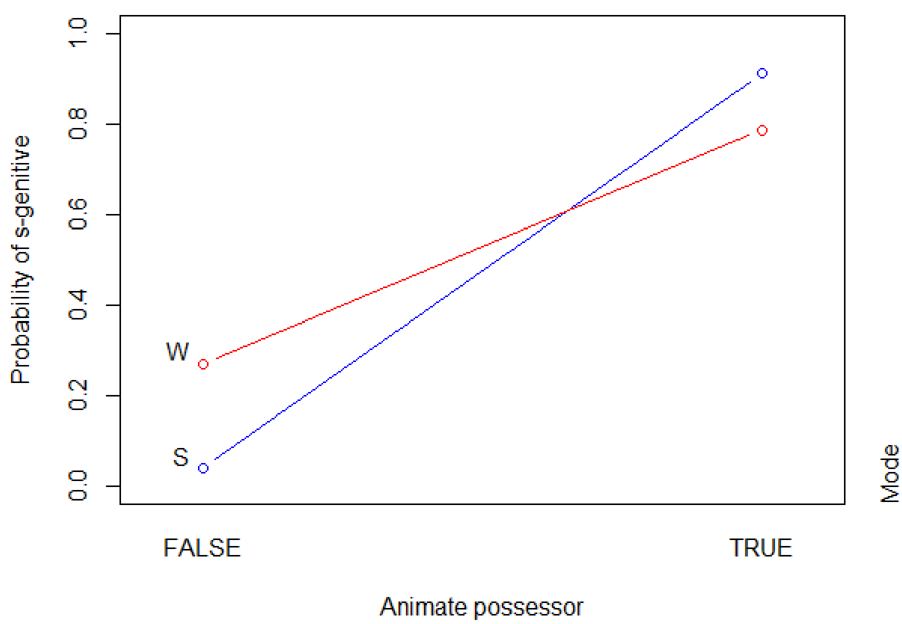


Figure 11: Interaction of MODE and ANIMATE in full model. Lines have been added to better illustrate the difference in the slopes.

	Animate possessor				Inanimate possessor			
	Spoken		Written		Spoken		Written	
	N	%	N	%	N	%	N	%
of-genitive	199	23.3	1431	28.2	1828	98.0	4243	71.3
s-genitive	654	76.6	3645	71.8	38	2.0	1704	28.7

Table 7: Use of s-genitives with inanimate possessors.

This apparently contradictory result may be a result of the way I have coded animacy. By using a binary animate/inanimate categorisation, finer-grained distinctions between levels of animacy have been lost. Jankowski and Tagliamonte (2014) found near-categorical use of the s-genitive with prototypically human possessors, but noted that with non-human animate possessors, s-genitive use is much lower. Jankowski (2013), using the animacy coding system developed by Zaenen et al. (2004) which distinguishes organisations (defined as having a collective voice or purpose) from other human collectives such as crowds (which merely act collectively), found that fewer s-genitives are used with organisations, and concluded that organisations are less animate than human collectives.

In my data, organisations were considered to be human collectives, so were coded as animate. It is likely that there are more mentions of organisations in the *Press* than in QuakeBox, as the QuakeBox participants are telling their personal stories, while the increasing trend in newspaper writing has been toward more often discussion of organisations than of individual people (Szmrecsanyi, 2013). If organisations are indeed behaving more like inanimate possessors, the result would be to skew the *Press* data towards more of-genitive use with possessors that are coded animate. And indeed, many examples of of-genitives occurring with organisation possessors can be found in the data (53).

(53)

a. the permission of Housing New Zealand

(Olivia Carville, ‘11 into four bedrooms won’t go’, *Press*, 31 March 2012)

b. the end of Ironhorse Hobbies

(Charlie Gates, ‘Firms rally on red-zone limit’, *Press*, 10 April 2012)

c. the operational arm of the Wood Council of New Zealand

(Jane Arnott, 'Case for wood rebuild is compelling', *Press*, 14 March 2012)

d. the assistance of Fisher and Paykel

(Lianne Dalziel, 'Community's response humbling', *Press*, 22 February 2012)

e. the relaunch of Ballantynes

(Charlie Gates, 'Red-zone video store open again', *Press*, 17 March 2012)

f. the home of Untouched World

(Charlie Gates, 'December quakes nearly fell Arts Centre landmark', *Press*, 24 February 2012)

Also contributing to the frequency with which of-genitives appear with organisation possessors may be the journalistic idiom of describing an employee of an organisation as *person, of organisation*. Again, many examples can be found in the data.

(54)

a. Tim Howe, of Ocean Partners

(Alan Wood, 'Flash marketing of city a waste – bank', *Press*, 9 July 2012)

b. Hamish Evans of Switch

(Pierre Changuion, 'Brighton Rising to the sounds of dub', *Press*, 14 January 2012)

c. organiser Janine Morrell-Gunn, of Whitebait TV

(Tina Law, 'Hope still shines', 5 March 2012)

d. Jonathan Barnett, of Beca

(Marc Greenhill, 'Victim had told wife not to worry', *Press*, 14 December 2011)

Not all of the tokens which fall into this animate/of-possessor intersection are organisations, however. Of interest when looking at the raw data is how often the of-genitive appears to be used with EQC staff members. EQC, the Earthquake Commission, is New Zealand's government-run insurance scheme covering land, home and contents damage caused by natural disasters. EQC has a very poor reputation in Canterbury due to their perceived mishandling of insurance claims arising from the 2010-2011 earthquakes, and this disfavour is reflected in the *Press*, which often prints articles critical of EQC. This leads me to wonder whether the of-genitive construction is being used (consciously or unconsciously) by the journalists in order to dehumanise the EQC staff, as animacy can be thought of as reflecting a speaker's degree of empathy with or connection to the entity being spoken about (Yamamoto, 1999). Investigating this question, and the wider issue of how animacy should best be categorised, is beyond the scope of this study, but it presents an interesting area for future research.

A further possibility for the difference in s-genitive use between speech and writing is suggested by considering the question not as "why are there so many s-genitives in writing?", but rather as "why are there so few s-genitives in speech?" QuakeBox participants are telling their personal stories, so it would be expected that they would frequently refer to human subjects in their speech, and therefore, given the strong effect of animacy, use a large number of s-genitives. But this is not what is found.

The answer to where the missing s-genitives are may lie in possessive pronouns. Where a human possessor has been previously mentioned in the monologue, it is likely that the speaker will express the relationship using a possessive pronoun rather than an s-genitive. For example, in (55) the speaker refers to *his wife*, not *my neighbour's wife*.

(55) I knew one of my neighbours was in hospital . um h~ with heart problems and **his wife** would be home on her own

(QuakeBox AP2510_Beryl)

Thus it may be that a large number of potential s-genitives in the QuakeBox data are being replaced by possessive pronouns. Possessive pronoun use was not recorded in my data, so no direct comparison can be made of the frequency of possessive pronouns in QuakeBox and the *Press*, but it seems likely that the frequency will be lower in the more formal language of the *Press*. I will return to the issue of possessive pronouns in Section 4.2 below.

No interaction was found between MODE and PROTOTYPICAL. This may be simply due to insufficient data, as when the *Press* data and QuakeBox data were modelled separately (Table 9 and Table 8 below), PROTOTYPICAL failed to reach significance in either model.

The model does however show significant interactions between POSSESSOR WEIGHT and MODE, and between POSSESSUM WEIGHT and MODE. In Figure 12, it appears that while the effect of POSSESSUM WEIGHT for speech and writing is similar at the lower and upper ends of the weight range, in the middle the effect is much stronger in writing[11]. However, this may not be the case. As discussed in the methodology, the weight distribution is skewed heavily towards the lower end of the range, even after extreme outliers were removed, and this is especially true for the spoken data, where more than three quarters of the possessums are less than 10 characters long (Figure 13). Thus, except at the extreme low end, the curves in Figure 12 are based on very sparse data, and must be treated with caution. And indeed, Table 8 shows that POSSESSUM WEIGHT fails to reach significance as a main effect in the QuakeBox data.

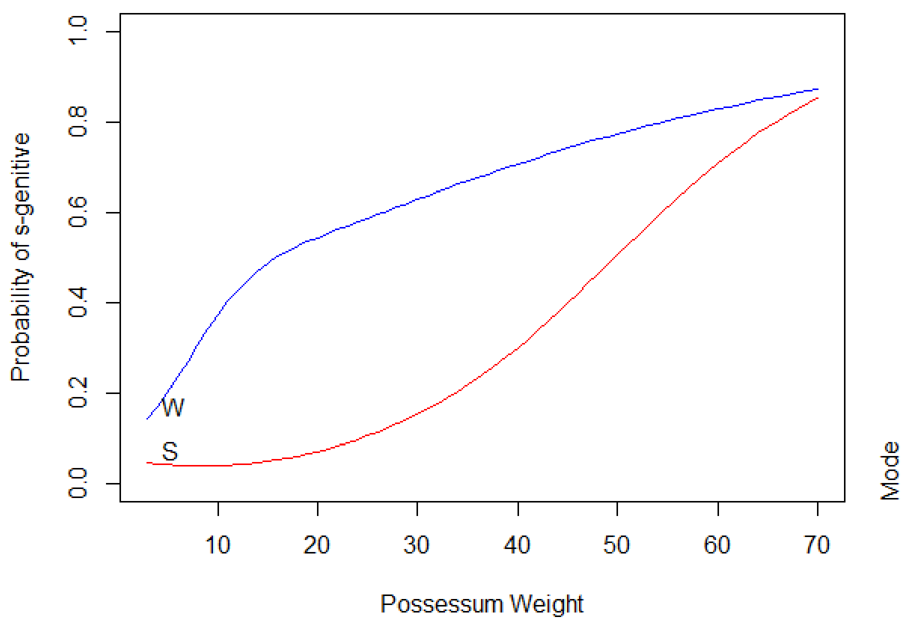


Figure 12: Interaction of MODE with POSSESSUM WEIGHT.

[11] The possibility was considered that this difference may have been influenced by the inclusion of punctuation in the weight measures for the written data, while pause markers were excluded from the spoken data. However, punctuation in possessors and possessums is relatively uncommon, and generally only found at the top end of the weights.

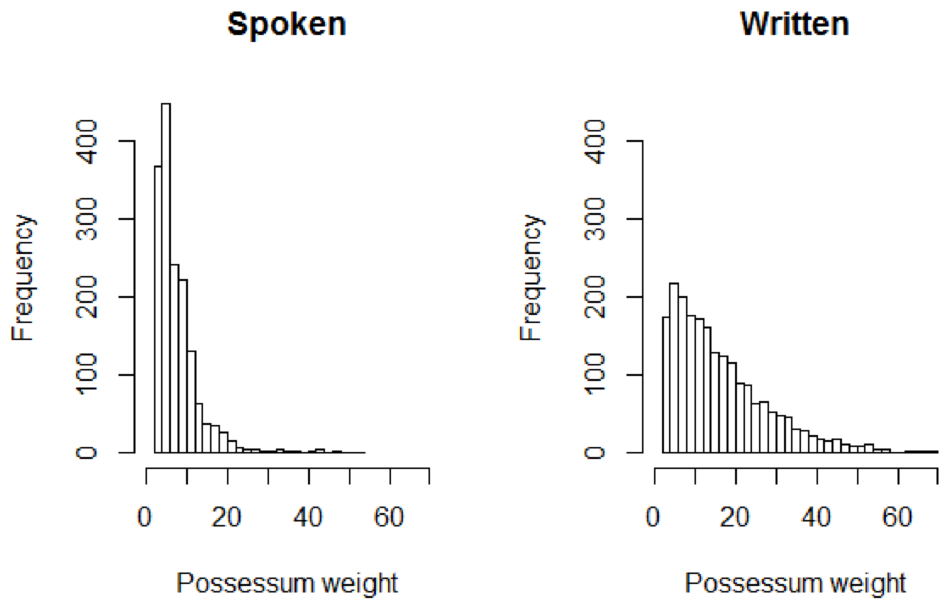


Figure 13: Frequency distribution of POSSESSUM WEIGHT *across spoken (left) and written (right) data.*

Similarly, Figure 14 shows the interaction between MODE and POSSESSOR WEIGHT, with POSSESSOR WEIGHT having no effect on the spoken data, while for the written data, a short possessor greatly increases the likelihood of an s-genitive. But again, POSSESSOR WEIGHT is very heavily skewed towards the low end of the weight range for the spoken data (Figure 15), with a median length of only 10 characters, so much of the curve is modelled on very sparse data.

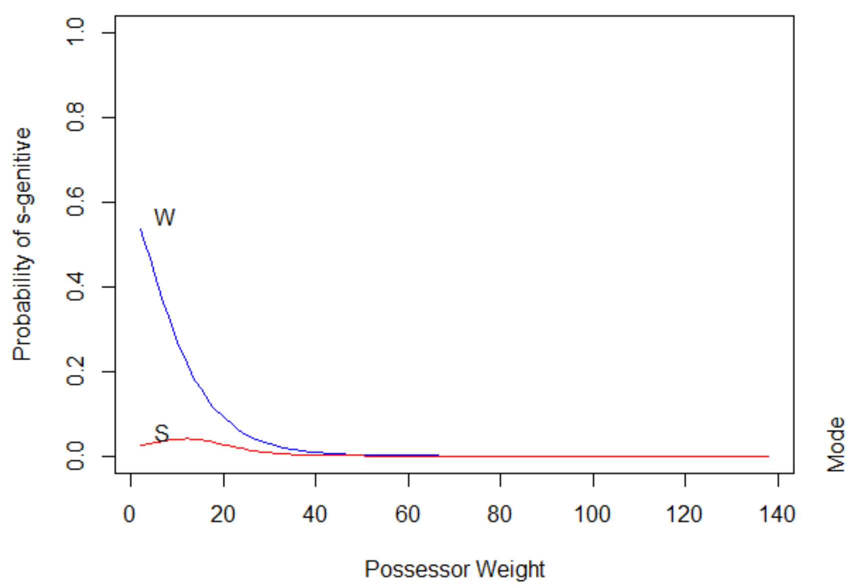


Figure 14: Interaction of MODE *with* POSSESSOR WEIGHT*.*

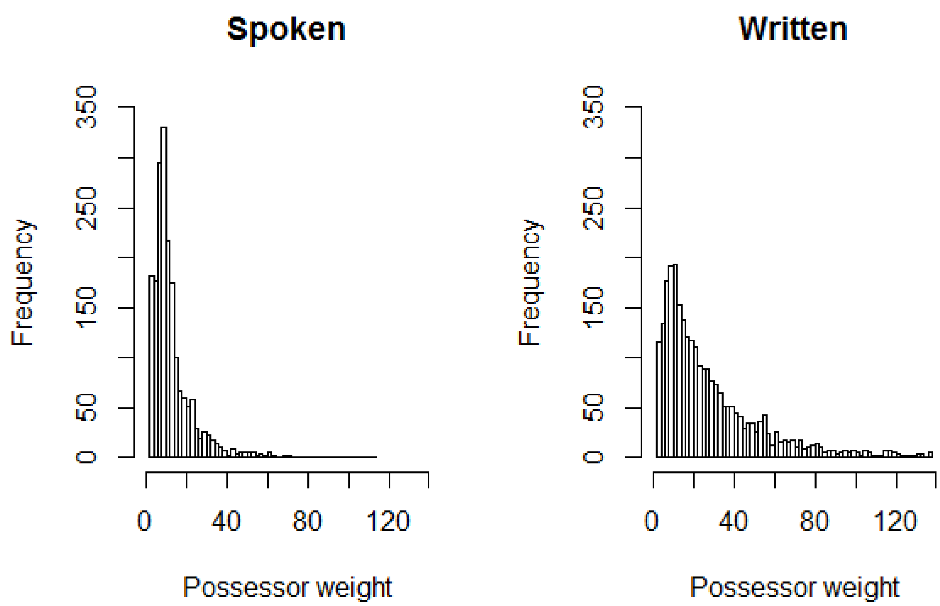


Figure 15: Frequency distribution of POSSESSOR WEIGHT *across spoken (left) and written (right) data.*

Furthermore, Wolk et al. (2013) found that the when both constituents are less than 8-12 characters long, the end-weight principle no longer applies. Many of the QuakeBox tokens fall into this category (e.g. (56)), with both possessum and possessor very short, so for these genitives, it would be expected that weight will not be a factor, and instead other factors such as animacy will be more important in determining genitive choice.

(56)

a. the boot of my car (QuakeBox NB167_Shaun)

b. David's legs (QuakeBox QB469LJ_MichelleHarrison)

c. my wife's friends (QuakeBox AP2516_JasonEager)

d. God's timing (QuakeBox NB176_ColleenPounsford)

e. the start of uni (QuakeBox WF2212_IvanIgnatov_A_ENG)

Modelling the QuakeBox and *Press* datasets separately (Table 8 and Table 9) allows for further investigation of the ways MODE interacts with the other factors affecting genitive choice. ANIMATE x POSSESSUM WEIGHT is a significant interaction for the *Press* data (Figure 16), but not for QuakeBox (as would be expected, given the scarcity of weight data for

QuakeBox). As in the full data model, in the *Press* model weight seems to have a greater effect when the possessor is inanimate, with the of-genitive favoured with short possessums, and the s-genitive favoured when the possessum is long. This difference is not as marked in the *Press* data as it is in the full model, reflecting the lower overall effect of animacy in writing.

