

Animacy in early New Zealand English*

Marianne Hundt and Benedikt Szmrecsanyi

University of Zurich / University of Freiburg

The literature suggests that animacy effects in present-day spoken New Zealand English (NZE) differ from animacy effects in other varieties of English. We seek to determine if such differences have a history in earlier NZE writing or not. We revisit two grammatical phenomena — progressives and genitives — that are well known to be sensitive to animacy effects, and we study these phenomena in corpora sampling 19th- and early 20th-century written NZE; for reference purposes, we also study parallel samples of 19th- and early 20th-century British English and American English. We indeed find significant regional differences between early New Zealand writing and the other varieties in terms of the effect that animacy has on the frequency and probabilities of grammatical phenomena.

Keywords: animacy, progressive, genitive alternation, regional variation, British, American and New Zealand English

1. Introduction

Animacy matters. Cross-linguistically, animacy effects in grammar are common (for example, Dahl and Fraurud 1996). In English, gradient degrees of NP animacy constrain the choice of various grammatical constructions, such as genitives (see e.g. Hinrichs and Szmrecsanyi 2007), datives (e.g. Bresnan and Hay 2008), or relativizers (e.g. Jaeger 2006). Animacy also plays a role in the later stages of the grammaticalisation of constructions such as the progressive in general and the progressive passive in particular: progressives with inanimate subjects increased before the construction's text frequency soared in the 20th century (Hundt 2004a),

* We would like to thank the participants of the workshop on areal and regional variation in English historical corpora at the *Helsinki Corpus Festival* (28 September — 2 October 2011). We are also grateful to Edgar W. Schneider and the anonymous reviewer of *EWV* for their helpful comments and constructive criticism on previous versions of this paper, which allowed us to further develop our ideas and fine-tune the interpretation of the data. This material is based upon work supported by the National Science Foundation under Grant No. BCS-1025602.

whereas the progressive passive shows a predilection to co-occur with animate subjects (Hundt 2004b: 88–9).

Studies based on corpora of Present-Day English (PDE) indicate that there is considerable regional variation in the effect that animacy has on the choice of constructional variants or the use of constructions such as the progressive (e.g. Hundt 2009a: 299). For speakers of contemporary New Zealand English (NZE), Bresnan and Hay (2008: 246) find that they are more sensitive to animacy of constituent NPs when choosing dative constructions (as in *we sent him a note* vs. *we sent a note to him*) than are speakers of contemporary American English (AmE). Hundt (1998: 45–9) demonstrates that the *s*-genitive is used significantly less frequently in NZE with personal nouns (as in [*Mr Smith*]_{personal noun}'s [*mansion*], as opposed to [*the mansion*] of [*Mr Smith*]_{personal noun}), a category high on the animacy scale, than in British English (BrE), AmE, or Australian English (AusE). Finally, Hundt (2009a: 299) observes that New Zealanders use the progressive and its passive variant significantly more often with inanimate subjects (as in [*the theory*]_{inanimate subject} *is being dealt a fatal blow*) than it is attested in AusE or BrE.

In light of these curious regional differences, the present paper features a pilot study exploring a small-scale historical corpus of early New Zealand writing, covering the period between the 1840s and the 1960s. Our research question is the following: are divergent animacy effects in NZE a recent development, or can they be found early on in the development of this variety? We investigate animacy in two very different constructions or constructional alternations (see Sec. 3): progressive VPs, and *of*- versus *s*-genitive NPs. We make use of comparative data for BrE and AmE from ARCHER (A Representative Corpus of Historical English Registers) and previous research (Hundt 2004a; Szmrecsanyi *fc.*).

The present study is an exercise in corpus-based historical linguistics. We wish to emphasize, however, that we are interested in shifting grammatical preferences (to what degree has animacy come to favour or disfavour particular grammatical constructions in NZE?). In other words, we are not exploring categorical grammatical change. Categorical grammatical change, such as the emergence of obligatory DO-support in Early Modern English (Ellegård 1953), can take centuries to reach completion (except under extreme circumstances, such as creole genesis). But from research investigating grammatical developments in the Brown family of corpora (e.g. Hundt and Mair 1999; Hinrichs and Szmrecsanyi 2007; Leech *et al.* 2009) we know that non-categorical (i.e. probabilistic) grammatical preferences can change significantly in a period of as little as 30 years. From this we conclude that it is a worthwhile task to explore changing grammatical preferences in a historical corpus that covers roughly a century.

Another issue that we should address right at the outset is why we should find regional differences at all concerning the role that animacy plays as a determinant

of grammatical variation. We submit that, unlike constraints such as the principle of end weight (in a VO language like English, long constituents are placed after short constituents), which are probably rooted in the human speech production architecture and thus fairly invariant across communities, animacy is a factor that ultimately boils down to human culture and ecological cognition. Consider gender systems in the world's languages: while these are all based on some notion of animacy, the exact shape and demarcation of animacy categories is highly variable. Some languages adopt sex-based systems, and other languages do not, due to different world views (see Corbett 2005). Add to this the effect of dialect contact that is characteristic of early NZE (see Sec. 2.1), and it becomes even less likely that Britons and New Zealanders — two communities that share roughly the same language but inhabit different hemispheres — interpret animacy categories (including their impact on grammatical variation) in exactly the same way. Instead, more or less subtle differences in, for instance, the extent to which inanimate subjects discourage the use of the progressive construction or inanimate possessors discourage the use of the *s*-genitive are likely to emerge. We also know that animacy is a determinant of grammatical variation that changes across time, so that variation across text type and regional variety (i.e. change at different speeds in different genres and varieties) might simply be a concomitant of diachronic change. In short, we seek to explore in this study the evolution of a culturally malleable determinant of grammatical variation, animacy, during the genesis of a new variety of English, NZE.

In part two of our paper, we will give a brief summary of the development of NZE as a background to our discussion of the corpus design. We also describe the corpora that this study is based on. In part three, we will define the linguistic variables and comment in some detail on the way we coded for the factor 'animacy'. In part four, we will present and discuss our results, including aspects of regional variation and the effect that text type may have. In our conclusion, we will speculate on possible reasons why animacy seems to affect the use of progressives and genitives differently in our historical New Zealand data.

2. Studying early New Zealand English

2.1 The beginnings of New Zealand English

After the Treaty of Waitangi, in which the Maori chiefs yielded sovereignty to the British crown, had been signed in 1840, New Zealand was settled exclusively by free settlers — mostly from Britain. The settlement of New Zealand was spearheaded by the New Zealand Company in the early colonial period. In fact, some of the earliest texts that we have are letters that were published by the New Zealand

Company to encourage further migration to the new crown colony (see Hundt, submitted, for a detailed study of this material). The question is whether these early letters constitute New Zealand writing or are simply a form of transplanted BrE or early AusE (quite a few people migrated to New Zealand via Australia). The issue is further complicated by the fact that settlers did not simply “stay put” once they arrived in New Zealand but travelled back to Europe. In the later colonial history, even those who were New Zealand-born travelled and at times spent lengthy periods overseas.² The diachronic data that our study is based on therefore represents both early New Zealand “in the making” as well as later stages in the development of this (post)-colonial variety of English.

Gordon *et al.* (2004) describe the origins of the New Zealand accent on the basis of a unique set of data in the form of recordings made throughout rural New Zealand in the 1940s with various informants, including some of the first New Zealand-born speakers of English. Gordon *et al.* (2004: 258) ultimately argue for multiple factors, namely input from settlers who came via Australia and swamping effects from large-scale immigration in the 1870s in addition to the founder principle, whereas Trudgill (2004: 158–60), on the basis of the same data, makes an argument for largely independent (but to a certain extent parallel) dialect formation. However, the descriptions based on the early spoken New Zealand data focus on the emergence of the New Zealand accent and do not discuss the development of morphosyntactic features. This constitutes a still largely uncharted territory of research. Furthermore, the data are difficult to compare with similar evidence for BrE or AmE (which is not available).³

2.2 A Corpus of Early New Zealand English (CENZE) and ARCHER

The data for this study come from a small corpus of early New Zealand texts (CENZE).⁴ Some were digitized at the *New Zealand Electronic Text Centre*, others

2. The problems that these facts may pose for the compilation of a corpus of early New Zealand texts, as well as some sample biographies, are discussed in Hundt (in preparation).

3. For further discussions of the history and development of NZE, including lexis and morphology, see e.g. Bauer (1994), Schneider (2007: 127–33) and Hundt (2012).

4. John Macalister at Victoria University, Wellington, has compiled a larger corpus of early NZE, but the genres he sampled do not allow for easy comparison with historical corpora of BrE or AmE. Furthermore, Macalister’s corpus does not extend back in time beyond the beginning of the 20th century and therefore does not permit comparison of late 19th-century usage of progressives across varieties. For information on his 5.5-million-word corpus of early New Zealand writing, see Macalister (2006) and <http://www.victoria.ac.nz/lals/research/grinze/publications/Corpora_of_NZ_English.pdf> (accessed 4 June 2012).

(the *Proceedings of the Royal Philosophical Society of New Zealand*) by the National Library of New Zealand. Early New Zealand newspaper texts and letters from emigrants are available in electronic form on the World Wide Web. The texts were collected with the intent of compiling a corpus comparable to existing diachronic corpora of BrE and AmE, notably ARCHER (see below). Tables 1a-d in the appendix give an overview of the text categories, subperiods and size (number of words) of the subsamples in the corpus.

To put the CENZE-based findings into cross-varietal perspective, we also tap into ARCHER (Version 3.1). The corpus covers the period between 1650 and 1999, spans about 1.8 million words of running text, and samples eight different registers (drama, fiction, sermons, journals/diaries, medicine, news, science, letters) and the two major varieties of English: BrE and AmE. The corpus design categorizes all texts into seven subperiods of 50 years, although coverage of AmE in ARCHER 3.1 is restricted to three of the seven periods (see Yáñez Bouza 2011). For reasons to be discussed below, different sections of ARCHER and the CENZE corpus were used for our two case studies.

An obvious difference between ARCHER and the corpus of early New Zealand texts is that the beginnings of NZE only reach as far back as the middle of the 19th century. With the exception of private letters, text categories that are represented in ARCHER become available only later in the 19th century. Furthermore, the availability of material makes sampling by 50-year subperiods as in ARCHER somewhat difficult. ARCHER samples personal letters by highly literate, published authors (including some letters that Queen Victoria wrote to her mother), whereas the CENZE includes letters that were written by early emigrants and New Zealanders from more diverse social backgrounds (i.e. a few gentlemen, mostly craftsmen, but also labourers and some people whose social background remains unknown; see Hundt, in preparation). This is a detail that will have to be borne in mind for the interpretation of our findings.

Finally, note that the frequency profiles of the two phenomena we investigate are very different. Progressives are relatively infrequent and we therefore studied all our early New Zealand texts and include diachronic evidence for the first half of the 20th century. *S-* and *of-*genitives, by contrast, are fairly frequent, and to obtain a manageable dataset we thus restricted attention to interchangeable genitives in the early New Zealand news and letters genres, analysing a total of 51 texts: 30 letters, which cover the period between 1842 and 1929, and 21 news texts, which cover the period between 1868 and 1900. Genitive variability in this sample we compare to genitive variability in a temporally matching set of genitives drawn from ARCHER's news and letters sections.

3. Definition of the linguistic variables and contextual conditioning factors

3.1 Progressives

To allow for comparability with previous ARCHER-based studies (Hundt 2004a, b), the same criteria were applied in the definition of the progressive, i.e. combinations of the auxiliary *be* with a present participle (allowing for material to occur between auxiliary and participle) were extracted from the early New Zealand texts. From these, all non-progressives were manually removed, including instances where the participle has adjectival rather than verbal function (e.g. *This news is shocking* or *His countenance was repulsive and forbidding*) and examples with participles that function as an apposition rather than as part of the verb phrase (e.g. *He was at home, repairing the roof*; see Hundt 2004a: 56). Instances with two participles were only counted once. As in Hundt (2004a), instances of *going to* as a future time expression were excluded from the datasets.