Figure 16: Interaction of ANIMATE *with* POSSESSUM WEIGHT *in the* Press *model.*

	Coefficient	SE	z	Pr(>|z|)	Sig
(intercept)	-1.591	0.968	-1.64	0.10	
ANIMATE = true	4.787	0.407	11.75	<0.0001	***
SIBILANT = true	-1.370	0.359	-3.82	0.0001	***
PROTOTYPICAL = true	0.030	0.516	0.06	0.95	
POSSESSUM WEIGHT	-0.150	0.107	-1.40	0.16	
POSSESSUM WEIGHT’	0.212	0.122	1.74	0.08	
POSSESSOR WEIGHT	0.072	0.055	1.32	0.19	
POSSESSOR WEIGHT’	-0.241	0.100	-2.42	0.02	*
GENITIVE MATCH = true	-1.663	0.376	-4.42	<0.0001	***
log(DISTANCE)	-0.210	0.107	-1.95	0.05	
POSSESSOR MATCH = true	-3.404	1.083	-3.14	0.002	**
PRIME TYPE = β	2.261	0.400	5.66	<0.0001	***
ANIMATE = true x PROTOTYPICAL = true	3.174	0.650	4.89	<0.0001	***
GENITIVE MATCH = true x POSSESSOR MATCH = true	3.842	1.103	3.48	0.0004	***
GENITIVE MATCH = true x PRIME TYPE = β	-4.852	0.573	-8.46	<0.0001	***
POSSESSOR MATCH = true x PRIME TYPE = β	2.555	0.901	2.84	0.005	**

* significant at p <.05, ** p <.01, *** p <.001

Table 8: Binomial mixed-effects model fit to the QuakeBox data, with GENITIVE TYPE as the dependent variable.

	Coefficient	SE	z	Pr(>|z|)	Sig
(intercept)	1.566	0.294	5.32	<0.0001	***
ANIMATE = true	3.046	0.242	12.574	<0.0001	***
SIBILANT = true	-1.324	0.130	-10.16	<0.0001	***
PROTOTYPICAL = true	0.089	0.122	0.73	0.46	
POSSESSUM WEIGHT	0.713	0.019	8.88	<0.0001	***
POSSESSUM WEIGHT’	-0.156	0.033	-4.78	<0.0001	***
POSSESSOR WEIGHT	-0.147	0.012	-12.52	<0.0001	***
POSSESSOR WEIGHT’	0.050	0.028	1.76	0.08	
GENITIVE MATCH = true	-2.443	0.321	-7.61	<0.0001	***
log(DISTANCE)	-0.468	0.044	-10.65	<0.0001	***
POSSESSUM MATCH = true	-7.916	1.462	-5.41	<0.0001	***
POSSESSOR MATCH = true	0.691	0.134	5.16	<0.0001	***
PRIME TYPE = β	2.685	0.118	22.72	<0.0001	***
SIBILANT = true x ANIMATE = true	0.094	0.172	0.55	0.58	
ANIMATE = true x PROTOTYPICAL = true	0.553	0.213	2.60	0.009	**
ANIMATE = true x POSSESSUM WEIGHT	-0.098	0.030	-3.32	0.0009	***
ANIMATE = true x POSSESSUM WEIGHT’	0.093	0.048	1.96	0.05	
SIBILANT = true x PROTOTYPICAL = true	1.024	0.307	3.33	0.0009	***
GENITIVE MATCH = true x log(DISTANCE)	0.427	0.065	6.57	<0.0001	***
GENITIVE MATCH = true x POSSESSUM MATCH = true	6.692	1.875	3.57	0.0004	***
GENITIVE MATCH = true x PRIME TYPE = β	-5.039	0.162	-31.03	<0.0001	***
log(DISTANCE) x POSSESSUM MATCH = true	1.394	0.323	4.32	<0.0001	***
ANIMATE = true x SIBILANT = true x PROTOTYPICAL = true	-0.958	0.432	-2.22	0.03	*
GENITIVE MATCH = true x log(DISTANCE) x POSSESSUM MATCH = true	-1.185	0.397	-2.99	0.003	**

* significant at p <.05, ** p <.01, *** p <.001

Table 9: Binomial mixed-effects model fit to the Press *data, with* GENITIVE TYPE *as the dependent variable.*

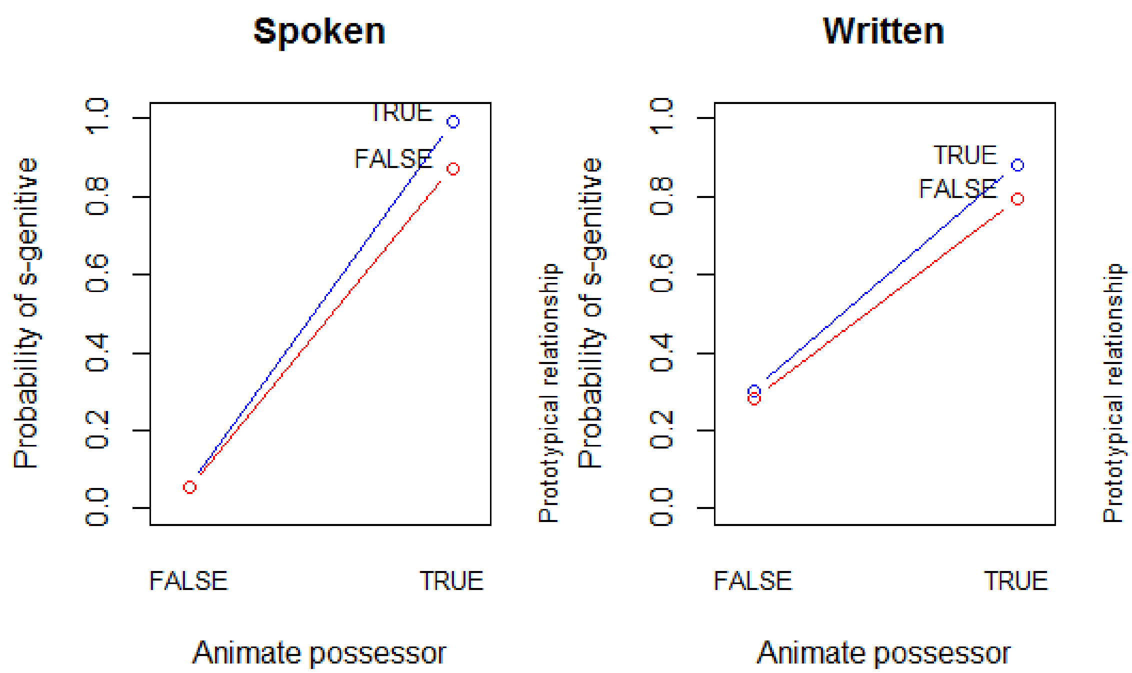


Figure 17: Interaction of ANIMATE *and* PROTOTYPICAL *in in QuakeBox (left) and* Press *(right) models. Lines have been added to the graphs in order to make the relative slopes clearer.*

The interaction between ANIMATE and PROTOTYPICAL is significant in both models (Figure 17). When the possessor is animate, the s-genitive is more strongly favoured when the relationship is prototypical, but this is not the case when the possessor is inanimate. The effect of animacy is again stronger in speech than in writing.

No other significant interactions among factors influencing genitive choice were found in the QuakeBox model. However, in the *Press* model there is a significant interaction between ANIMATE, SIBILANT and PROTOTYPICAL (Figure 18). When the relationship between possessor and possessum is prototypical, animate possessors ending in a sibilant are less likely to appear in an s-genitive than animate possessors which don't end in a sibilant. Where the possessor is inanimate, final sibilance makes less difference to genitive choice. In a non-prototypical relationship, animacy makes little difference to the effect of sibilance.

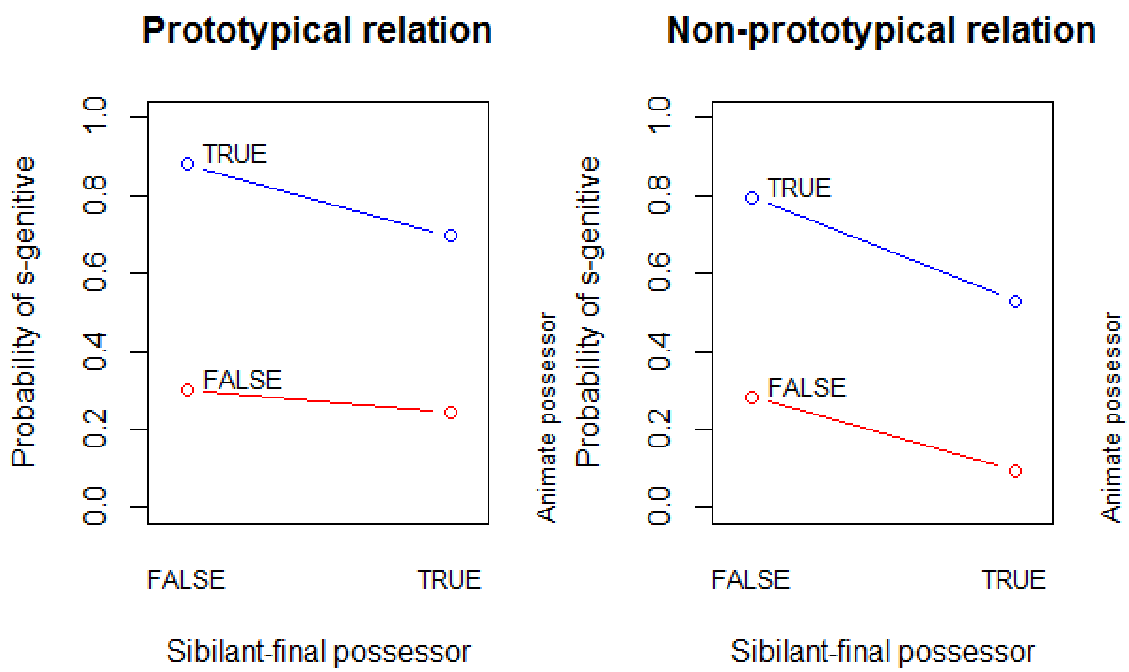


Figure 18: Interaction of SIBILANT, ANIMATE *and* PROTOTYPICAL *in the* Press *model. Lines have been added to the graphs in order to make the relative slopes clearer.*

4.2 Factors influencing priming

This section discusses GENITIVE MATCH, i.e. the effect of priming on genitive choice. As was seen in Table 5 (Section 4.1), GENITIVE MATCH remains significant after all other factors influencing genitive choice are taken into account by the model, meaning that priming is indeed present in the data. The factors influencing the strength of this priming effect will also be discussed in this section.

Around half (56%) of the genitives in the full dataset match their preceding genitives (Table 10). The proportion of matching genitives is much higher in the QuakeBox data (73.6%) than in the *Press* data (52%) ($\chi^2 = 413$, $p < 0.0001$).

	Match		No match		Total
	N	%	N	%	
Press	5736	52.0	5287	48.0	11023
QuakeBox	2001	73.6	718	26.4	2719
Full dataset	7737	56.3	6005	43.7	13742

Table 10: Proportion of genitives matching the immediately preceding genitive.

	Coefficient	SE	z	Pr(>|z|)	Sig
(intercept)	1.642	0.211	7.77	<0.0001	***
log(DISTANCE)	-0.054	0.033	-1.63	0.10	
POSSESSUM MATCH = true	-0.994	0.894	-1.11	0.27	
POSSESSOR MATCH = true	1.963	0.656	2.99	0.003	**
PRIME TYPE = β	1.812	0.171	10.62	<0.0001	***
MODE = written	-1.046	0.116	-9.00	<0.0001	***
GENITIVE TYPE = S	-3.532	0.308	-11.49	<0.0001	***
PROTOTYPICAL = true	0.363	0.096	3.77	0.0001	***
WEIGHT RATIO	-0.034	0.009	-3.62	0.0003	***
log(DISTANCE) x GENITIVE TYPE = S	0.279	0.044	6.31	<0.0001	***
log(DISTANCE) x POSSESSOR MATCH = true	-0.237	0.122	-1.94	0.05	
log(DISTANCE) x POSSESSUM MATCH = true	0.982	0.165	5.95	<0.0001	***
POSSESSUM MATCH = true x MODE = written	-2.779	0.663	-4.19	<0.0001	***
POSSESSUM MATCH = true x POSSESSOR MATCH = true	7.409	2.518	2.94	0.003	**
POSSESSOR MATCH = true x GENITIVE TYPE = S	1.041	0.225	4.622	<0.0001	***
PRIME TYPE = β x MODE = written	0.642	0.193	3.33	0.0009	***
PRIME TYPE = β x GENITIVE TYPE = S	-3.775	0.286	-13.19	<0.0001	***
MODE = written x GENITIVE TYPE = S	1.760	0.192	9.15	<0.0001	***
GENITIVE TYPE = S x PROTOTYPICAL = true	-0.427	0.130	-3.29	0.00099	***
GENITIVE TYPE = S x WEIGHT RATIO	0.088	0.035	2.54	0.01	*
log(DISTANCE) x POSSESSUM MATCH = true x POSSESSOR MATCH = true	-1.513	0.468	-3.24	0.001	**
PRIME TYPE = β x MODE = written x GENITIVE TYPE = S	-1.228	0.314	-3.91	<0.0001	***

* significant at p <.05, ** p <.01, *** p <.001

Table 11: Binomial mixed-effects model fit to the entire dataset, with GENITIVE MATCH as the dependent variable.

Table 11 shows the binomial mixed effects model fit over the full dataset, with GENITIVE MATCH as the dependent variable. This model predicts the likelihood that a genitive in a choice condition will match the immediately preceding genitive, i.e. be primed. Factors with a positive coefficient indicate conditions which are more likely to produce priming, and factors with a negative coefficient indicate those likely to inhibit priming.

The model in Table 11 includes a significant main effect for POSSESSOR MATCH, meaning that a genitive is more likely to match its predecessor when the possessors of the two genitives also match. However, POSSESSUM MATCH failed to reach significance as a main effect. This is the opposite to what would be expected if, as Cleland and Pickering (2003) suggested, a match between the heads of the NPs increases the strength of priming. However, Szmrecsanyi (2006) suggests another possibility. He found a main effect for POSSESSOR MATCH and POSSESSUM MATCH on genitive choice, with the s-genitive being more likely to be used when the possessor matches the previous genitive, and the of-genitive being more likely when it is the possessums that match. He interprets this as the effect of information status, with discourse-old information being placed earlier in the NP than new information (Altenberg, 1982). When POSSESSOR MATCH is true, the possessor is discourse-old information, so comes first, generating an s-genitive. Conversely, when POSSESSUM MATCH is true, an of-genitive will be used to ensure that the discourse-old possessum comes first. This explanation fits well with my data. The model in Table 5 (Section 4.1) shows a significant main effect for both POSSESSOR MATCH and POSSESSUM MATCH, with the directions of the slopes being as would be predicted if the factors are indeed measuring information status (Figure 19).

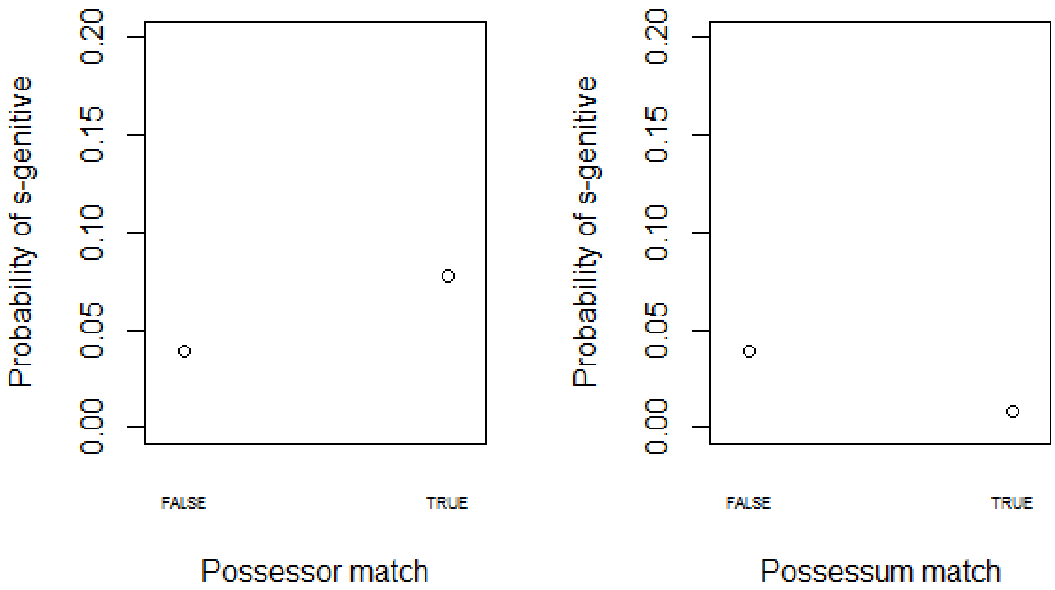


Figure 19: Effect of POSSESSOR MATCH *and* POSSESSUM MATCH *on genitive choice.*

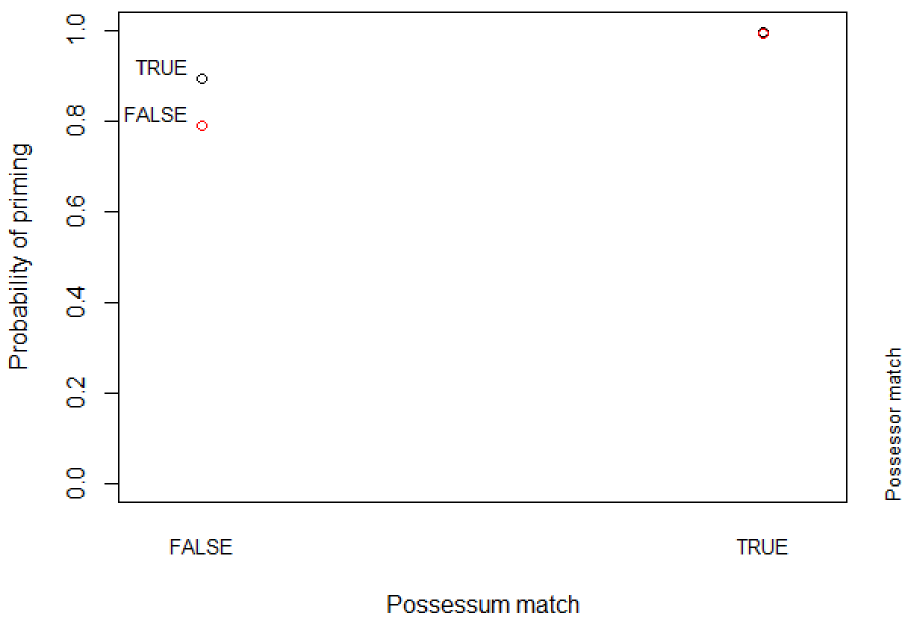


Figure 20: Interaction of POSSESSOR MATCH *and* POSSESSUM MATCH.

It is also worth keeping in mind that when possessors or possessums match, many of the factors which influence genitive choice will also match. Figure 20 shows the interaction between POSSESSOR MATCH and POSSESSUM MATCH in the priming model: when both possessor and possessum match, the likelihood of the genitive structure also matching approaches certainty. While this could be a result of priming, an examination of examples such as (57) and (58) shows that when possessum and possessor match, so do many of the conditions affecting genitive choice, such as animacy and the prototypicality of the relationship. If these conditions match, then naturally the likelihood is that the genitives will match as well. Further supporting this interpretation is the fact that it is precisely the effect of POSSESSOR MATCH on GENITIVE MATCH which reaches significance, while that of POSSESSUM MATCH does not. It is the possessor which carries animacy, which is the strongest factor affecting genitive choice, so the matching of the animacy condition will of course have an effect on the likelihood of the genitive matching[12].

(57) mum went to **my sister's place** which is really isolated and of course she could be there but **my sister's place** was not the place for mum and she just slowly petered away there

(QuakeBox EG111_InaWit)

(58) Reay said he probably would, as long as he knew there would be **a peer review of the work**.

The only review of Harding's work at the time was that done by the city council.

(Joelle Dally, 'Firm's owner asked to do homework', *Press*, 2 August 2012)

[12] It might be expected that if this is the case, the matching conditions would show up in the model as, for example, an interaction between POSSESSOR MATCH and ANIMATE. However the scarcity of data for POSSESSOR MATCH and POSSESSUM MATCH is likely to have prevented any such interactions, if present, from reaching significance.

	Coefficient	SE	z	Pr(>|z|)	Sig
(intercept)	2.524	0.436	5.79	<0.0001	***
log(DISTANCE)	-0.225	0.074	-3.04	0.002	**
POSSESSUM MATCH = true	2.917	0.664	4.39	<0.0001	***
POSSESSOR MATCH = true	2.171	0.355	6.12	<0.0001	***
PRIME TYPE = β	0.407	0.658	0.62	0.53	
GENITIVE TYPE = S	-2.809	0.652	-4.31	<0.0001	***
ANIMATE = true	-0.272	0.335	-0.81	0.42	
SIBILANT = true	0.193	0.213	0.91	0.36	
WEIGHT RATIO	-0.006	0.038	-0.17	0.87	
log(DISTANCE) x PRIME TYPE = β	0.260	0.116	2.24	0.03	*
PRIME TYPE = β x GENITIVE TYPE = S	-3.689	0.298	-12.40	<0.0001	***
GENITIVE TYPE = S x ANIMATE = true	0.701	0.689	1.02	0.31	
GENITIVE TYPE = S x SIBILANT = true	1.779	1.142	1.56	0.12	
GENITIVE TYPE = S x WEIGHT RATIO	0.251	0.095	2.66	0.008	**
ANIMATE = true x SIBILANT = true	-0.105	0.592	-0.18	0.86	
GENITIVE TYPE = S x ANIMATE = true x SIBILANT = true	-2.854	1.332	-2.14	0.03	*

* significant at p <.05, ** p <.01, *** p <.001

Table 12: Binomial mixed-effects model fit to the QuakeBox data, with GENITIVE MATCH as the dependent variable.

	Coefficient	SE	z	Pr(>|z|)	Sig
(intercept)	0.694	0.196	3.54	0.0004	***
log(DISTANCE)	-0.075	0.037	-2.00	0.046	*
POSSESSUM MATCH = true	-5.718	0.907	-6.30	<0.0001	***
POSSESSOR MATCH = true	1.706	0.846	2.02	0.044	*
PRIME TYPE = β	2.450	0.091	27.06	<0.0001	***
GENITIVE TYPE = S	-2.165	0.251	-8.61	<0.0001	***
ANIMATE = true	-0.009	0.094	-0.10	0.92	
SIBILANT = true	0.048	0.084	0.57	0.57	
PROTOTYPICAL = true	0.365	0.106	3.43	0.0006	***
WEIGHT RATIO	-0.034	0.009	-3.61	0.0003	***
log(DISTANCE) x GENITIVE TYPE = S	0.374	0.049	7.60	<0.0001	***
log(DISTANCE) x POSSESSUM MATCH = true	1.411	0.217	6.51	<0.0001	***
log(DISTANCE) x POSSESSOR MATCH = true	-0.213	0.160	-1.33	0.18	
POSSESSUM MATCH = true x POSSESSOR MATCH = true	4.849	2.604	1.86	0.06	
POSSESSUM MATCH = true x GENITIVE TYPE = S	7.388	1.714	4.31	<0.0001	***
POSSESSOR MATCH = true x GENITIVE TYPE = S	1.051	0.244	4.32	<0.0001	***
PRIME TYPE = β x GENITIVE TYPE = S	-5.012	0.131	-38.36	<0.0001	***
GENITIVE TYPE = S x ANIMATE = true	0.094	0.125	0.76	0.45	
GENITIVE TYPE = S x SIBILANT = true	-0.646	0.237	-2.73	0.006	**
GENITIVE TYPE = S x PROTOTYPICAL = true	-0.375	0.142	-2.65	0.008	**
ANIMATE = true x SIBILANT = true	-0.042	0.160	-0.26	0.80	
log(DISTANCE) x POSSESSUM MATCH = true x POSSESSOR MATCH = true	-0.972	0.494	-1.97	0.049	*
log(DISTANCE) x POSSESSUM MATCH = true x GENITIVE TYPE = S	-1.493	0.357	-4.18	<0.0001	***
GENITIVE TYPE = S x ANIMATE = true x SIBILANT = true	0.637	0.295	2.15	0.03	*

* significant at p <.05, ** p <.01, *** p <.001

Table 13: Binomial mixed-effects model fit to the Press *data, with GENITIVE MATCH as the dependent variable.*

POSSESSUM MATCH cannot be ruled out entirely as an effect on priming, however. When models were fit to the *Press* data and QuakeBox data independently (Table 12 and Table 13), both showed significant main effects for POSSESSUM MATCH (Figure 21). Thus there may be some evidence for matching heads increasing priming strength. This result should be treated with caution, though, as the frequency of POSSESSUM MATCH and POSSESSOR MATCH in both datasets is extremely low (Table 14), so the effect found by the model is based on very little data. Therefore no firm conclusion can be reached on the question of whether the lexical boost effect reported for other alternating structures is also present for the genitive alternation.

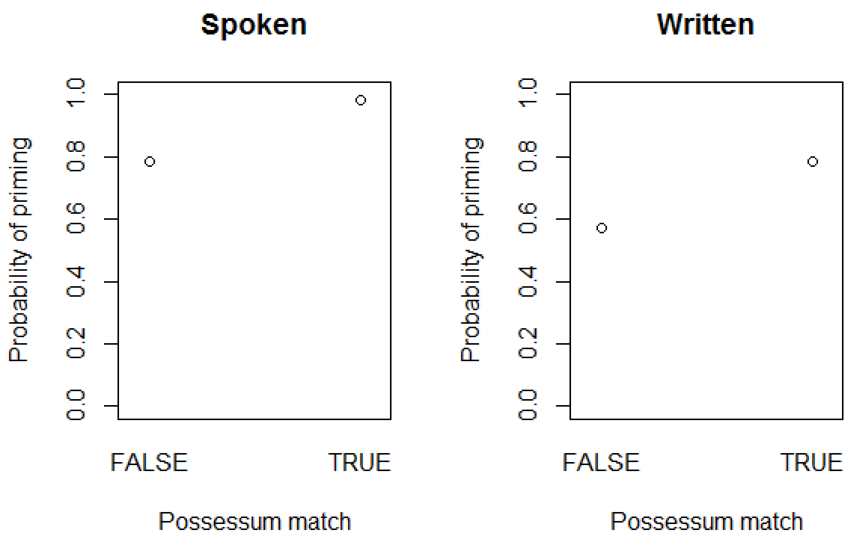


Figure 21: Effect of POSSESSUM MATCH *on priming in the QuakeBox (left) and* Press *(right) models.*

	Possessor matches	Possessum matches	Both match	Neither matches	Total[13]
Press	575	281	90	10257	11023
QuakeBox	184	149	71	2457	2719
Full dataset	759	430	161	12714	13742

Table 14: Frequencies of possessor and possessum matches across datasets.

[13] Note that the counts for POSSESSOR MATCH and POSSESSUM MATCH include those cases where both match, so the total number of tokens is less than the sum across the columns.

If the effect of GENITIVE MATCH does indeed indicate the presence of syntactic priming, then it would be expected that the priming effect would decay logarithmically (cf. Gries, 2005; Szmrecsanyi, 2006); that is, there should be a linear relationship between GENITIVE MATCH and log(DISTANCE). Contrary to this expectation, log(DISTANCE) failed to reach significance as a main effect in the model presented in Table 11. However, in the GENITIVE TYPE model presented above (Table 5, in Section 4.1) the interaction between GENITIVE MATCH and log(DISTANCE) did reach significance. This suggests that DISTANCE is indeed affecting GENITIVE MATCH, as would be expected, but that the main effect is perhaps being hidden by the interaction of distance with other factors.

For example, there is an interaction between log(DISTANCE), POSSESSUM MATCH and POSSESSOR MATCH (Figure 22). When possessums do not match, then priming decreases approximately linearly with the log of distance, as would be expected. However, if both possessors and possessums match, then it appears that distance has almost no effect on priming, and when the possessum matches but the possessor does not, then increasing distance appears to actually *increase* the likelihood of priming.

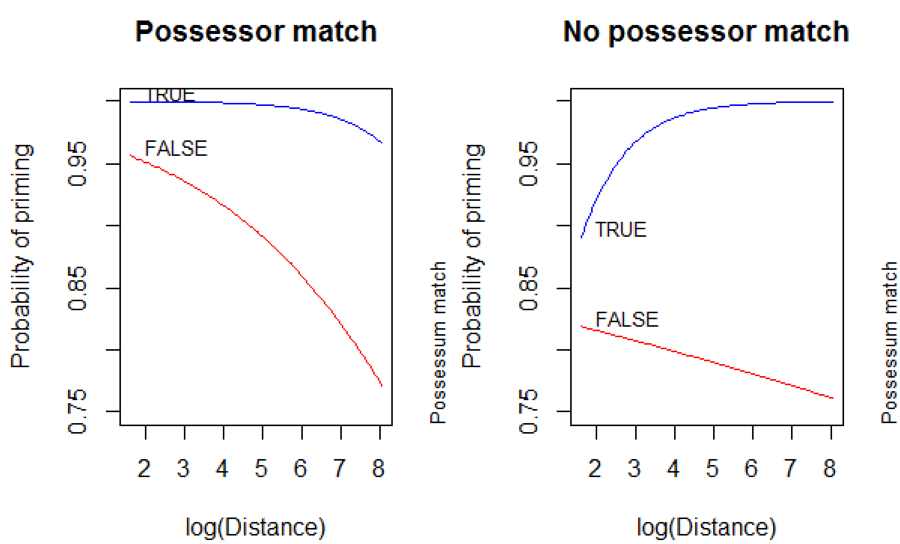


Figure 22: Interaction between POSSESSUM MATCH, POSSESSOR MATCH, and log(DISTANCE). Note that the vertical scale has been adjusted to better show the slopes.

This seemingly unlikely result can perhaps be explained by the low frequencies for POSSESSUM MATCH and POSSESSOR MATCH seen in Table 14. The matches occur relatively rarely, meaning that except in the case where neither possessums nor possessors match, the curves are based on very little data. This is especially so around the end points, as the distribution of log distances (Figure 23) is approximately normal, with most tokens clustering at the centre, and very few at the extremes. Thus little can be deduced about the true shape of the relationship of priming to distance when either possessum or possessor matches.

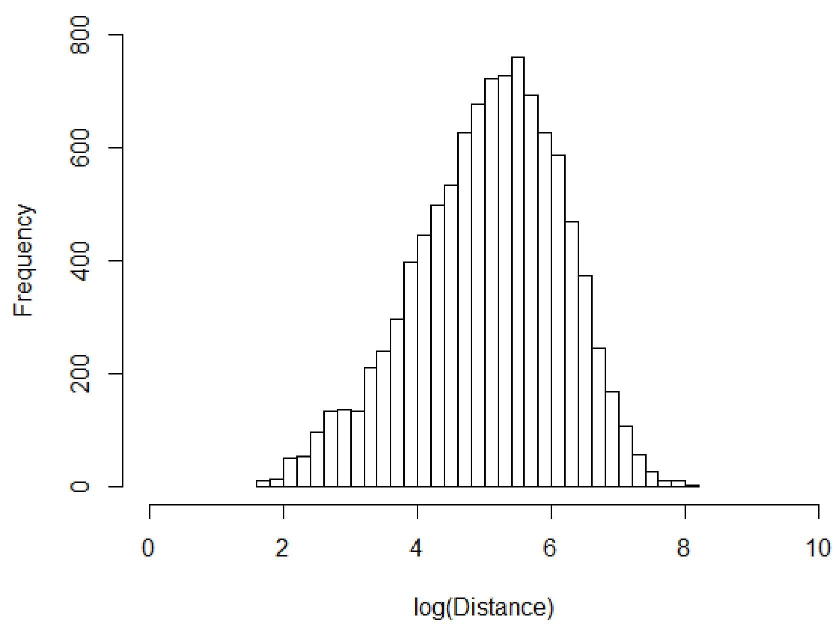


Figure 23: Distribution of log distances across the full dataset.

When the QuakeBox and *Press* data are modelled separately, log(DISTANCE) does emerge as a main effect in each (although in the *Press* model, log(DISTANCE) is only significant at the $p < .05$ level), and approximates the expected linear relationship with GENITIVE MATCH which would indicate priming (Figure 24). Branigan, Pickering, and Cleland (1999) suggested that priming in writing decays more rapidly than in speech. Figure 24 would seem to suggest that the opposite is true in my data, but as the interaction between MODE and log(DISTANCE) did not reach significance, no meaningful comparison can be made of the effect of distance in the two corpora[14].

---

[14] It should also be noted that although I have shown that textual distance is closely correlated to temporal distance in the QuakeBox data, it is unlikely that the relationship between textual distance and temporal distance in the *Press* data is identical to that in the QuakeBox data. As speech is generally more rapid than writing, it is

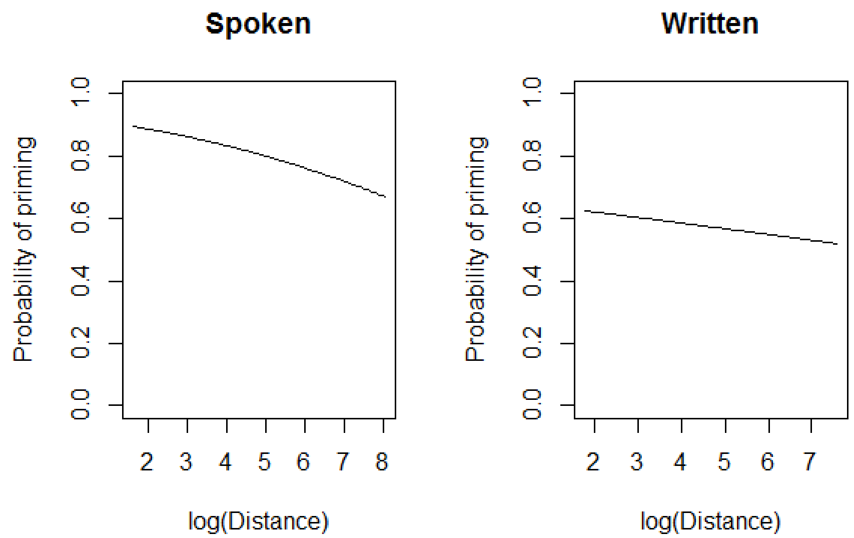


*Figure 24: Effect of log(*DISTANCE*) on priming in the QuakeBox (left) and* Press *(right) models.*

Log(DISTANCE) also interacts with GENITIVE TYPE (Figure 25). Of-genitives show the expected decay relationship, with the probability of priming reducing as the log of distance increases. However, s-genitives have the opposite relationship, with the probability of priming *increasing* with distance.

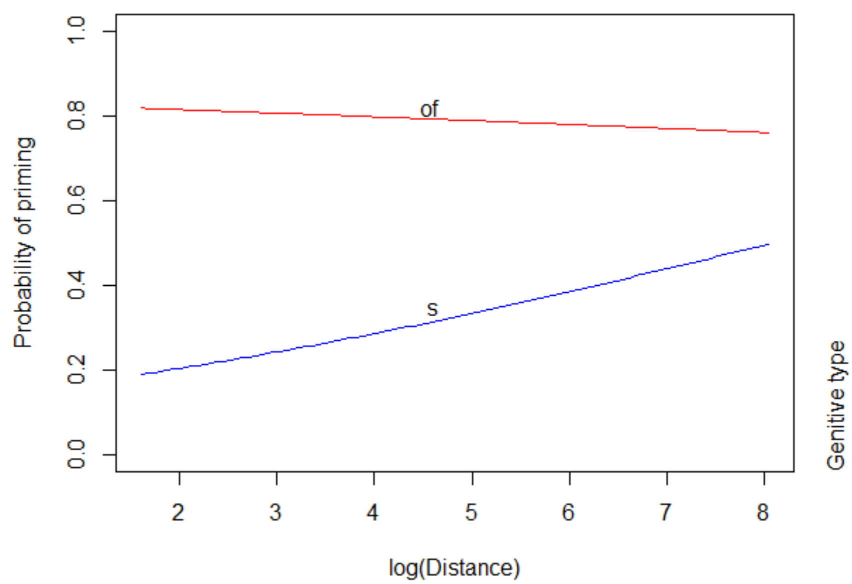


Figure 25: Interaction of GENITIVE TYPE *and log(*DISTANCE*).*

---

likely that the actual elapsed time represented by a particular textual distance in the QuakeBox data is lower than that represented by the same textual distance in the *Press*. Therefore care must be taken in comparing the x-axis scales of the two graphs in Figure 24.