The contextual variable ‘animacy’ is not a binary choice but a gradient (see also Sec. 3.2). In her seminal study on the spread of the progressive in Late Modern English, Strang (1982) distinguishes between ‘human’, ‘quasi-human or animal’ and ‘inanimate’ subjects. Collective nouns like *army* or *crew*, for instance, refer to entities that are not prototypically ‘human’ in that the neuter pronoun *it* can be used to refer to them, but they are certainly human-like because they are made up of a group of human beings (see Hundt 2004a: 48–51).⁵ Among the collective nouns, we also included the following use of *South Island* because it is used to refer to the population of the region rather than to the geographical area:

- (1) ... the South Island has hithero [sic!] made large sacrifices for the North, and is still, and is likely for years to come *to be making* such sacrifices.
(CENZE, NorthOtagoTimes_1868.new)

In Hundt (2004a: 56), nouns that were used metonymically for humans (e.g. *chair*, *house*) were also included amongst the animate subjects, but no such examples were found in our New Zealand data. Despite the gradience of the contextual variable ‘animacy’, a binary classification is applied in our analysis of progressives, distinguishing between animate nouns (humans, animals, collective nouns) and inanimate nouns (including body parts). The reason for this decision is that the relatively low text frequency of progressives in our data would not have supported

5. Note that Hundt (2004a: 59–61) found no difference in the overall diachronic trend towards a more frequent use of ‘inanimate’ and ‘non-agentive’ subjects, i.e. the general development was the same irrespective of the way in which the contextual variable was defined.

a more fine-grained classification of subjects. Examples in (2) illustrate typical animate subjects in our data, those in (3) typical inanimate subjects:

- (2) a. ... there have been bundles of telegrams. *Miss Maud* is opening them.
(CENZE, 1889_voge_f.fic)
- b. One thing is certain, *the Major* was doing his duty too stringently to retain his office.
(CENZE, TuapekaTimes_1869new)
- (3) a. ... and every one admitted, when the dessert was laid, and the *wine* was passing, that the result produced was glorious, and worthy of the occasion.
(CENZE, 1866_farj_f.fic)
- b. These are but a few names of rare birds from a *list* that is annually increasing;
(CENZE, 1869_Pott_VIII_z.sci)

3.2 Genitives

We are interested in interchangeable *s*-genitives, as in (4a), *vis-à-vis* interchangeable *of*-genitives, as in (4b).

- (4) a. ... it will be much better when *Captain King's fat cattle* comes from Sydney
(~ the fat cattle of Captain King; CENZE, 1842_x_new1)
- b. ... coal he has bought of *the captain of the Timandra*
(~ the Timandra's Captain; CENZE; 1842_x_new1)

Our definition of the genitive variable and of its most crucial conditioning factors follows the guidelines in Szmrecsanyi (*fc.*) and Wolk *et al.* (submitted). To obtain our dataset, we first identified genitive construction candidates in the dataset, and then hand-coded the material for interchangeability. Interchangeability is, of course, a Labovian criterion according to which either genitive construction must be paraphrasable by the alternative construction with no semantic change (see Sankoff 1998). This exercise yielded a dataset of $N=1\ 525$ interchangeable Early New Zealand genitives (81.6% of which are *of*-genitives, and 18.4% of which are *s*-genitives). The observations cover the period between 1842 and 1929. We supplement this dataset, for benchmarking purposes, with a parallel dataset drawn from ARCHER's BrE news and letters sections, which likewise covers the period between 1842 and 1929, and which has $N=1\ 020$ interchangeable genitives (*of*-genitives: 80.7%; *s*-genitives: 19.3%).⁶

We subsequently added a layer of rich contextual information. What will take centre stage in this paper is our annotation for possessor animacy, which

6. The ARCHER genitive dataset we analyze here is a proper subset of the more comprehensive dataset analyzed in Wolk *et al.* (submitted).

according to the literature is the most crucial conditioning factor in the genitive alternation: more animate possessors favour the *s*-genitive, less animate possessors favour the *of*-genitive. We distinguish between five animacy categories, which form a hierarchy in terms of their favouring effect on *s*-genitive choice (Rosenbach 2008: 153): animate > collective > temporal > locative > inanimate. We reiterate here that animacy effects are gradient and not restricted to the animate / inanimate distinction (which merely represents the extreme poles of the hierarchy). Thus, animacy effects can also manifest in the effect that intermediate categories (such as collective, locative, and temporal possessors) have on genitive choice.⁷ The categories we are considering were hand-coded based on the guidelines in Zaenen *et al.*'s (2004) coding scheme.

Animacy, however, is not the only factor playing an important role in genitive alternation, so we additionally annotated each instance of a genitive construction in our dataset for the following contextual variables:

- WEIGHT of the possessor and possessum phrase, in orthographic characters: according to the principle of end weight (Behaghel 1909/1910), long possessors should prefer the *of*-genitive while long possessums should prefer the *s*-genitive.
- GENITIVE RELATION: following Rosenbach (2002) we distinguish between prototypical semantic relations (legal ownership, body parts, kinship, and part-whole), which are supposed to prefer the *s*-genitive, and non-prototypical relations (e.g. *the law of the country*, *the name of God*), which are supposed to favour the *of*-genitive.
- FINAL SIBILANCY OF THE POSSESSOR: a final sibilant in the possessor can discourage use of the *s*-genitive (Altenberg 1982).
- REGISTER CATEGORIZATION (news *vs.* letters) of the source corpus text.

We note that the technicalities behind the annotation described in this section are exactly parallel to those discussed in detail in Wolk *et al.* (submitted).

3.3 Progressives and genitives: an interim discussion

The two phenomena that we focus on are interestingly different in nature. The use of animate and inanimate subjects with the progressive cannot be approached in a strictly variationist fashion because the progressive is undergoing

7. In addition, note that the intermediate categories are not marginal in terms of frequency; in the combined ARCHER/CENZE dataset, observations break down as follows: animate possessors — $N=741$; collective possessors — $N=322$; locative possessors — $N=211$; temporal possessors — $N=159$; inanimate possessors — $N=1\ 088$.

grammaticalisation and spread (including new grammatical contexts) in the period under investigation. It is therefore impossible to define a neat envelope of variation to measure the probabilistic impact of the factor ‘animacy’ on the use of progressives *vs.* simple verb phrases. Our discussion of animacy effects in the progressive construction will thus be based on a straightforward frequency analysis. The genitive alternation, on the other hand, is — in the period of time under consideration here — a well-behaved linguistic variable (*s-* vs. *of-*genitives) in the Labovian sense, which allows us to approach variable-internal changes with multivariate statistical modelling.⁸

Note that the way the variables are defined also has repercussions for the way they may be affected by differences in choice of topic of the texts in our corpora, and thus by potential skewing effects inherent in the sampling. Our findings on the co-occurrence of inanimate subjects with progressives could be subject to different overall text frequencies of different semantic types of subject in our corpus. By contrast, the multivariate modelling we apply to the genitive alternation is immune to differential input frequencies of animacy categories in the textual habitat. This is why choice of topic and related (potential) differences in the overall text frequency of inanimate subjects cannot skew the results we obtain for the genitive alternation.