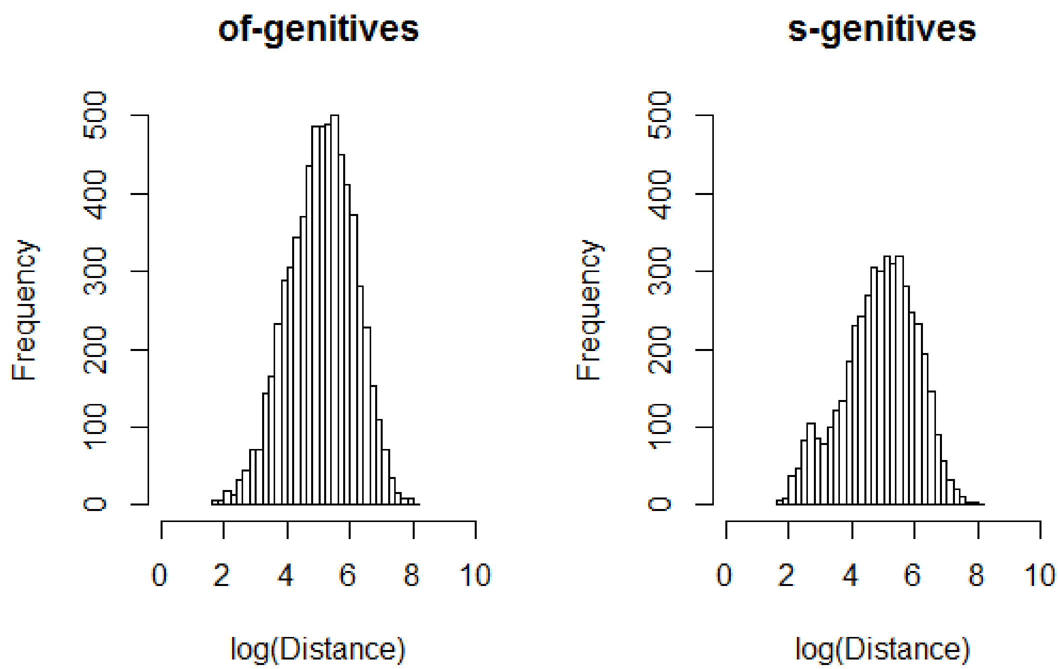


Figure 26: Comparison of weight distribution for of-genitives and s-genitives.

This anomaly cannot be attributed to the distribution of distances in the data, as of-genitives and s-genitives have the same distribution (Figure 26: m_{of} = 5.07, sd_{of} = 1.03; m_s = 4.88, sd_s = 1.14; $t = 9.0$, $p < .001$).

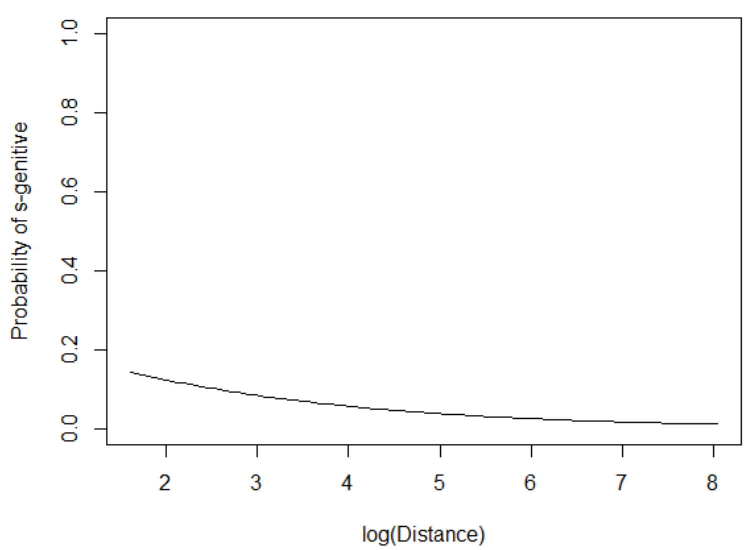


Figure 27: Decrease in s-genitive probability with log(DISTANCE).

Table 5 (in Section 4.1) shows a main effect for log(DISTANCE), with a negative slope, meaning that, having accounted for all other factors, s-genitives become less likely as distance increases (Figure 27). There seems to be no reason why genitive choice should be sensitive to the distance from the previous genitive, but looking at an example of a token with a large distance score (59) may provide a clue. *Parker's argument* has a textual distance of 1374[15] characters from the previous full-NP genitive, *Parker's comments*. However, there are five pronominal possessors intervening between the two s-genitives, with the nearest at a distance of only 19 characters.

(59) Organiser Nigel Salsbury said yesterday **Parker's comments** gave him hope the council was prepared to listen.

"It does make me think [the campaign] we launched is for a worthwhile cause. It's not something unjust, we have a damn good case," he said.

"What [Parker's] saying is, 'I was comfortable with [the decision], but if this is brought to **our attention**, it's a bigger issue than what we thought it was'. That's great news rather than just being told, 'no, no no'."

Parker said the "fundamental difference" between homeowners receiving a full rates remission, such as those in rockfall-threatened properties, was that the red-zoners could access **their property** and had "things they could do there".

Salsbury said owning an uninhabitable home had no benefits.

"If you had a home that was broken and there are no services there whatsoever, why would you gain access to that property?

"It's uninhabitable for a damn good reason - the roof's fallen off, the walls are fallen down, you've got a six-inch crack through the floor. Yes, I still have a key, yes, I can still go to **my property**, but why would I want to do that?"

---

[15] Due to the way paragraph breaks are marked up in the *Press* database, this calculated figure is slightly higher than would be obtained by manually counting characters.

The campaign had attracted support from those outside the red zone, Salsbury said.

"I'm getting just as much, if not more support, from people still able to live in **their homes**."

Christchurch East MP Lianne Dalziel said yesterday she did not believe the council wanted to review **its decision**.

Parker's argument for the distinction between 40 per cent and 100 per cent relief was "weak", she said.

(Marc Greenhill, 'Parker offers hope on rates', *Press*, 16 April 2012)

Although I argued in Section 3 above that possessive pronouns cannot prime s-genitives, as they have a different syntactic structure[16] (Lowe, 2016), there may be more than just syntactic priming in play. Bock and Loebell (1990) showed that syntactic priming does not occur when the underlying structures differ, even if the phrases produced are superficially similar. However, in a study comparing priming of the passive structure in English and Russian, Vasilyeva and Waterfall (2012) found results suggesting that it is not only the syntactic structure that is primed, but also the order of the thematic roles, independent of the syntax[17]. If this is the case, then this may mean that possessive pronouns can indeed prime s-genitives to some extent, as they share the same ordering of thematic roles.

---

[16] Other approaches to the syntax of genitives have argued that s-genitives and possessive pronouns do share the same syntactic structure. For example, Adger (2003, p. 274) places both full-NP possessors and possessive pronouns in the specifier of DP. Under this approach, possessive pronouns would be expected to syntactically prime s-genitives.

[17] If this priming of thematic role ordering can occur, then the possibility arises of priming between s-genitives and nominal compounds that express possessive relationships, such as (i), where *Avondale School principal* could equally have been expressed as *Avondale School's principal*, or *(the) principal of Avondale School*.

i. **Avondale School principal** Mark Scown
(Tina Law, 'Student numbers remain in freefall, teacher jobs to go', *Press*, 8 September 2012)

This suggests that in addition to factoring in the effect of possessive pronouns, a full account of s-genitive priming should perhaps also take into account intervening nominal compounds which express possessive relationships.

In (59), then, it may not be the distant *Parker's comments* that is priming the s-genitive in *Parker's argument*, but instead the much closer *its decision*. Alternatively, the syntactic priming from *Parker's comments* may have been strengthened by the thematic priming from each of the intervening possessive pronouns, so that the priming effect decayed more slowly. Investigating the effect of possessive pronouns and thematic priming on the strength of s-genitive priming is beyond the scope of this study, but it suggests an opportunity for future research.

At the other end of the distance scale, very short distances between genitives are often due to the genitives being nested, as in (60).

(60)
  a. our mate's dad's work (QuakeBox UC011AM_PerryHyde)

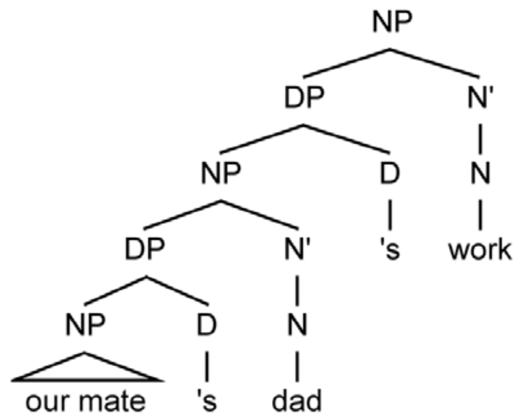
  b. the work of the dad of our mate

In this study, nested genitives were treated as if there were two distinct genitive structures, able to prime each other in the same way as successive non-nested genitives. So in (60a), the prime is *our mate's dad*, and the target [*our mate's dad*]*'s work*, with a textual distance between the two of 6 characters.

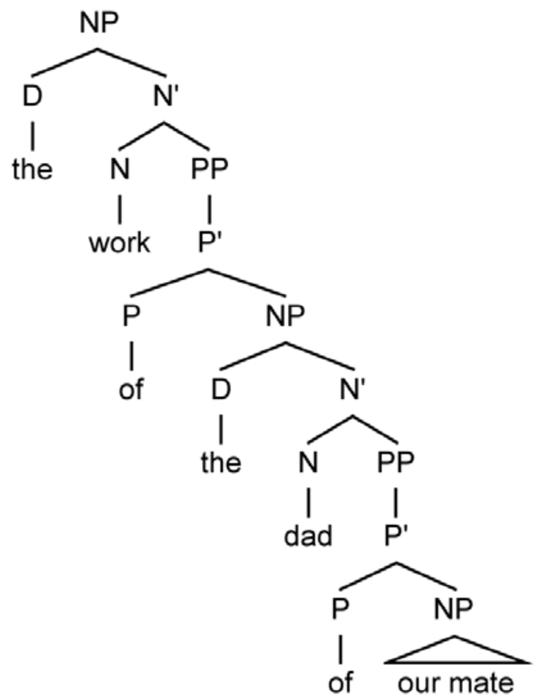
However, consider the alternative structure in (60b). Here, the assumption would be that *the work of* [*the dad of our mate*] is the prime, and the target *the dad of our mate*. So the initial choice of genitive appears to determine which structure will act as the prime, and which as the target – which seems counter-intuitive. Indeed, Pickering, Branigan, Cleland, and Stewart (2000) suggest that in nested structures, priming may not be a linear process, but instead hierarchical, with a matrix structure able to prime child structures, independent of distance. For nested s-genitives, this would imply that the direction of priming can be reversed, with the target being produced before the prime, as can be seen in (61), where the genitive [*our mate's dad*]*'s work* is hierarchically selected earlier than *our mate's dad*, although linearly it is produced later (I follow Lowe's (2016, p. 175) analysis of the s-genitive structure here, but the same argument can be applied to other proposed analyses, as it is the left-branching nature of the tree which is of importance, not the specific details of the analysis).

(61)

a.



b.



Whether or not this interpretation of priming in nested structures is correct, it is clear that priming within nested structures is more complex than my coding system allows for, so it is likely that my model does not accurately depict the true situation at the lower end of the distance scale.

Figure 28 shows the interactions between PRIME TYPE, GENITIVE TYPE and MODE. The interaction between PRIME TYPE and GENITIVE TYPE is of particular interest. If the genitive is *'s*, then the strength of α-priming is stronger than that of β-priming, but if the genitive is *of*, then the effect is reversed, and β-priming is stronger than α-priming.

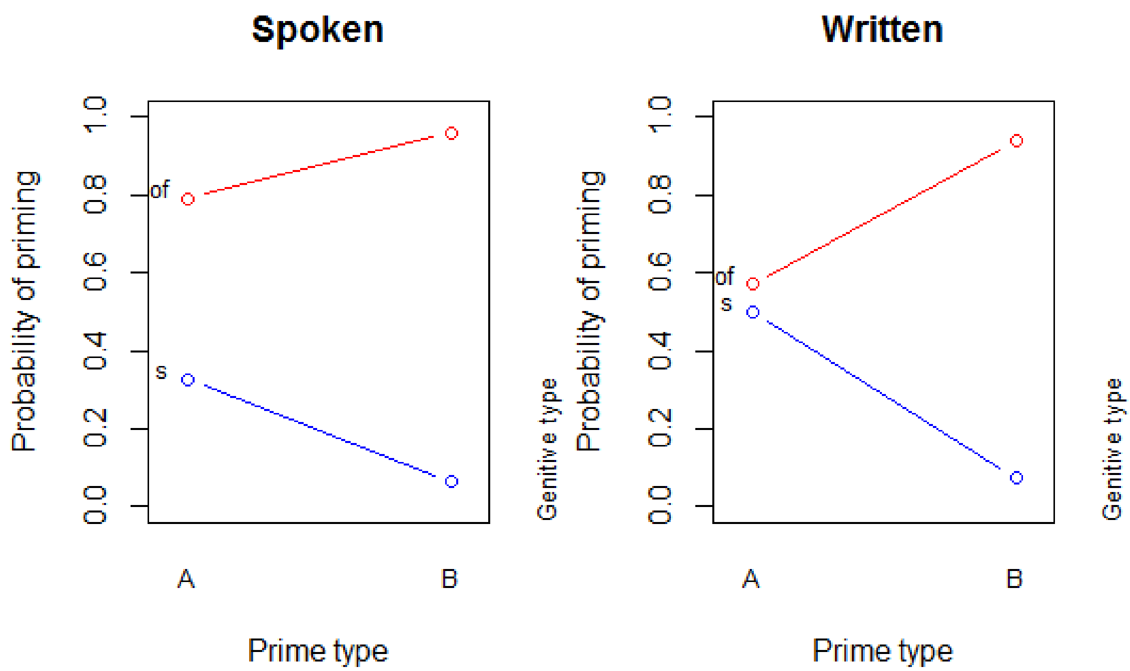


Figure 28: Interaction between PRIME TYPE, GENITIVE TYPE and MODE.

A possible explanation for this interaction arises from a consideration of what is being activated in the α-priming and β-priming conditions. In an α-priming condition such as (62), the prime *the future of the city* is in a choice context: both options, *the future of the city* and *the city's future* are available to the speaker. Segaert, Menenti, Weber, and Hagoort (2011) suggest that when two syntactic structures are available to a speaker, both structures are activated during the selection stage of production, until eventually the activation of one of the structures passes a selection threshold, and is produced. According to this account, in (62), both the s-genitive and of-genitive structures were activated, but the of-genitive was first to reach the selection threshold level of activation, so the form produced was *the future of the city*. However, the activation of the alternative form, *the city's future*, remains, and may still be sufficient to allow a priming effect. That this is possible was shown experimentally by van Gompel, Pickering, Pearson, and Jacob (2006), who found that structures inappropriately activated by garden path sentences, although discarded after the ambiguity is resolved, remain activated, and are able to prime subsequent production.

(62) ... stories from Christchurch residents who shared their feelings about **the future of the city**.

At **the end of the service**, 185 monarch butterflies will be released ...

(Rachel Young, 'Musicians, hero volunteer have remembrance roles', *Press*, 20 February 2012)

In a β-priming condition, such as (63), however, the prime *a great symbol of hope* is in a categorical context, so only the of-genitive structure will be activated. Thus the level of activation of the priming structure in the β condition will be higher than in the α condition, where the activation is shared between the two structures. It should be expected, then, that β-priming will have a stronger effect than α-priming.

(63) I think we do need **a great symbol of hope** back in **the centre of our city**

(Steve Graham, 'Give people hopeful future', *Press*, 2 April 2012)

If this difference in activation levels does mean that priming is more likely in the β-priming condition, why then do s-genitives show the opposite pattern, with priming more likely in the α condition? For syntactic priming to occur, the two structures must be identical (Bock & Loebell, 1990). However, not all categorical genitives share the same structure as choice genitives.

Two types of s-genitives were classed as categorical in this study: classifying genitives such as (64), and fixed phrases such as (65). Taylor (1996) argues that classifying genitives, which he terms "possessive compounds" (p. 287), have a different syntactic structure to other s-genitives, being syntactically and semantically closer to nominal compounds such as *baby clothing* (p. 303).

(64) children's clothing

(Tina Law, 'Crafty women open new Purple Patch store', *Press*, 10 April 2012)

(65) Christchurch Women's Refuge

(Olivia Carville, 'Few options left for vulnerable women', *Press*, 30 June 2012)

The structure of these classifying genitives therefore cannot syntactically prime the s-genitive structure[18]. Thus some proportion of the target s-genitives which I have classed as being in β-priming conditions are not actually being primed (or, if they are, it is by a more distant genitive). I did not code categorical genitives by type, so it is not possible to determine what this proportion actually is. However, as will be noted in (65), many fixed phrases (although not all, e.g. (66)) are also classifying genitives, so will share this non-genitive structure. Thus it seems likely that the majority of "β-primed" s-genitives have not actually been primed by a categorical genitive structure.

(66) New Year's Eve

(Marc Greenhill, 'Wallets open as mullet hits the floor for charity', *Press*, 3 January 2012)

This same argument might apply to some categorical of-genitives. For example, Lehrer (1986) classifies quantitative genitives, such as (67), as pseudo-partitives. Lehrer failed to find compelling evidence that pseudo-partitives and of-genitives have a different syntactic structures, so interpreted the *of* in a pseudo-partitive as the head of a prepositional phrase PP. However, more recently it has been suggested that pseudo-partitive *of* in fact heads a functional phrase FP (e.g. Rutkowski, 2007; Stickney, 2009).

(67) the huge amount of infrastructure

(Alan Wood, 'Firms at terminal produce over $3b', *Press*, 7 August 2012)

If quantitative genitives do have a different structure than of-genitives, then quantitative genitives cannot act as β-primes. But there are other types of categorical of-genitives where the structure does match that of choice of-genitives. For example, (68) has an indefinite possessum, so is categorical, but has the same structure as an of-genitive in a choice condition. These indefinite of-genitives are therefore able to syntactically prime the of-

[18] The thematic priming effect discussed above will of course not have an effect here, as the 'possessor' and 'possessum' of a classifying genitive do not have the same thematic roles as an s-genitive expressing an actual possessive relationship.

genitive structure. Again, no accurate count is possible, but my impression from the raw data is that these indefinite of-genitives made up a large proportion of the genitives classed as categorical.