Previous studies (e.g. Leech *et al.* 2009) have further shown that the two variables take part in different kinds of stylistic change. The progressive, for instance, is a construction that has been discussed in the context of the ongoing colloquialisation of written language in BrE and AmE (Leech *et al.* 2009:239), whereas the shift towards *s-*genitives contributes to the densification, or “eco-nomization” (Hinrichs and Szmrecsanyi 2007; Szmrecsanyi and Hinrichs 2008), of content observed in some written genres (Leech *et al.* 2009:250). Both trends are apparent in 20th-century newspaper writing, whereas a text type like personal letters is unlikely to be affected by densification.

4. Results and discussion

4.1 Progressives

Previous studies have shown that the progressive was used more frequently with inanimate subjects before its text frequency increased in the 20th century (Hundt

8. Notice, however, that the *s-*genitive can be argued to have been subject to degrammaticalisation in the Late Modern English period (see Szmrecsanyi *fc.* for a discussion). However, any such degrammaticalisation effects are extremely subtle and are not associated with major re-analysis processes, which is why positing a unifying envelope of variation for the past few hundred years is unproblematic.

2004a). Furthermore, NZE has been shown to use the progressive significantly more often than BrE and AmE in the 20th century (Hundt 1998; Collins 2009). Taken together, the results of previous studies lead us to expect that the progressive occurs more frequently with inanimate subjects in early New Zealand texts than in parallel texts from Britain and the US. The question is whether this goes hand in hand with an overall higher text frequency of progressives. In other words, we expect early NZE to be ahead of developments in the spread of the progressive to the environment that fosters an increase in its text frequency precisely because the progressive is used more frequently in this southern hemisphere variety in the 20th century.

The overall proportion of inanimate subjects in the BrE part of ARCHER for the relevant subperiod (1850–1899) is 21.5%; there is no statistically significant regional difference between BrE and AmE at this point in time, with AmE files in ARCHER yielding 22.5% inanimate subjects (Hundt 2004a:60, 69). But one important finding of the previous study was that the proportion of inanimate subjects with progressives differed quite markedly from one text type to another. For better comparability with the early New Zealand data, we therefore only use the results from the newspaper texts, fiction, scientific writing and private letters that fall within the second half of the 19th century. Note that for the New Zealand sample, twice as many newspaper texts (approximately 20 000 words each for the 1860s and 1890s) are included than in the ARCHER subsample (a total of approximately 20 000); nevertheless, overall raw frequencies of progressives in the New Zealand corpus are considerably lower than in the ARCHER subsample. This is a somewhat surprising result, because progressives in 20th-century NZE are significantly more frequent than in standard northern hemisphere varieties (see Hundt 1998; Collins 2009). As far as the proportion of inanimate subjects is concerned, it is highest in our New Zealand data (see Table 1).

Table 1. Semantics of the subject with progressives (including passive progressives) in ARCHER and CENZE (1850–1899); figures based on categories n, f, s and x (ARCHER) only⁹

	animate	inanimate
BrE	238	43 (15.3%)
AmE	163	52 (24.2%)
NZE	129	56 (30.3%)

Note that on the basis of a subset of the ARCHER data, the difference between BrE and AmE does prove significant (at $p=0.0175$, 1df) in a chi-square test. The

9. Complete tables for the New Zealand data that show developments for individual text types (in some cases beyond the second half of the 19th century) are given in the appendix.

difference between the BrE and NZE sample is also highly significant ($p = 0.0002$, 1 df). The difference between our AmE and NZE data, however, is below the level of statistical significance. In other words, even though the overall text frequency of progressives in the AmE and NZE data is below that found in the BrE sample, the proportion of inanimate subjects is significantly higher. This confirms that the text frequency of progressives as such is not necessarily the most important factor for the semantics of the subject. However, the point made in Hundt (2004a) is that an increase of inanimate subjects with the progressive preceded the subsequent spread of the construction in both BrE and AmE. The data in Table 1 do not support this view: the fact that in the second half of the 19th century BrE has an overall higher text frequency of progressives than AmE and NZE, but a lower proportion of the construction occurring with inanimate subjects, throws some doubt on this causal interpretation. Moreover, our results do not confirm a split between northern and southern hemisphere varieties. Both NZE and AmE are more advanced in the spread of progressives to inanimate subjects than BrE is in the second half of the 19th century.

We already pointed out that an important factor in the previous study was text type. Hundt (2004a: 62) found that progressives with inanimate subjects were particularly frequent in medical and scientific texts (57% and 38% respectively), even though the overall text frequency of progressives in these registers was lower than in prose fiction. Table 2c in the appendix shows that in the science part of CENZE, inanimate subjects occur with an overall proportion of 54%.

In Table 2, the data are presented according to text type. They confirm the general trend from the previous ARCHER-based study: scientific texts show a very high propensity for progressives with inanimate subjects despite the overall low text frequency of progressives. Fiction, on the other hand, has the lowest proportion of inanimate subjects even though it has very high overall frequencies of progressives (the highest in the British and American samples). An obvious explanation lies in the very different subject matter of these text types: scientific texts are more likely to be about topics that trigger a high text frequency of inanimate or abstract nouns, whereas fictional writing and personal letters are likely to contain a high proportion of nouns that refer to humans. Similarly, certain sections of newspapers focus on topics with a 'human' interest and are therefore bound to contain more animate nouns. One would expect these characteristics of text types to be similar across different regional varieties (see Leech *et al.* [2009: 213] on changes in noun types in 20th-century BrE and AmE). The results in Table 2, however, indicate that this may not be the case. (Note that for the science genre, raw frequencies are too low to make the calculations of percentages meaningful.)