(68) a director of the commercial arm

(Alan Wood, ‘More “family” sought for firm’, *Press*, 3 July 2012)

If it is the case that a much greater proportion categorical of-genitives than s-genitives have identical structures to the choice genitives, this may explain the observed difference in direction of the effect of PRIME TYPE: categorical s-genitives are rarely able to have a β-priming effect, so the overall strength of effect is higher for α-priming. Categorical of-genitives, however, are more often able to prime, so the stronger activation level of β-priming can be detected.

This may also explain why Szmrecsanyi (2006) found a stronger effect for α-priming than β-priming. Szmrecsanyi’s criteria for β-priming constructions were broader than mine, considering any phrase containing the word *of* as a potential β-prime, while I restricted my β-priming candidates to the more genitive-like structures described by Rosenbach (2002) as genitives in categorical contexts. Thus, Szmrecsanyi will have had many more non-genitive structures potentially β-priming his targets, so it is unsurprising that he found a weaker overall effect for β-priming (just as I did for β-priming with s-genitives).

If, as I have suggested, β-primes activate the syntactic structure more strongly than α-primes, then it would be expected that α-primes would decay more rapidly. However, this is not what is seen in the interaction between PRIME TYPE and log(DISTANCE) found in the QuakeBox data. Figure 29 shows that α-priming decays logarithmically as expected, but DISTANCE appears to have no effect on β-priming.

This apparently contradictory result may of course be due to the non-genitive structures included in the β-prime data muddying the picture, as discussed above. Or, given that the interaction between PRIME TYPE and log(DISTANCE) failed to reach significance in the full dataset and in the *Press* data, it is possible that the interaction found in the QuakeBox data is simply a Type I error caused by overfitting the model. Further research restricting potential β-primes to only those categorical genitives which share a syntactic structure with choice-condition s-genitives or of-genitives may shed light on this issue of the decay rate of β-primes, but is beyond the scope of this study.

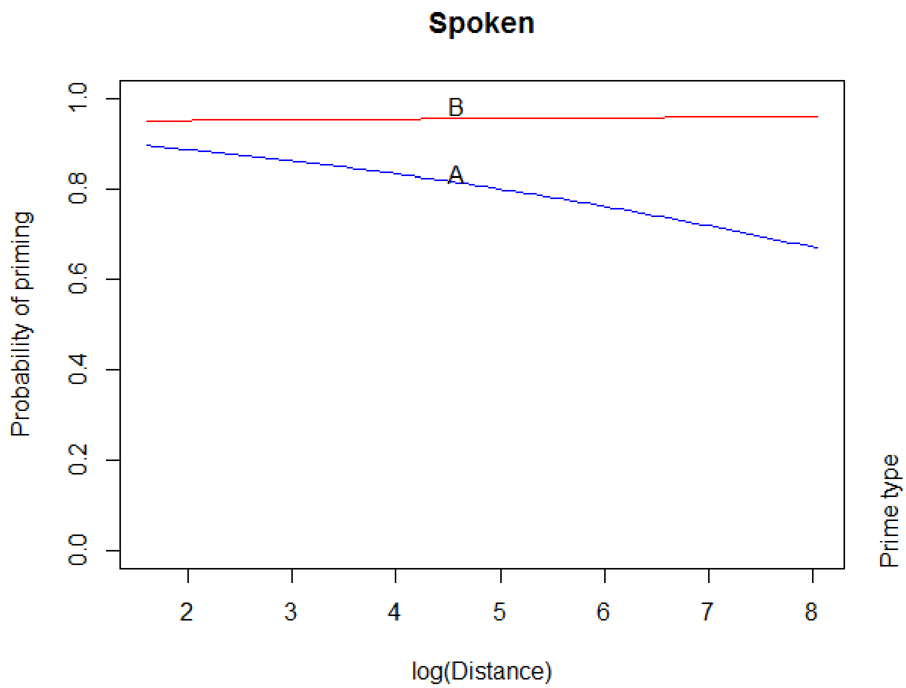


Figure 29: Interaction of PRIME TYPE *and log(*DISTANCE*) in QuakeBox data.*

Figure 28 above also shows an interaction with MODE. While the effect of β-priming is similar in the spoken and written data, in the α-priming condition, of-genitives are more likely to be primed in speech than in writing, while s-genitives are more likely to be primed in writing than in speech.

Szmrecsanyi (2006) found that syntactic priming decreases as the level of formality increases, and attributes this to the increased level of monitoring and planning in more formal registers, which leads to a conscious avoidance of repetition. This would suggest that there will also be less priming in writing than in speech, as writing is more closely monitored than speech (witness the absence in writing of the false starts and corrections that are a feature of spoken language) and allows more time for such stylistic choices to be made[19].

Szmrecsanyi (2006) also found that priming decreases in informationally dense environments, where the type-token ration (TTR) is high. I did not code for TTR in my data, so cannot say whether this is a contributing factor, but it seems reasonable to assume that newspaper writing, with its pressures of column-inch space, will by necessity be more informationally dense than the spoken language of the QuakeBox recordings, where participants were free to take as long as they wished to tell their stories.

Differences in formality and TTR between writing and speech therefore predict a stronger priming effect for speech than writing, and this is what is seen in Table 11, where there is a main effect for MODE, with priming more likely in speech than writing. Although this result does not seem to hold for s-genitives, as discussed above there are a number of issues with the way s-genitives were coded which may have skewed the results. Thus it can be reasonably concluded that the effect of priming is indeed stronger in speech.

---

[19] For β-priming, Szmrecsanyi conversely found an *increase* in more formal registers. He argues that this is because, unlike α-priming, β-priming cannot be consciously avoided. My data does not show an increase in β-priming in writing, but this may be due to the more restrictive way I have defined β-priming, as discussed above.

5 General Discussion

This study set out to discover whether syntactic priming is present in extended monologue, and whether there are differences between speech and writing. This chapter reviews the findings for each of these research questions, and discusses some of the implications. Section 5.3 discusses limitations of this study, and suggests areas for future research.

5.1 Is priming present in extended monologue?

Clear evidence for priming was found in both the QuakeBox and *Press* data, as well as in the combined dataset, confirming the existence of priming in extended monologue. However, conflicting evidence was found for factors expected to influence priming, such as distance, head matching, and α- versus β-priming.

In activation models of syntactic priming (e.g. Pickering & Branigan, 1998), lexical items and their syntactic and other properties are stored as nodes within an activation network. When a lexical item is activated, by being produced or comprehended, connected nodes such as the structure it is used in are also activated, and are therefore more easily accessible for the next utterance, creating a priming effect.

Activation decays rapidly, so these models predict that priming will diminish with distance (Levelt and Kelter (1982) found a significant reduction in priming after just one intervening sentence, for example). A relationship between priming and distance was found in my data, but only under certain conditions – with the of-genitive, for example, or where there was no match between the heads of possessums. Over the dataset as a whole, the effect of distance was not significant, although distance was a significant main effect in both corpora when considered individually. Although no firm conclusion can be drawn about the rapidity of decay in my data, some priming effect does appear to still be operating even at great distances (up to around 2000 orthographic characters, which corresponded to 13 intervening sentences in one example). Whether this would remain the case if intervening possessive pronouns were taken into account is uncertain, however.

Alignment models (e.g. Pickering & Garrod, 2004) and prediction models (e.g. Jaeger & Snider, 2013; Pickering & Garrod, 2013) describe priming as a way of aligning speakers' linguistic representations in dialogue, building a shared model and way of describing the world. These models view priming as a phenomenon that arises from dialogue, with priming in monologue only happening incidentally, either because of self-monitoring (Pickering & Garrod, 2004) or because production and prediction make use of the same systems (Dell & Chang, 2014). Thus the presence of priming in monologue does not rule out these models, but it also does not provide any particular support for them.

Alignment models rely on alignment (and hence priming) occurring at all linguistic levels. Therefore, it would be expected that syntactic priming should be accompanied by lexical and semantic priming. Although my data does show some evidence for matching possessors increasing the strength of priming, there is no clear increase in priming from matching possessums, and it seems likely that Szmrecsanyi (2006) is correct in suggesting that the effect of POSSESSOR MATCH is due more to the possessor being discourse-old than to a lexical boost effect.

In implicit learning models (e.g. Bock & Griffin, 2000), priming is described as the result of long-term strengthening of connections between syntactic structures and the messages they can convey. As with alignment models, implicit learning models do not explicitly predict priming in monologue, but neither do they rule it out. A prediction arising from implicit learning models is that if one of two competing structures is less common than the other, then it will be primed more frequently, as it is more unexpected and therefore will be learnt more easily (Ferreira & Bock, 2006). My data does not support this prediction, as more priming was found for the s-genitive, which is the less common structure. However, as was discussed in Section 4.2 above, methodological issues may have skewed the results for s-genitives, so this is not sufficient evidence to entirely rule out implicit learning as a mechanism.

A number of proposed models combine short-term activation and long-term implicit learning. For example, Hartsuiker et al. (2008) argue that explicit memory is activated when there is a lexical boost from matching heads, and this activation decays rapidly, but in conditions where there is no lexical boost, only the implicit learning mechanism applies. Therefore it would be expected that when the heads match, priming would diminish much more rapidly with distance than when there is no match. I found no clear evidence for a lexical boost in my data, and moreover only found the expected decay of priming with distance in the condition where there was no match between possessum heads.

That no lexical boost effect was found does not entirely rule out these models, however. Lexical boost has been found most often in priming of verb phrase structures such as datives (e.g. Branigan et al., 2000; Hartsuiker et al., 2008). Only a limited number of verbs can be used in dative structures, while almost all nouns are able to appear in genitives. Therefore, it might be expected that there would be a stronger link between verb and dative structure than between noun and genitive. Additionally, Bernolet, Collina, and Hartsuiker (2016) suggest that explicit memory activation may not be restricted only to conditions of lexical boost, but may occur when there is any lexical or semantic similarity between the target and the prime. This assertion is unable to be tested in my data, which is only coded for head matching, but provides an interesting avenue for future research.

None of the models discussed here can be either fully supported or completely ruled out by the results of this study. What is evident though is that any model which purports to explain priming must be able to account not only for the mechanisms present in dialogue, but also for those of monologue.

5.2 Does priming differ between speech and writing?

This study found that overall, priming is more likely in speech than in writing. However, it is not clear how this difference arises. Differences in levels of formality and information density between the journalism of the *Press* and the personal stories of the QuakeBox archive may be sufficient to account for the difference in priming, or it could be that more fundamental differences in processing are involved.

It would be expected that any processing differences should be able to be detected via differences in the decay of priming over time. No significant difference was found for the effect of distance, but this result is not conclusive, as the measure of distance between prime and target used, textual distance, is unlikely to be entirely equivalent for speech and writing, as speech is generally produced more quickly than writing.

Other methodological problems, such as the exclusion of pronominal possessives, may also have contributed to the differences observed. In addition, the difference in size between the two corpora may account for at least some of the factors and interactions which were found to be significant in the *Press* data but which did not reach significance in QuakeBox. Equally though, this size difference may have masked genuine differences between the corpora. Therefore no firm conclusions can be drawn about whether priming differs between speech and writing.

An interesting possible difference, though, is that seen in the interaction of GENITIVE TYPE, PRIME TYPE, and MODE. For the s-genitive, differences between α-priming and β-priming appear to be largely attributable to methodological issues. However, for the of-genitive, a much more interesting pattern appears. In contrast to Szmrecsanyi (2006), who found less priming in the β condition (i.e. when the target is in a choice condition, but the prime is categorical), I found a significant increase in priming in the β condition, and this difference was more marked in writing than in speech. I hypothesise that this difference between α-priming and β-priming may be due to the way syntactic structures are activated in choice and categorical contexts. In a choice condition, activation is shared between the two available structures, so the activation of the selected structure is at a lower level than in a categorical condition, where the single available structure is fully activated. This lower level of initial activation would lead to choice genitives being less likely than categorical genitives to prime subsequent genitives.

If this explanation is correct, then it would be expected that α-priming would decay more rapidly than β-priming. I did not find the expected interaction between PRIME TYPE and DISTANCE in my results, so my hypothesis cannot be confirmed. However, if, as I supect, shortcomings in my methodology have weakened the overall impact of β-priming, then it may be that improvements to the methodology would allow such a difference to be detected. This possibility is discussed further in Section 5.3 below.

Although my aim was primarily to compare the strength of genitive priming in speech and writing, I was also able to show that there is a difference in genitive choice between the two modalities. The s-genitive is less likely to be used in speech than in writing, and the effect of animacy is much stronger in speech. These results agree with those of Grafmiller (2014), supporting his assertion that although his comparison of speech and writing used corpora collected several decades apart, the differences he found were not due to change over time.

5.3 Future research

Grafmiller (2014) suggests that the s-genitive is more common in newspaper writing than speech because the force of economisation in journalism prompts greater use of the more compact s-genitive with inanimate possessors. Although this study replicated Grafmiller's result, the written data also showed a small but significant increase in of-genitive use with animate possessors. I have suggested that this may be due my choice to code animacy as a binary animate/inanimate categorisation. It would be interesting to repeat this study using a finer-grained approach to animacy, such as that developed by Zaenen et al. (2004), in order to discover whether the frequency of lower-animacy possessors such as organisations can account for this difference in of-genitive use. Expanding the study to include other structures which express possessive relationships, such as pronominal possessives and nominal compounds, could also shed light on the differences in genitive use between speech and writing, as well as discovering whether there is any evidence for the thematic role priming proposed by Vasilyeva and Waterfall (2012).

Related to the question of animacy is the issue of dehumanisation via genitive choice. It appears possible that of-genitives are more likely to be used when a speaker is expressing criticism of an organisation or individual. Further research is needed to discover whether this is indeed a measurable effect. Can genitive choice reflect the attitude of the speaker to the possessor? And if so (and perhaps more importantly), is the speaker consciously aware of the dehumanising effect of using an of-genitive, and making the decision to use that structure deliberately, or is the choice happening at a deeper linguistic level?

This study used textual distance as a convenient measure of temporal distance. While textual distance proved to be closely correlated to temporal distance within the spoken corpus, the question remains as to its validity when comparing spoken and written corpora. It would be of value to future corpus studies comparing priming in speech and writing to develop a measure of distance that can more accurately represent the passage of time in both modalities.

Although this study used spoken and written corpora that were closely matched in geographical location, time period and topic, the *Press* articles and the QuakeBox stories still differed in other potentially important ways. To fully compare priming in speech and writing, it will be necessary to also attempt to match levels of formality and information density. Social media such as Facebook may provide a rich source of data for future research in this area. If a future significant event such as a natural disaster were to give the opportunity to again gather a QuakeBox-like spoken corpus focussed on a single topic, its value would be greatly enhanced by a companion written corpus of social media posts describing the same event.

While a number of studies have found strong evidence that priming is increased when heads of verb phrases match, the evidence for a lexical boost from matching noun phrase heads in the genitive alternation is inconclusive. If, as Szmrecsanyi (2006) suggests, what appears to be lexical boost in genitive priming is actually an effect of givenness, this may imply there is a fundamental difference between nouns and verbs in the way that lexical items are connected to syntactic structures within the mental representation of language. Further research is needed to tease apart the effect of information status from that of lexical boost, and to investigate the possibility raised by Bernolet et al. (2016) that lexical boost could be coming from other lexical or semantic matching, not just matching heads.

This study did not adequately address the issue of nested genitives. It is apparent that treating nested genitives in the same way as non-nested genitives muddies the relationship between priming and distance. Future research taking nesting into account would provide a more accurate picture of the effect of distance on priming. Additionally, if priming in nested structures is found to be hierarchical rather than linear, as Pickering et al. (2000) suggest, this would provide important evidence for the hierarchical structure of syntactic representations.

Finally, β-priming offers a promising area for future research. If α-priming involves a lower level of activation than β-priming, then restricting potential β-primes to only those categorical genitives which share a syntactic structure with choice-condition s-genitives or of-genitives should result in both s-genitives and of-genitives showing a greater degree of priming in the β condition. It would also be interesting to more accurately model activation decay for α- and β-primes, by taking into account any thematic role priming from intervening pronominal possessives and nominal compounds. If β-primes have a higher initial level of activation than α-primes, then α-primes should decay more rapidly. Investigating this difference between α- and β-priming will enable greater insights into the mechanism of activation, and the role that activation plays in syntactic priming.

6 Conclusion

This study investigated within-speaker priming in extended monologue. Using the QuakeBox and *Press* corpora as my data sources allowed me to directly compare priming in spoken and written monologues that are closely matched in geographical location, time period and topic.

My study found clear evidence for syntactic priming in monologue, with priming occurring in both the spoken and written corpora, and priming more likely in speech than in writing. However, the evidence is less clear for other factors that might be expected to influence syntactic priming, such as distance, head matching, and α- versus β-priming.

Although this study was primarily concerned with syntactic priming, I was also able to confirm the influence of a number of other factors previously reported to affect genitive choice (cf. Altenberg, 1982; Grafmiller, 2014; Quirk et al., 1985; Rosenbach, 2005). Animacy has a particularly strong effect in my data, with animate possessors greatly favouring the s-genitive, and the effect of animacy was found to be greater in speech than in writing. Prototypicality and sibilant sounds were also found to affect genitive choice, with sibilant-final possessors favouring the of-genitive, and the s-genitive being more likely to be used to express a prototypical possessive relationship.

As discussed in Section 5.3 above, the findings of this study raise many important questions that deserve to be investigated further. The areas of β-priming and priming of nested structures offer particularly promising directions for future study, and to gain a comprehensive understanding of genitive priming we may need to expand the range of structures included in the analysis to pronominal possessives as well as other structures which express a possessive relationship.

References

Adger, D. (2003). *Core syntax: a minimalist approach*. New York: Oxford University Press.

Altenberg, B. (1982). *The genitive v. the of-construction: a study of syntactic variation in 17th century English*. Lund: LiberLäromedel/Gleerup.

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. doi: http://dx.doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious Mixed Models. Retrieved 19 December 2016, 2016, from https://arxiv.org/abs/1506.04967v1

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi: 10.18637/jss.v067.i01

Bernolet, S., Collina, S., & Hartsuiker, R. J. (2016). The persistence of syntactic priming revisited. *Journal of Memory and Language, 91*, 99-116. doi: 10.1016/j.jml.2016.01.002

Biber, D. (2003). Compressed noun-phrase structures in newspaper discourse: The competing demands of popularization vs. economy. In J. Aitchison & D. M. Lewis (Eds.), *New Media Language* (pp. 169-181). London: Routledge.

Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*(3), 355-387.

Bock, K., & Griffin, Z. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General, 129*(2), 177-192. doi: 10.1037//0096-3445.129.2.177

Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition, 35*(1), 1-39.

Branigan, H. P. (2007). Syntactic Priming. *Language and Linguistics Compass, 1*(1-2), 1-16.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (1999). Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin & Review, 6*(4), 635-640. doi: 10.3758/BF03212972

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*(2), B13-B25.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, H. (2007). Predicting the Dative Alternation. In G. Bouma, I. Krämer & J. Zwarts (Eds.), *Cognitive Foundations of Interpretation* (pp. 69-94). Amsterdam: Royal Netherlands Academy of Science.

Bresnan, J., & Hay, J. (2008). Gradient grammar: An effect of animacy on the syntax of give in New Zealand and American English. *Lingua, 118*(2), 245-259.

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

Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language, 49*(2), 214-230.

Dahl, L. (1971). The S-Genitive with Non-Personal Nouns in Modern Journalistic Style. *Neuphilologische Mitteilungen: Bulletin de la Societe Neophilologique/Bulletin of the Modern Language Society, 72*, 140-172.

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), 20120394-20120394. doi: 10.1098/rstb.2012.0394

Dziak, J. J., Coffman, D. L., Lanza, S. T., & Li, R. (2012). Sesnsitivity and specificity of information criteria. *Technical Report Series* 12-119. from https://methodology.psu.edu/media/techreports/12-119.pdf

Ferreira, V. S., & Bock, K. (2006). The functions of structural priming. *Language and Cognitive Processes, 21*(7-8), 1011-1029.

Frank, A. (2011). Diagnosing collinearity in mixed models from lme4. Retrieved from https://hlplab.wordpress.com/2011/02/24/diagnosing-collinearity-in-lme4/

Fromont, R., & Hay, J. (2008). ONZE Miner: the development of a browser-based research tool. *Corpora, 3*(2), 173-193. doi: doi:10.3366/E1749503208000142

Fromont, R., & Hay, J. (2012). LaBB-CAT: an Annotation Store. *Proceedings of Australasian Language Technology Workshop*, 113-117.

Grafmiller, J. (2014). Variation in English genitives across modality and genres. *English Language & Linguistics, 18*(3), 471-496.

Gries, S. T. (2005). Syntactic Priming: A Corpus-based Approach. *Journal of Psycholinguistic Research, 34*(4), 365-399.

Gries, S. T., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics, 9*(1), 97-129.

Harrell, F. E. J. (2016). Regression Modeling Strategies (Version 4.5-0). Retrieved from https://CRAN.R-project.org/package=rms

Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., & Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language, 58*(2), 214-238. doi: 10.1016/j.jml.2007.07.003

Hartsuiker, R. J., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition, 75*(2), B27-B39. doi: 10.1016/S0010-0277(99)00080-3

Hinrichs, L., & Szmrecsanyi, B. (2007). Recent changes in the function and frequency of Standard English genitive constructions: a multivariate analysis of tagged corpora. *English Language & Linguistics, 11*(3), 437-474. doi: 10.1017/S1360674307002341

Hundt, M., & Szmrecsanyi, B. (2012). Animacy in early New Zealand English. *English World-Wide, 33*(3), 241-263. doi: 10.1075/eww.33.3.01hun

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition, 127*(1), 57-83.

Jankowski, B. L. (2013). *A Variationist Approach to Cross-register Language Variation and Change.* (PhD PhD dissertation), University of Toronto.

Jankowski, B. L., & Tagliamonte, S. A. (2014). On the genitive's trail: data and method from a sociolinguistic perspective. *English Language & Linguistics, 18*(2), 305-329. doi: 10.1017/S1360674314000045

Jucker, A. H. (1993). The Genitive versus the Of-Construction in Newspaper Language. *Anglistik & Englischunterricht, 49*, 121-136.

Kootstra, G. J., van Hell, J. G., & Dijkstra, T. (2010). Syntactic alignment and shared word order in code-switched sentence production: Evidence from bilingual monologue and dialogue. *Journal of Memory and Language, 63*(2), 210-231. doi: 10.1016/j.jml.2010.03.006

Leech, G., & Smith, N. (2006). Recent Grammatical Change in Written English 1961-1992: Some Preliminary Findings of a Comparison of American with British English. In A.

Kehoe & A. Renouf (Eds.), *Changing Face of Corpus Linguistics* (pp. 185-204). Amsterdam; New York: Rodopi.

Lehrer, A. (1986). English classifier constructions. *Lingua, 68*(2), 109-148. doi: 10.1016/0024-3841(86)90001-X

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology, 14*(1), 78-106. doi: 10.1016/0010-0285(82)90005-6

Lowe, J. J. (2016). English possessive 's: clitic and affix. *Natural Language & Linguistic Theory, 34*(1), 157-195. doi: 10.1007/s11049-015-9300-1

O'Connor, C., Maling, J., & Skarabela, B. (2013). Nominal categories and the expression of possession. In K. Börjars, D. Denison & A. Scott (Eds.), *Morphosyntactic categories and the expression of possession* (pp. 89-121). Amsterdam and Philadelphia: John Benjamins.

Pickering, M. J., & Branigan, H. P. (1998). The Representation of Verbs: Evidence from Syntactic Priming in Language Production. *Journal of Memory and Language, 39*(4), 633-651. doi: 10.1006/jmla.1998.2592

Pickering, M. J., Branigan, H. P., Cleland, A. A., & Stewart, A. J. (2000). Activation of Syntactic Information During Language Production. *Journal of Psycholinguistic Research, 29*(2), 205-216. doi: 10.1023/A:1005149129259

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169-190. doi: 10.1017/S0140525X04000056

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

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive grammar of the English language*. New York; London: Longman.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rosenbach, A. (2002). *Genitive Variation in English: Conceptual Factors in Synchronic and Diachronic Studies*. Germany: Mouton de Gruyter.

Rosenbach, A. (2003). Aspects of Iconicity and Economy in the Choice between the S-Genitive and the Of-Genitive in English. In G. n. Rohdenburg & B. Mondorf (Eds.), *Determinants of Grammatical Variation in English* (pp. 379-412). Berlin, New York: De Gruyer Mouton.

Rosenbach, A. (2005). Animacy versus Weight as Determinants of Grammatical Variation in English. *Language, 81*(3), 613-644.

Rosenbach, A. (2014). English genitive variation - the state of the art. *English Language & Linguistics, 18*(2), 215-262.

Rutkowski, P. (2007). The syntactic structure of grammaticalized partitives (pseudo-partitives). *University of Pennsylvania Working Papers in Linguistics, 13*(1), 337-350.

Segaert, K., Menenti, L., Weber, K., & Hagoort, P. (2011). A Paradox of Syntactic Priming: Why Response Tendencies Show Priming for Passives, and Response Latencies Show Priming for Actives. *PLoS One, 6*(10), e24209. doi: 10.1371/journal.pone.0024209

Shih, S., Grafmiller, J., Futrell, R., & Bresnan, J. (2015). Rhythm's role in genitive construction choice in spoken English. In R. Vogel (Ed.), *Rhythm in Cognition and Grammar: A Germanic Perspective* (pp. 207-234). Berlin, München, Boston: De Gruyter Mouton.

Smithies, J., Millar, P., & Thomson, C. (2015). Open Principles, Open Data: The Design Principles and Architecture of the UC CEISMIC Canterbury Earthquakes Digital

Archive. *Journal of the Japanese Association for Digital Humanities, 1*(1), 10-36. doi: 10.17928/jjadh.1.1_10

Stefanowitsch, A. (2003). Constructional semantics as a limit to grammatical alternation: The two genitives of English. In G. n. Rohdenburg & B. Mondorf (Eds.), *Determinants of Grammatical Variation in English* (pp. 413-441). Berlin/New York: Mouton de Gruyter.

Stickney, H. (2009). *The Emergence of DP in the Partitive Structure.* (Doctor of Philosophy (PhD) PhD dissertation), University of Massachusetts - Amherst. Retrieved from http://scholarworks.umass.edu/open_access_dissertations/131

Szmrecsanyi, B. (2006). *Morphosyntactic persistence in spoken English: a corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis* Retrieved from http://www.ebrary.com

Szmrecsanyi, B. (2010). The English genitive alternation in a cognitive sociolinguistics perspective. In D. Geeraerts, G. Kristiansen & Y. Peirsman (Eds.), *Advances in Cognitive Sociolinguistics* (pp. 139-166). Berlin, New York: De Gruyter Mouton.

Szmrecsanyi, B. (2013). The Great Regression: Genitive variability in Late Modern English news texts. In D. Denison, K. Börjars & A. Scott (Eds.), *Morphosyntactic categories and the expression of possession* (pp. 59-88). Amsterdam; Philadelphia: John Benjamins Publishing Company.

Szmrecsanyi, B., & Hinrichs, L. (2008). Probabilistic determinants of genitive variation in spoken and written English: A multivariate comparison across time, space and genres. In T. Nevalainen (Ed.), *The dynamics of linguistic variation: corpus evidence on English past and present* (pp. 291-309). Amsterdam; Philadelphia: John Benjamins.

Tannen, D. (1987). Repetition in Conversation: Toward a Poetics of Talk. *Language, 63*(3), 574-605.

Taylor, J. R. (1996). *Possessives in English: an exploration in cognitive grammar*. Oxford: Clarendon Press.

van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language, 55*(3), 335-362. doi: 10.1016/j.jml.2006.06.004

Vasilyeva, M., & Waterfall, H. (2012). Beyond syntactic priming: Evidence for activation of alternative syntactic structures. *Journal of Child Language, 39*(2), 258-283. doi: 10.1017/S0305000911000055

Walsh, L., Hay, J., Bent, D., Grant, L., King, J., Millar, P., Papp, V., & Watson, K. (2013). The UC QuakeBox Project: Creation of a community-focused research archive. *New Zealand English Journal, 27*, 20-32.

Wolk, C., Bresnan, J., Rosenbach, A., & Szmrecsanyi, B. (2013). Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. *Diachronica, 30*(3), 382-419.

Yamamoto, M. (1999). *Animacy and Reference*. Philadelphia, Netherlands: John Benjamins Publishing Company.

Zaenen, A., Carletta, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Nikitina, T., O'Connor, M. C., & Wasow, T. (2004). Animacy encoding in English: why and how. In D. Byron & B. Webber (Eds.), *Proceedings of the 2004 ACL Workshop on Discourse Annotation, Barcelona, July 2004* (pp. 118-125). Stroudsburg PA: Association for Computational Linguistics.

Appendices

A Microsoft Excel Visual Basic script

Textual distance measures were not available at the time of the initial data extraction, but were added later by repeating the original searches on the transcript layer. This script was used to align tokens from the new set of search results to those in the working dataset by matching speaker id and timestamp. The textual distances from the new search results were then appended to the working data.

```
Sub AddTextualDistance()
' Searches through new search results for tokens already identified as being of interest and transfers the textual distance value for each token to the cleaned up data.

' Before running:
' 1) back up working data file
' 2) copy new search results csv data into a new sheet in the working file
' 3) write 'stop' at the bottom of the Speaker column on both sheets
' 4) change sheet and range values where indicated by ********* below

Dim Working As Worksheet 'worksheet that holds current working copy of data
Dim Search As Worksheet 'worksheet that holds the newly downloaded search results
Dim SpeakerColWork As Range 'first cell in the column which contains speaker names in working sheet
Dim SpeakerColSearch As Range 'first cell in the column which contains speaker names in results sheet
Dim TimeColWork As Range 'first cell in the column which contains end timestamps in working sheet
Dim TimeColSearch As Range 'first cell in the column which contains end timestamps in results sheet
Dim DistColWork As Range 'first cell in the column which contains character count in working sheet
Dim DistColSearch As Range 'first cell in the column which contains character count in results sheet

'variables to hold values from the above cells:
Dim SpeakerWork
Dim SpeakerSearch
Dim TimeWork
Dim TimeSearch
Dim DistSearch

Dim WorkRow As Integer 'the row I'm up to in the working sheet
Dim SearchRow As Integer 'the row I'm up to in the results sheet

Dim FoundIt As Boolean

'************************************
'    CHANGE VALUES HERE
'************************************
Set Working = Sheets("Sheet1")
Set Search = Sheets("Sheet2")

'Initialise starting cell in each column of interest
Set SpeakerColWork = Working.Range("D1") 'Speaker
Set SpeakerColSearch = Search.Range("D1") 'Speaker
Set TimeColWork = Working.Range("AL1") 'Target transcript end
Set TimeColSearch = Search.Range("R1") 'Target transcript end
Set DistColWork = Working.Range("AO1") 'where I want the character count end to go
Set DistColSearch = Search.Range("M1") 'where the character count is
'************************************
```

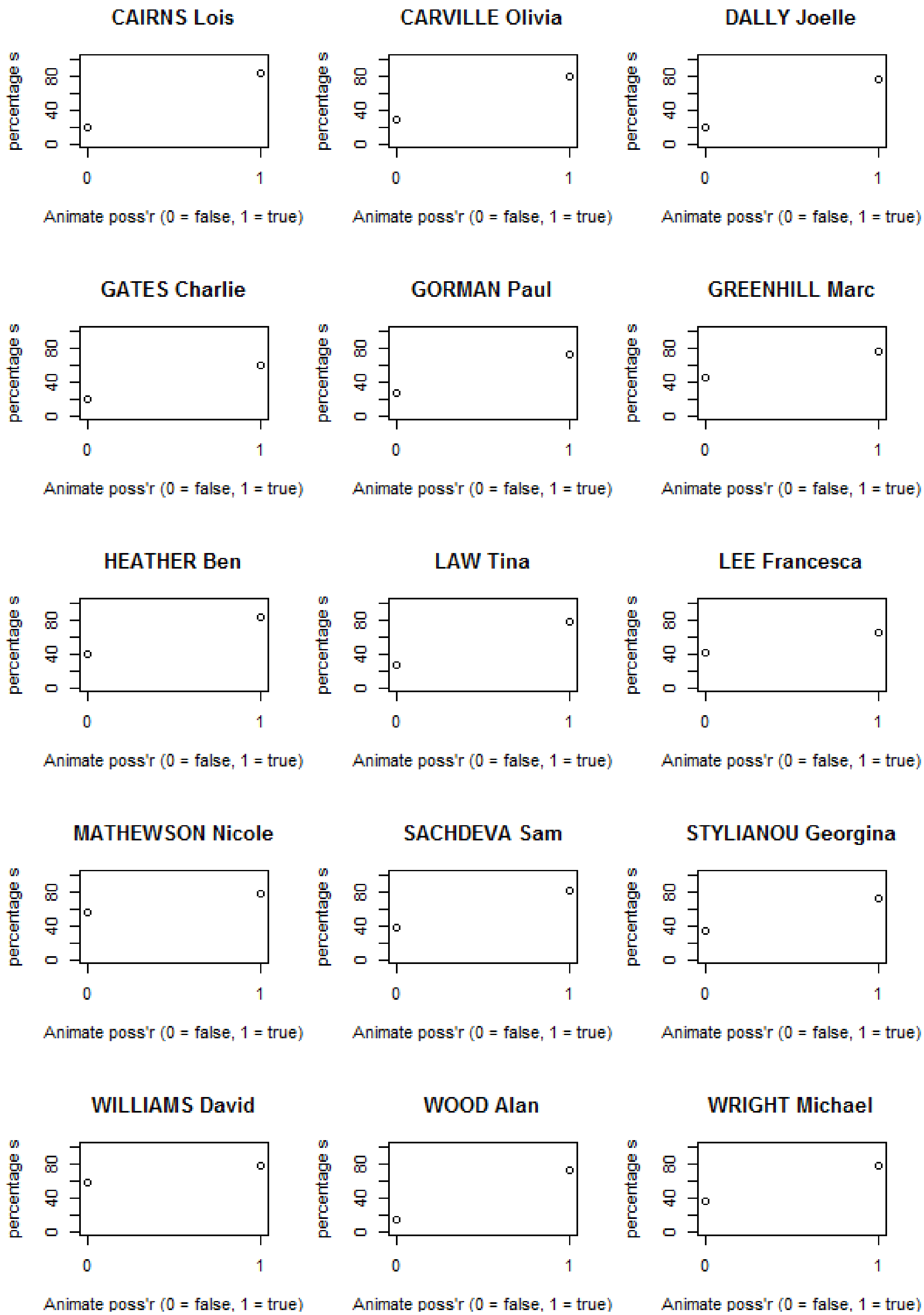
```
WorkRow = 1 'actually is the row number - 1, because row 1 is the headers
Do 'work down main sheet row by row
  SpeakerWork = SpeakerColWork.Offset(WorkRow, 0).Value
  TimeWork = TimeColWork.Offset(WorkRow, 0).Value

  'reset values for looping through the results page
  SearchRow = 1
  FoundIt = False

  Do 'work down results sheet looking for a match
    SpeakerSearch = SpeakerColSearch.Offset(SearchRow, 0).Value
    TimeSearch = TimeColSearch.Offset(SearchRow, 0).Value
    If SpeakerSearch = SpeakerWork And TimeSearch = TimeWork Then
            'copy value if it's found the matching row
      FoundIt = True 'so loop can be stopped prematurely
      DistSearch = DistColSearch.Offset(SearchRow, 0).Value
      DistColWork.Offset(WorkRow, 0).Value = DistSearch
    End If
    SearchRow = SearchRow + 1
  Loop While SpeakerSearch <> "stop" 'stop when it hits a blank row
  If Not FoundIt Then 'message so I know it actually didn't find it
       DistColWork.Offset(WorkRow, 0).Value = "Not found"
  End If
  WorkRow = WorkRow + 1
Loop While SpeakerWork <> "stop"
'main loop through working sheet ends if it hits a blank row

End Sub
```