Table 2. Semantics of progressive subjects (animate : inanimate) across different text types in ARCHER (1850–99) and CENZE

	BrE (% inanimate)	AmE (% inanimate)	NZE (% inanimate)
News	91:14 (13.3%)	1:12 (22.6%)	53:25 (32.1%)
Fiction	114:15 (11.6%)	102:32 (23.8%)	43:11 (22.3%)
Science	6: 7	0: 5	2: 6
Private letters	27: 7 (20.6%)	20: 3 (13%)	31:14 (31.1%)

There are some regional differences in the use of progressives with inanimate subjects across different text types: both AmE and NZE newspapers have higher proportions of progressives with inanimate subjects than BrE, but it is only the difference between BrE newspapers and NZE newspapers that proves statistically significant ($p=0.004$, 1df). There also appears to be a tendency for New Zealand writers to use the progressive more often with inanimate subjects in private letters, but the raw frequencies are rather low. It is less likely that the social background of the letter writers may have affected the semantics of the subject in progressive constructions: despite the similar social backgrounds of the British and American subsamples in ARCHER, there are slightly more inanimate subjects in the British than in the American subsample. With fiction, it is in the American part of ARCHER that we find significantly more inanimate subjects than in the British part ($p=0.051$, 1df). Whether these differences are genuine regional differences in the effect that animacy has on the use of progressives or whether they reflect differences in the choice of topics is a different matter. Impressionistic evidence suggests, for instance, that newspaper articles and letters from New Zealand often mention geographical features of the newly established colony (*land claims, drainage, paddocks, etc.*) and its crops (*flax, wool, gold, etc.*):

- (5) a. The left-hand branch of Topsy Creek is being sluiced in a face by Crow and Co. (CENZE, GreyRiverArgus_1868.new)
- b. The country in this neighbourhood is only becoming known; no sooner is one valley explored and surveyed, than another is discovered contiguous to it. (1842_wel2.let)
- (6) a. Over the area on which this flax was bleaching were a number of short posts ... (CENZE, TimaruHerald_1868.new)
- b. ... the grass was getting very dry and the creeks low, ... (CENZE, 1868_barkxxxiii_x.let)

Table 2d in the appendix indicates that in private letters, for instance, the topics that people write about may play a significant role: the proportion of progressives with inanimate subject in this text type decreases over time from around 30% to

only around 15%. The question is therefore whether the results obtained for the co-occurrence of the progressive with animate subjects simply reflect a change in the proportion of animate subjects overall or one that is specific to this grammatical construction.

One way to verify whether the animacy effect observed across different text types is due to regional preferences rather than choice of topic is to compare the frequency of different semantic classes of nouns across corpora. Such a comparison, if done automatically, would need to combine semantic classes with syntactic relation, i.e. give a measure of animate *vs.* inanimate nouns in the subject function across the different corpora. This was beyond the scope of the present paper. As a first approximation, however, 100 sentences were randomly sampled from the newspapers and letters section of CENZE and ARCHER and manually analysed according to the semantics of the subject. (Note that these sentences include non-progressive VPs as well.)

Table 3. Text frequency of inanimate subject (100 randomly sampled sentences)

	BrE (ARCHER)	NZE (CENZE)
News	46%	50%
Letters (1850–99)	22%	29%
Letters (1920s)	—	32%

The results in Table 3 indicate that the text frequencies of inanimate subjects in the two corpora are very similar, and therefore the differences in the type of subject with the progressive reported in Tables 1 and 2 are likely to be due to regional preferences. Larger, more socially stratified samples are needed to verify whether the social background of the writers may have had an influence on the semantics of the subject. We also note that the proportion of inanimate subjects in the letters part of the corpus appears to be rather stable, i.e. the decrease in the proportion of inanimate subjects with progressives cannot be attributed to there being fewer animate subjects overall. Finally, we investigated the overall text frequency of animate subjects in the science part of the CENZE corpus from the second half of the 19th century. The results confirm that science texts yield a higher proportion of inanimate subjects than news and letters: 61% of the subjects in the randomly sampled set of sentences were inanimate, and the likelihood that a progressive could be used with an inanimate subject is thus higher in this text type, despite the overall lower discourse frequency of progressives in science texts.

As far as diachronic developments are concerned, the data in our New Zealand corpus indicate that the progressive is used less frequently with inanimate subjects at the beginning of the 20th than in the second half of the 19th century. This trend

can be seen in a comparison of letters and newspaper writing from the 1860s and 1920s (Table 4) as well as in a comparison of all text types in our corpus (Table 5).

Table 4. Semantics of the subject with progressives (including passive progressives) in letters and newspaper texts from CENZE — diachronic development

	animate	inanimate
1860s	51	31 (38.3%)
1920s	122	41 (25.2%)

Table 5. Semantics of the subject with progressives (including passive progressives) in CENZE — diachronic development

	animate	inanimate
1850–99	129	56 (30.3%)
1900–49	213	67 (23.9%)

4.2 Genitives

Recall that the genitive alternation is a linguistic variable in the Labovian sense as it is sensitive to a number of conditioning factors (or: predictors). To gauge the direction and effect of animacy, we utilize binary logistic regression, a multivariate procedure that integrates probabilistic statements into the description of performance (see Szmrecsanyi 2006 for an accessible introduction).¹⁰ The procedure is applicable “wherever a choice can be perceived as having been made in the course of linguistic performance” (Sankoff 1998: 151). Logistic regression analysis has the following key features: 1) it seeks to predict a binary outcome (i.e. a linguistic choice) given several independent (or predictor) variables; 2) it quantifies the influence of each predictor; 3) it specifies the direction of the effect of each predictor; and 4) it states how well the model fares in predicting actual speakers’ choices.

Table 6 sketches the fixed effects in a logistic regression model that predicts *s*-genitive outcomes in the combined ARCHER and CENZE dataset. Overall, the model is a fairly accurate one: it correctly predicts 89% of all genitive outcomes in the dataset. Somers’ D_{xy} is .86, which indicates that the model discriminates satisfactorily between genitive types.

The figures reported in the table are so-called odds ratios (henceforth: ORs), which quantify the magnitude and the direction of the effect of each predictor

10. We are specifically using a modern refinement of logistic regression analysis, mixed-effects logistic regression (Pinheiro and Bates 2000). In addition to classical fixed effects, as shown in Table 6, the procedure can accommodate random effects, such as author idiosyncrasies, which we approximate by feeding corpus file IDs into the model.