B By-speaker random effects

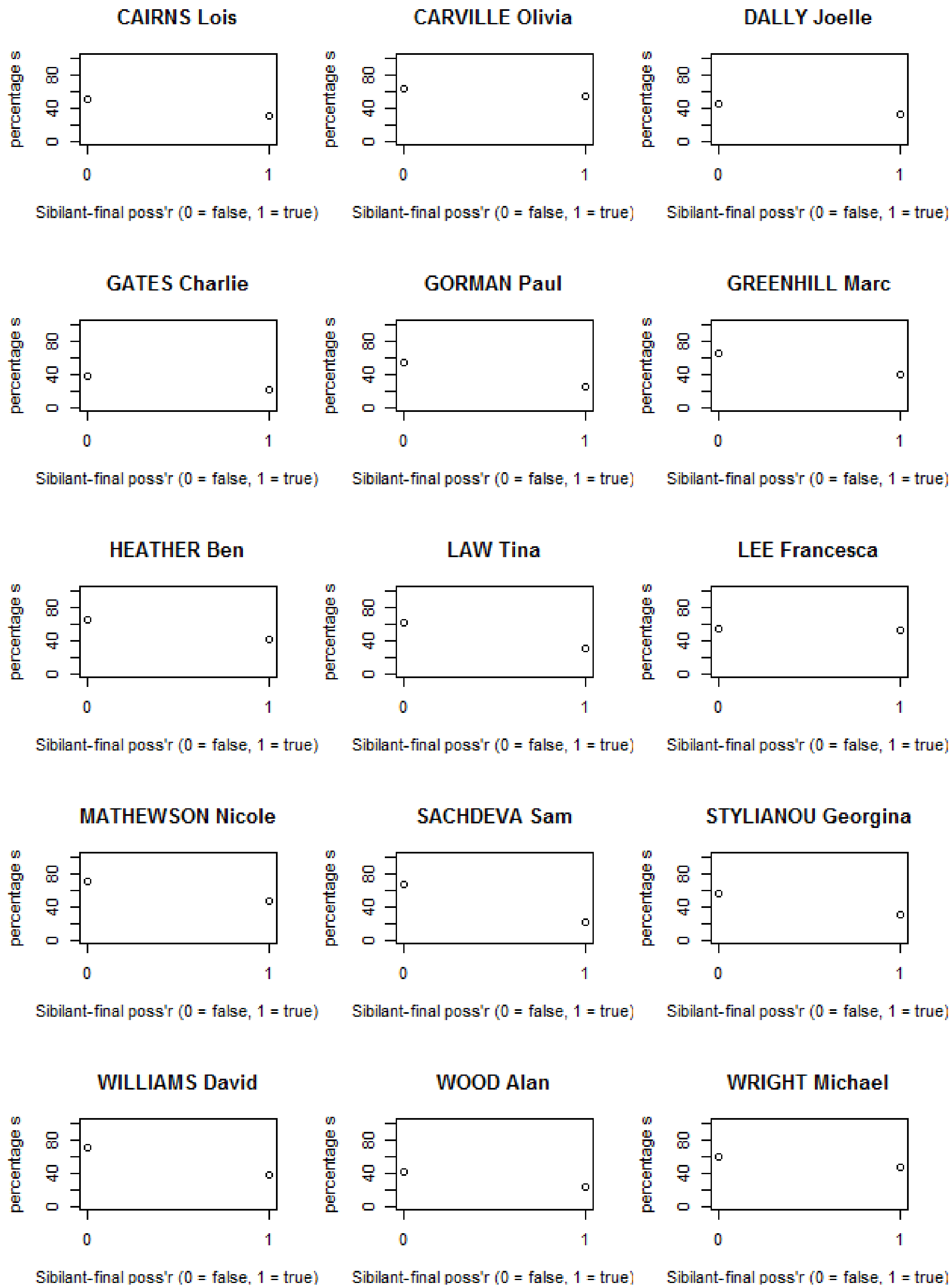
These graphs of the raw frequencies of factors across the 15 speakers with the largest number of tokens suggest that by-speaker random slopes may be present for some factors. However, because so many of the speakers provided only a small number of tokens each, there was insufficient data to allow models with random slopes to converge, so only the random intercept was included in the final models.

Note that in the graphs of binary factors (e.g. ANIMATE, SIBILANT), the x-axis represents false as 0, and true as 1.

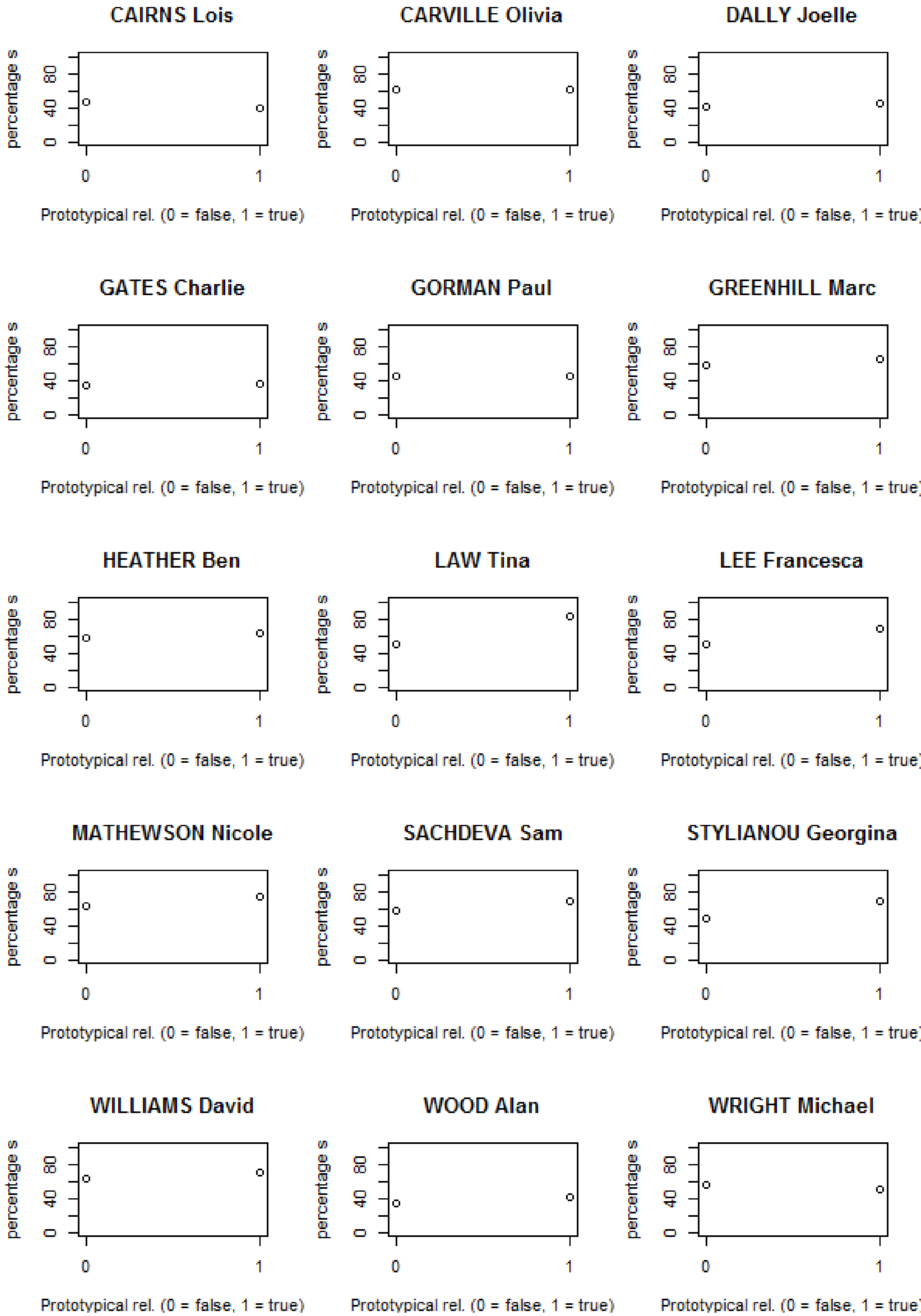
ANIMATE


CAIRNS Lois
CARVILLE Olivia
DALLY Joelle
GATES Charlie
GORMAN Paul
GREENHILL Marc
HEATHER Ben
LAW Tina
LEE Francesca
MATHEWSON Nicole
SACHDEVA Sam
STYLIANOU Georgina
WILLIAMS David
WOOD Alan
WRIGHT Michael
percentage s
0
40
80
0
1
Animate poss'r (0 = false, 1 = true)

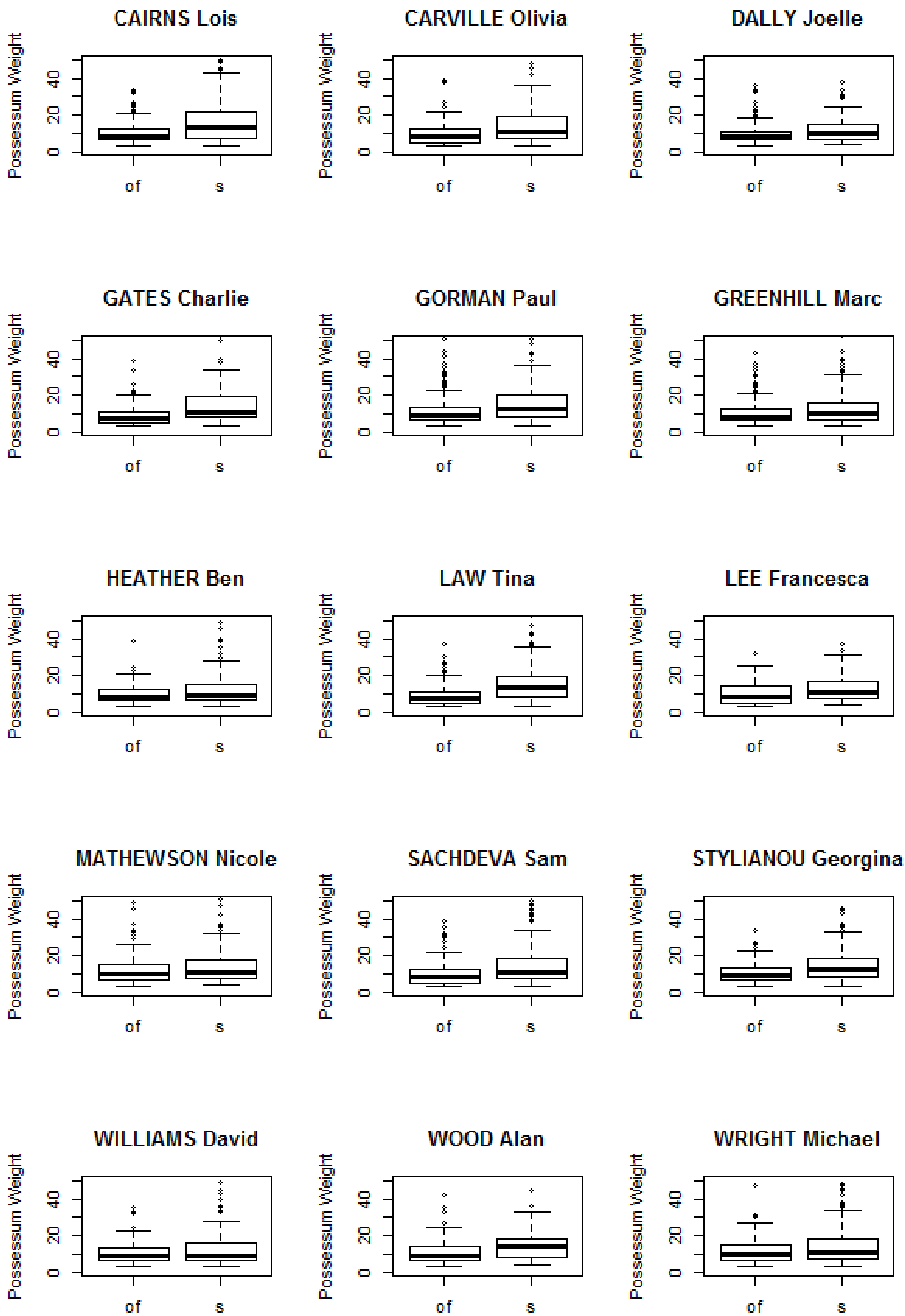
SIBILANT


CAIRNS Lois
CARVILLE Olivia
DALLY Joelle
GATES Charlie
GORMAN Paul
GREENHILL Marc
HEATHER Ben
LAW Tina
LEE Francesca
MATHEWSON Nicole
SACHDEVA Sam
STYLIANOU Georgina
WILLIAMS David
WOOD Alan
WRIGHT Michael
percentage s
0 40 80
0 1
Sibilant-final poss'r (0 = false, 1 = true)