Table 6. Genitive choice: logistic regression estimates for the combined ARCHER/CENZE dataset.

	odds ratio (OR)	
(Intercept)	2.00	**
possessor animacy (default: 'animate')		
temporal	.65	
collective	.24	***
locative	.09	***
inanimate	.03	***
genitive relation is prototypical (default: 'non-prototypical')		
possessor length (1 unit = 1 character)	.87	***
possessum length (1 unit = 1 character)	1.04	***
possessor has a final sibilant (default: 'final sibilant absent')	.56	***
register: news (default: 'letters')		
corpus / variety: CENZE (default: 'ARCHER')	1.41	.
possessor animacy × corpus (default: 'animate' / 'ARCHER')		
temporal × CENZE	.69	
collective × CENZE	.32	**
locative × CENZE	.06	*
inanimate × CENZE	.83	
N	2 521	
Somers' Dxy	.86	
AIC	1 435	
correctly predicted (baseline)	89% (81%)	

Predicted odds are for the *s*-genitive. Significance codes: '***' significant at $p=0.001$, '**' 0.01, '*' 0.05, '.' 0.1.

on genitive outcomes. ORs indicate how the presence or absence of a contextual linguistic feature (for categorical predictors *vis-à-vis* their default values, which are specified in Table 6 for each predictor¹¹) or a one-unit increase in a scalar predictor influences the odds for an outcome. In other words, for categorical predictors (such as CORPUS / VARIETY), ORs measure an effect against the backdrop of a default value; for scalar predictors (such as POSSESSOR LENGTH), ORs gauge the

11. Note that the selection of default values is arbitrary in principle, though typically the reference category is set as default. Thus, because we consider ARCHER the reference point in this study, we set 'ARCHER' as default for the categorical predictor CORPUS / VARIETY and subsequently determine the extent to which the value 'CENZE' makes a significant difference.

effect of one-unit increases. In any case, because ORs can take values between 0 and ∞ , three cases can be distinguished: 1) if $OR < 1$, the predictor makes a specific outcome less likely; 2) if $OR = 1$, the predictor has no effect whatsoever on the outcome; 3) if $OR > 1$, the predictor makes a specific outcome more likely.

With this information in mind, let us have a look at the effect directions of the major language-internal determinants of genitive choice in Table 6 (rows POSSESSOR ANIMACY through POSSESSOR HAS A FINAL SIBILANT). The predictors dis-favouring the *s*-genitive (because their ORs are less than 1) include all POSSESSOR ANIMACY categories except ‘animate’, POSSESSOR LENGTH, and FINAL SIBILANCY. The predictors favouring the *s*-genitive (because their ORs are greater than 1) include PROTOTYPICAL GENITIVE RELATIONS (recall from Sec. 3.2 that these comprise legal ownership, body part, kinship, and part-whole relations) and POSSESSUM LENGTH. This division of labour is precisely in line with the literature.

The two language-external predictors included in the model — REGISTER: NEWS and CORPUS/VARIETY: CENZE — both exhibit an OR greater than 1. This is another way of saying that the news genre and the CENZE material both favour the *s*-genitive compared to the letters genre and the ARCHER material, respectively (notice that ‘letters’ and ‘ARCHER’ are the default values here). The variety effect is only marginally significant, but as we shall see below it interacts in a significant fashion with the animacy constraint.

The magnitude of the probabilistic effects can be read off Table 6 as follows. Take the OR of 2.29 associated with the predictor GENITIVE RELATION IS PROTOTYPICAL. As we are dealing with a categorical variable, this OR indicates that compared to a non-prototypical genitive relation, a prototypical genitive relation increases the odds for an *s*-genitive by a factor of 2.29. Conversely, the OR of .87 associated with the continuous predictor POSSESSOR LENGTH indicates that for every one-character increase in the length of the possessor phrase, the odds for the *s*-genitive decrease by a factor of .87, i.e. by 13%.

What, then, is the effect that animacy has on genitive choice? Here, we are dealing with a significant difference between earlier BrE (ARCHER) and early NZE (CENZE), as can be seen from the interaction term POSSESSOR ANIMACY \times CORPUS in Table 6.¹² Precisely because of this interaction, and because the default value that the model uses for CORPUS is ‘ARCHER’, the main effects of animacy reported in rows 3 through 6 in Table 6 (“temporal” through “inanimate”) actually describe the situation in ARCHER. The coefficients reported in rows 14 through 17 of Table 6 (“temporal \times CENZE” through “inanimate \times CENZE”)

12. We note that except for final sibilancy, which unlike in the ARCHER data for some reason does not play a role in the CENZE data, the other language-internal predictors are not subject to a variety effect.

describe how animacy effects in CENZE differ from animacy effects in ARCHER. In ARCHER, therefore, inanimate possessors (as in *the quality of the soil*; CENZE, 1842_x_nel2) disfavour the *s*-genitive most robustly; the OR of .03 indicates that if the possessor is inanimate instead of animate, the odds for the *s*-genitive decrease by 97%. Locative possessors (as in *the interior of New Zealand*; CENZE, 1842_x_well1), with an OR of .09, are almost as disfavouring. Collective possessors (as in *the monthly meeting of the North School Committee*; CENZE, 1897_n_B North Otago Times) also disfavour the *s*-genitive (OR: .24). The ranking of these constraints thus roughly conforms to Rosenbach's (2008) animacy hierarchy, though in our dataset temporal possessors do not appear to have a significant effect.

It turns out that the CENZE material differs significantly from the ARCHER material in terms of the effect of two possessor animacy categories: 'collective' and 'locative'. While in ARCHER collective and locative possessors disfavour the *s*-genitive with probabilistic weights of .24 and .09, respectively, the corresponding weights in CENZE are $.24 \times .32 = .08$ and $.09 \times .06 = .01$, respectively. In other words, early NZE writers are substantially less inclined to use the *s*-genitive when the possessor is a collective or locative noun — i.e. when the possessor is not fully animate. Thus, early NZE writers' probabilistic genitive grammars exhibit a stronger animacy constraint. This is so even though this stronger animacy constraint finds expression primarily in intermediate animacy categories, such as collective or locative, which nonetheless are card-carrying members of the animacy hierarchy (see Sec. 3.2).