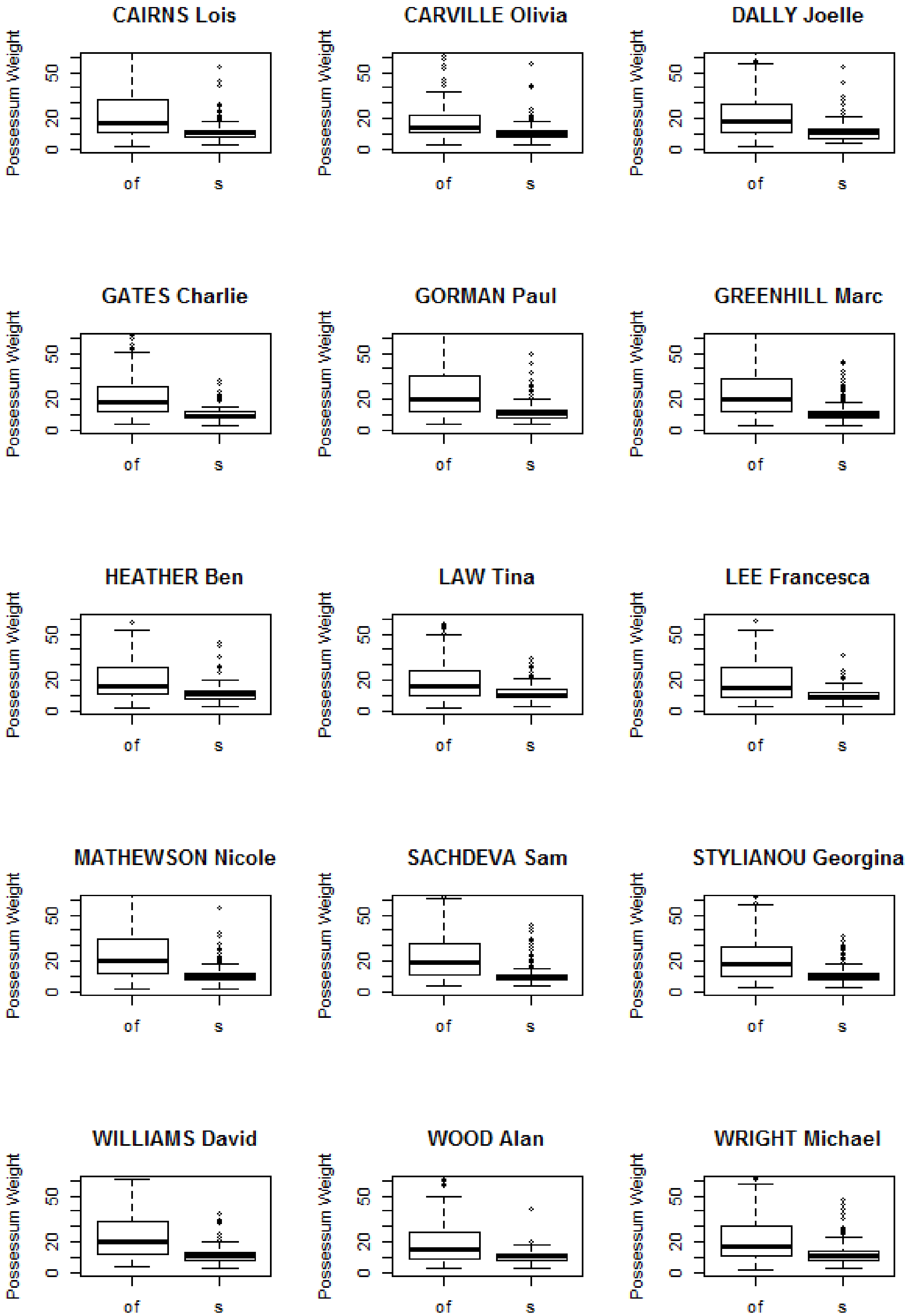
PROTOTYPICAL



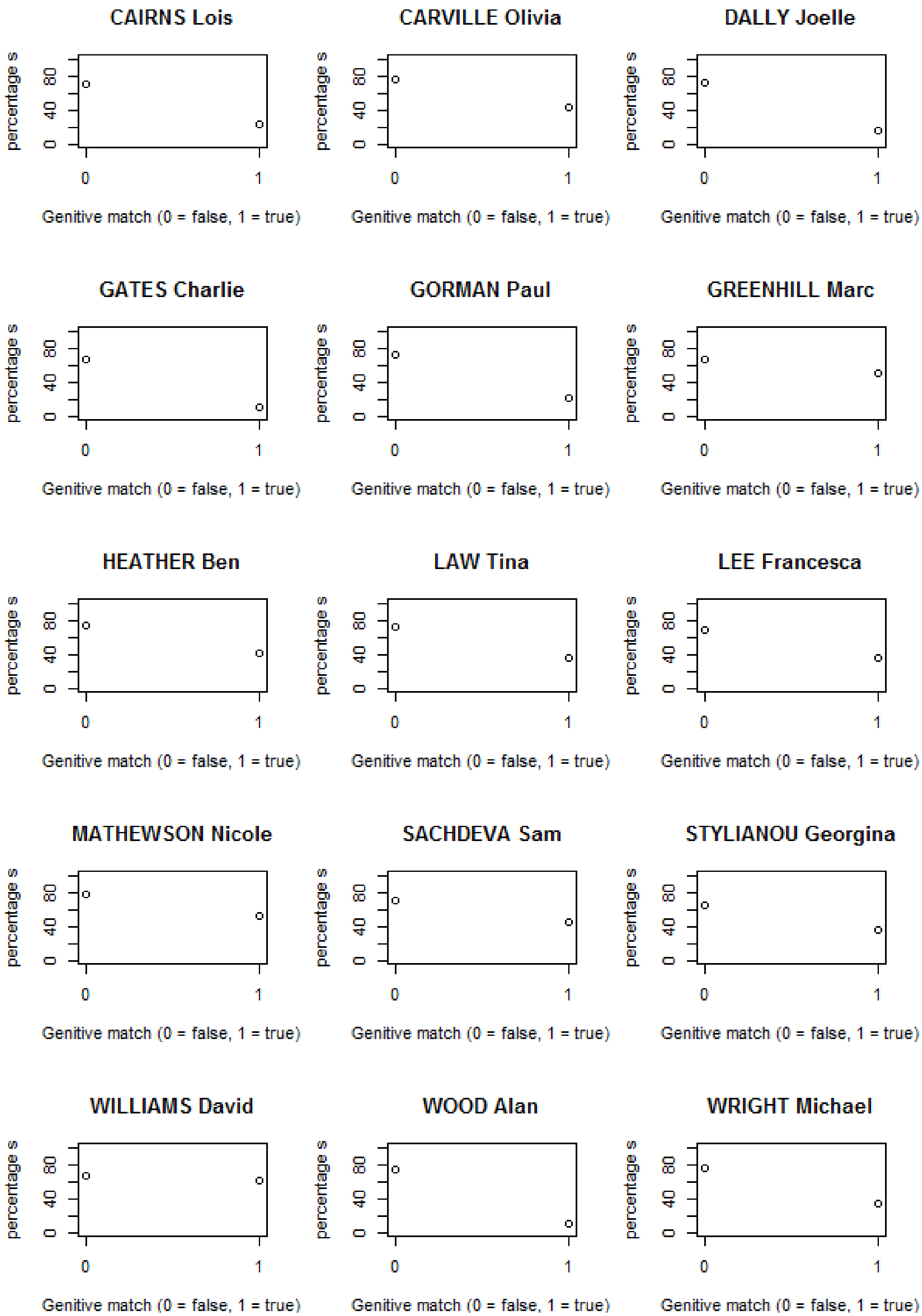
POSSESSUM WEIGHT



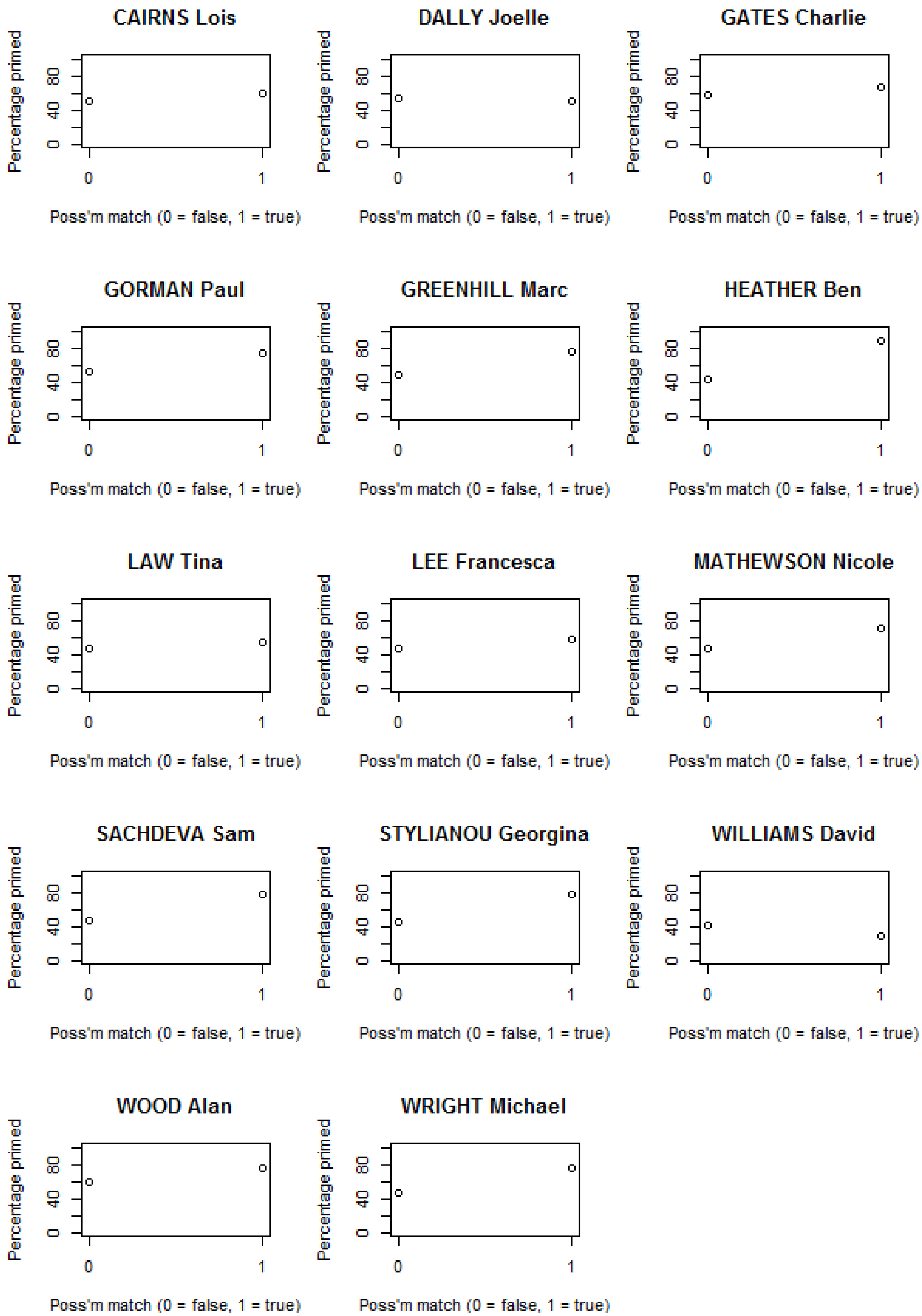
POSSESSOR WEIGHT



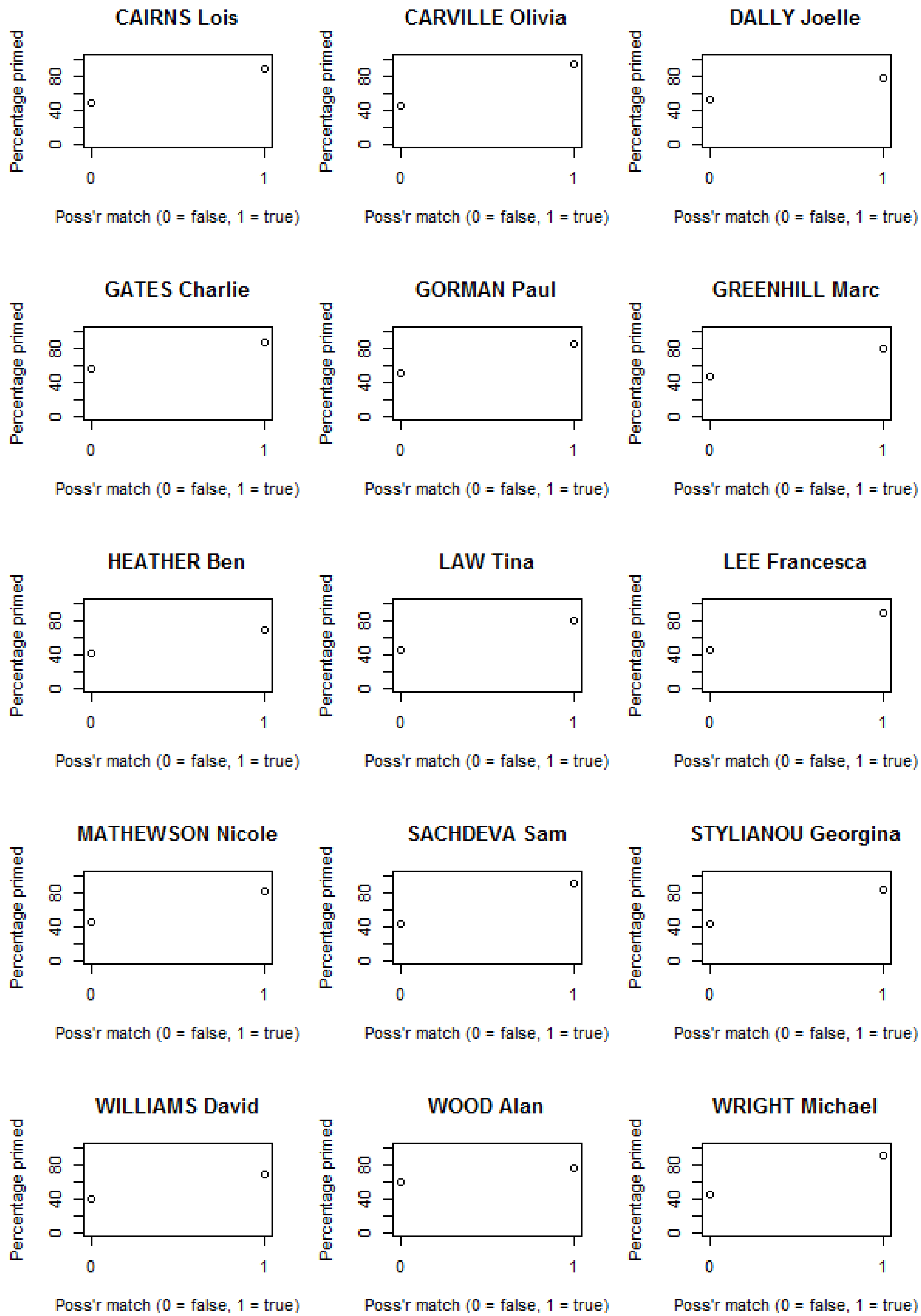
GENITIVE MATCH



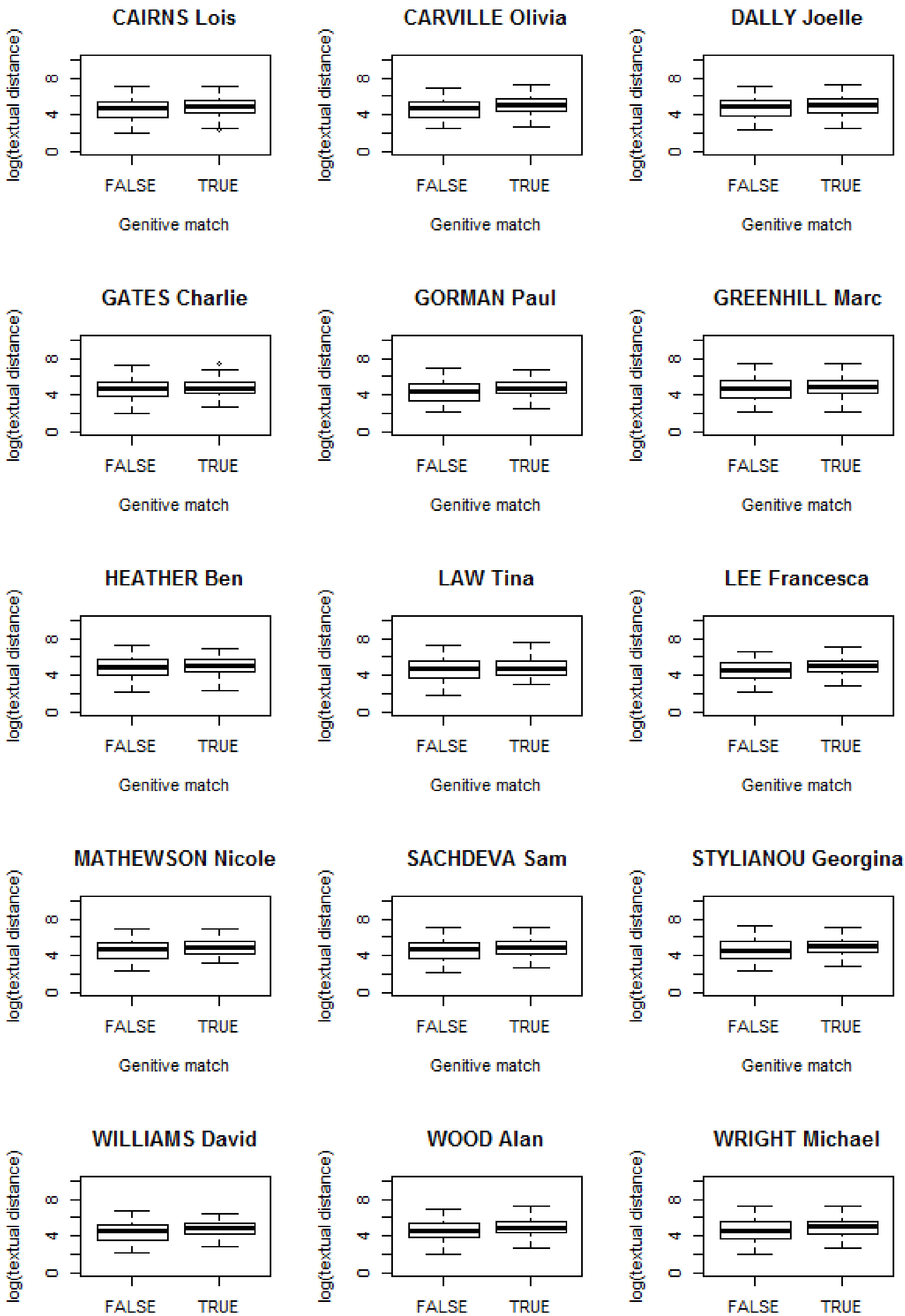
POSSESSUM MATCH



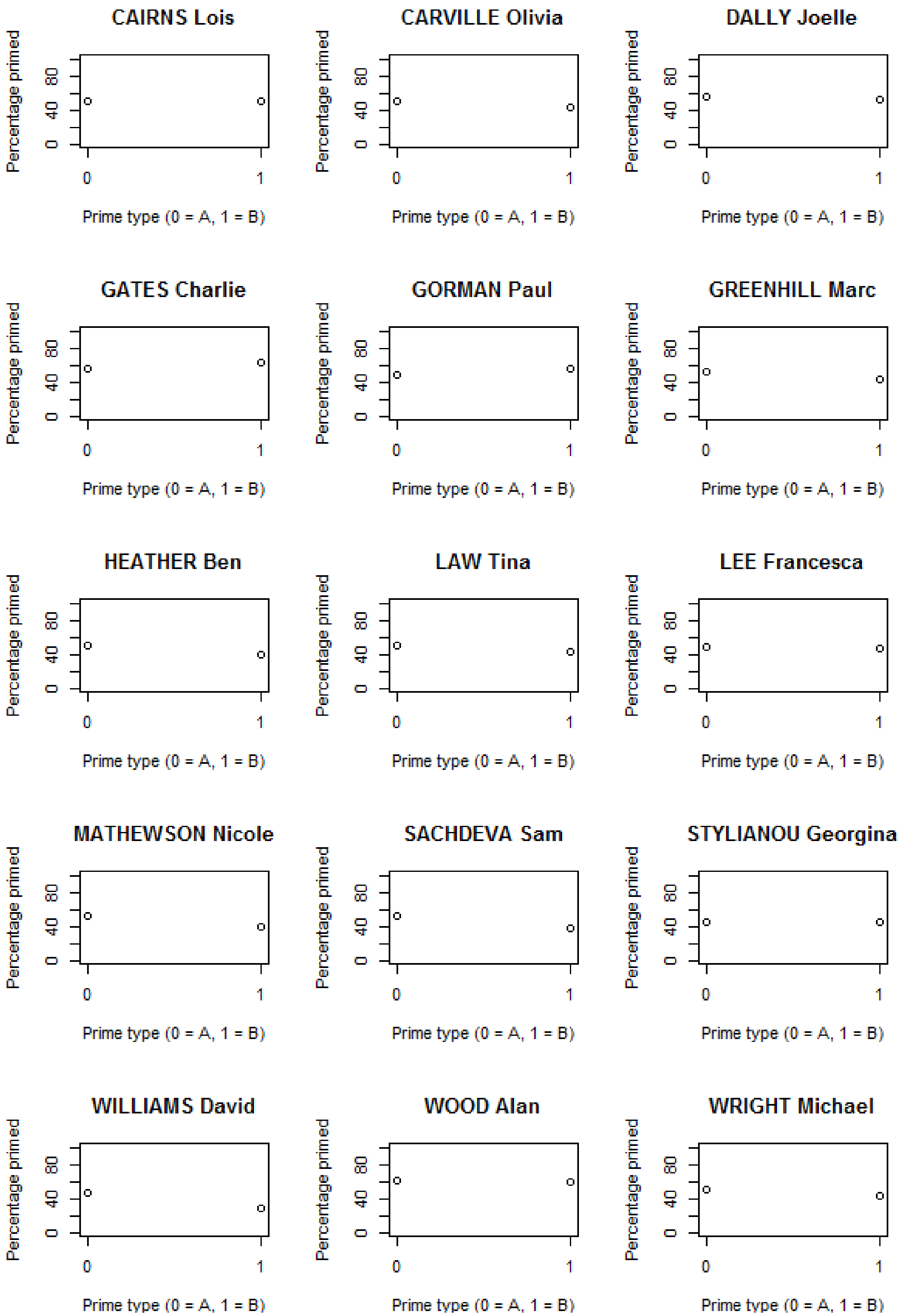
POSSESSOR MATCH



log(DISTANCE)



PRIME TYPE


CAIRNS Lois
CARVILLE Olivia
DALLY Joelle
GATES Charlie
GORMAN Paul
GREENHILL Marc
HEATHER Ben
LAW Tina
LEE Francesca
MATHEWSON Nicole
SACHDEVA Sam
STYLIANOU Georgina
WILLIAMS David
WOOD Alan
WRIGHT Michael
Percentage primed
0
40
80
0
1
Prime type (0 = A, 1 = B)