The discussion in the previous paragraph has centred on the effects that individual conditioning factors have. Let us now explore the overall importance of the factors in the model (and thus, their overall importance for predicting genitive choice). Figure 1 considers the joint impact of the predictors in the model and plots the increase in the model's *Akaike Information Criterion* (AIC; see Sakamoto and Akaike 1978) if a predictor is removed from the model; higher values correspond to increased overall importance. POSSESSOR ANIMACY (which in terms of Figure 1 captures the full-fledged five-category distinction) is thus the overall most crucial variable for predicting genitive outcomes in both ARCHER and CENZE. The important difference to note, though, is that while POSSESSOR ANIMACY in the ARCHER material is ahead of the other predictors but not by an exceedingly wide margin, POSSESSOR ANIMACY is hugely more important than the other predictors in the CENZE material. This finding dovetails nicely with Bresnan and Hay's (2008:246) observation, based on different datasets, that "New Zealand English speakers appear to be more sensitive [than AmE speakers; MH and BS] to the role of animacy".

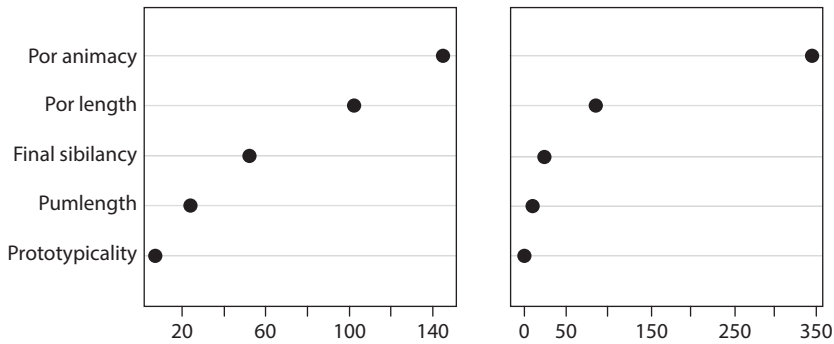


Figure 1. Importance of factors in model: increase in Akaike Information Criterion (AIC) if factor is removed. Left: ARCHER (BrE), right: CENZE (NZE).

5. Conclusion

Previous research has tentatively suggested that animacy effects in present-day spoken NZE differ from animacy effects in other varieties of English. In this study, we were interested in whether such differences have a history in earlier NZE or not. Thus, we revisited two grammatical phenomena — frequency variability of progressive constructions and constructional genitive variability — that are well known to be sensitive to animacy effects, and studied these phenomena in corpora sampling 19th- and early 20th-century written NZE as well as BrE and AmE. The verdict is that yes, there are indeed significant differences between earlier NZE and the other varieties in terms of the effect that animacy has — so the divergence, in terms of animacy, between NZE and BrE/AmE seems to be long-established.

Progressives are used more frequently with inanimate subjects in early New Zealand texts than in contemporary British writing, even though the overall text frequency of the construction is lower in the southern hemisphere variety. Thus, early NZE turns out to be conservative with respect to the ongoing spread of the progressive, but innovative in its extension to inanimate subjects when compared with BrE. This illustrates that a regional variety can simultaneously exhibit colonial lag and innovation (see Hundt 2009b). Somewhat surprisingly, there are no significant differences between our early New Zealand writing and contemporary American texts in the proportion of progressives with inanimate subjects. Contrary to the previous study (Hundt 2004a), significant differences between BrE and AmE emerged from the subcorpus used for this study. This suggests that the composition of the corpus (text type) may have an important influence on regional variation. Our study confirms that text type is an important factor, generally: scientific writing is more likely to yield high proportions of inanimate subjects with progressives than fiction. Preliminary evidence suggests that regional differences

in the use of the progressive with inanimate subjects are not due to differences in the overall text frequency of inanimate subjects. Because of the overall low text frequency of progressives, the results obtained in our pilot study need to be verified against larger amounts of data.

In the genitive alternation, we find a significant interaction of regional variety with animacy, such that some non-animate possessor categories are less likely to be used with an *s*-genitive in earlier NZE material compared to the BrE material. Also, animacy is overall clearly a more important factor for predicting genitive outcomes in the earlier NZE material compared to the BrE material. All this is another way of saying that the animacy constraint in genitive choice appears to be stronger in the early New Zealand data than in the BrE data. This is exactly what one would expect to see against the backdrop of the literature — in particular Bresnan and Hay (2008), who studied a different phenomenon (the dative alternation) in a different dataset and likewise obtained a stronger animacy effect in NZE than in another variety of English. Because we know from previous ARCHER-based research (Wolk *et al.* submitted) that in AmE and BrE animacy has been subject to weakening in the Late Modern English period, NZE can be argued to be rather conservative in terms of the comparatively strong effect that possessor animacy has on genitive choice.

References

- Altenberg, Bengt. 1982. *The Genitive v. the Of-Construction. A Study of Syntactic Variation in 17th Century English*. Malmö: CWK Gleerup.
- Bauer, Laurie. 1994. "English in New Zealand". In Robert Burchfield, ed. *The Cambridge History of the English Language*. Vol. 5: *English in Britain and Overseas: Origins and Development*. Cambridge: Cambridge University Press, 382–429.
- Behaghel, Otto. 1909/1910. "Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern". *Indogermanische Forschungen* 25: 110–42.
- Bresnan, Joan and Jennifer Hay. 2008. "Gradient grammar: an effect of animacy on the syntax of give in New Zealand and American English". *Lingua* 118: 245–59.
- Collins, Peter. 2009. "The progressive". In Pam Peters, Peter Collins and Adam Smith, eds. *Comparative Studies in Australian and New Zealand English: Grammar and Beyond*. Amsterdam: John Benjamins, 115–23.
- Corbett, Greville. 2005. "Sex-based and non-sex-based gender systems". In Martin Haspelmath, Matthew S. Dryer, David Gil and Bernhard Comrie, eds. *The World Atlas of Language Structures*. Oxford: Oxford University Press, 134–7.
- Dahl, Östen and Kari Fraurud. 1996. "Animacy in grammar and discourse". In Thorstein Fretheim and Jeanette K. Gundel, eds. *Reference and Referent Accessibility*. Amsterdam: John Benjamins, 47–64.
- Ellegård, Alvar. 1953. *The Auxiliary Do: The Establishment and Regulation of Its Use in English*. Stockholm: Almqvist & Wiksell.

- Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury and Peter Trudgill. 2004. *New Zealand English. Its Origin and Evolution*. Cambridge: Cambridge University Press.
- Hinrichs, Lars and Benedikt Szmrecsanyi. 2007. "Recent changes in the function and frequency of standard English genitive constructions: a multivariate analysis of tagged corpora". *English Language and Linguistics* 11: 437–74.
- Hundt, Marianne. 1998. *New Zealand English Grammar. Fact or Fiction?* Amsterdam: John Benjamins.
- Hundt, Marianne. 2004a. "Animacy, agency and the spread of the progressive in eighteenth- and nineteenth-century English". *English Language and Linguistics* 8: 47–69.
- Hundt, Marianne. 2004b. "The passival and the progressive passive — a case study of layering in the English aspect and voice systems". In Hans Lindquist and Christian Mair, eds. *Corpus Approaches to Grammaticalisation in English*. Amsterdam: John Benjamins, 79–120.
- Hundt, Marianne. 2009a. "Global feature — local norms? A case study on the progressive passive". In Thomas Hoffmann and Lucia Siebers, eds. *World Englishes — Problems, Properties and Prospects*. Amsterdam: John Benjamins, 287–308.
- Hundt, Marianne. 2009b. "'Colonial lag', 'colonial innovation' or simply 'language change'?" In Günter Rohdenburg and Julia Schlüter, eds. *One Language, Two Grammars? Differences between British and American English*. Cambridge: Cambridge University Press, 13–37.
- Hundt, Marianne. 2012. "Australian/New Zealand English". In Alexander Bergs and Laurel J. Brinton, eds. *English Historical Linguistics. An International Handbook*. (HSK 34–2). Berlin: Mouton de Gruyter.
- Hundt, Marianne. Submitted. "Heterogeneity vs. homogeneity". In Anita Auer, Daniel Schreier and Richard J. Watts, eds. *Letter Writing and Language Change*. Cambridge: Cambridge University Press.
- Hundt, Marianne. In preparation. "Towards a corpus of early New Zealand English — news from *Erewhon*?"
- Hundt, Marianne and Christian Mair. 1999. "'Agile' and 'uptight' genres: the corpus-based approach to language change in progress". *International Journal of Corpus Linguistics* 4: 221–42.
- Jaeger, T. Florian. 2006. "Redundancy and syntactic reduction in spontaneous speech". Ph.D. dissertation, Stanford University.
- Leech, Geoffrey, Marianne Hundt, Christian Mair and Nicholas Smith. 2009. *Changes in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.
- Macalister, John. 2006. "The Maori lexical presence in New Zealand English: constructing a corpus for diachronic change". *Corpora* 1: 85–98.
- Pinheiro, José C. and Douglas M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. New York: Springer.
- Rosenbach, Anette. 2002. *Genitive Variation in English: Conceptual Factors in Synchronic and Diachronic Studies*. Berlin, New York: Mouton de Gruyter.
- Rosenbach, Anette. 2008. "Animacy and grammatical variation — findings from English genitive variation". *Lingua* 118: 151–71.
- Sakamoto, Yosiyuki and Hirotugu Akaike. 1978. "Analysis of cross classified data by AIC". *Annals of the Institute of Statistical Mathematics* 30: 185–97.
- Sankoff, David. 1998. "Sociolinguistics and syntactic variation". In Frederick Newmeyer, ed. *Linguistics: The Cambridge Survey*. Vol. 4. *Language: The Socio-cultural Context*. Cambridge: Cambridge University Press, 140–61.

- Schneider, Edgar W. 2007. *Postcolonial English: Varieties Around the World*. Cambridge: Cambridge University Press.
- Strang, Barbara M.H. 1982. "Some aspects of the history of the be + ing construction". In John Anderson, ed. *Language Form and Linguistic Variation: Papers Dedicated to Angus MacIntosh*. Amsterdam: John Benjamins, 427–74.
- Szmrecsanyi, Benedikt. 2006. *Morphosyntactic Persistence in Spoken English. A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis*. Berlin, New York: Mouton de Gruyter.
- Szmrecsanyi, Benedikt. fc. "The great regression: genitive variability in Late Modern English news texts". In Kersti Börjars, David Denison and Alan Scott, eds. *Morphosyntactic Categories and the Expression of Possession*. Amsterdam: John Benjamins.
- Szmrecsanyi, Benedikt and Lars Hinrichs. 2008. "Probabilistic determinants of genitive variation in spoken and written English: a multivariate comparison across time, space, and genres". In Terttu Nevalainen, Irma Taavitsainen, Päivi Pahta and Minna Korhonen, eds. *The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present*. Amsterdam: Benjamins, 291–309.
- Trudgill, Peter. 2004. *New Dialect Formation: The Inevitability of Colonial Englishes*. Edinburgh: Edinburgh University Press.
- Wolk, Christoph, Joan Bresnan, Anette Rosenbach and Benedikt Szmrecsanyi. Submitted. "Dative and genitive variability in Late Modern English: exploring cross-constructural variation and change".
- Yáñez Bouza, Nuria. 2011. "ARCHER past and present (1990–2010)". *ICAME Journal* 35: 205–36.
- Zaenen, Annie, Jean Carlette, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, M. Catherine O'Connor and Tom Wasow. 2004. "Animacy encoding in English: why and how". In Donna Byron and Bonnie Webber, eds. *Proceedings of the 2004 ACL Workshop on Discourse. Barcelona: Association for Computational Linguistics*, 118–25.

Appendix

Table 1a. Newspapers in *CENZE*

	words
1860s	20 437
1890s	20 372
1920s	21 215

Table 1b. Fiction texts in *CENZE*

	words
1850–99	20 969
1900–49	20 855

Table 1c. Science texts in *CENZE*

	words
1870s	20 266
1900s	14 390
1930s	20 776
1960s	20 429

Table 1d. Private letters in *CENZE*

	words
1840s	20 364
1860s	20 790
1920s	20 709

Table 2a. Semantics of the subject in progressives (NZ newspaper texts)

	animate	inanimate	total
1860s	20	15 (42.9%)	35
1890s	33	10 (23.3%)	43
1920s	32	25 (43.9%)	57
Total	85	50	135

Table 2b. Semantics of the subject in progressives (NZ fiction)

	animate	inanimate	total
1850–99	43	11 (20.4%)	54
1900–49	73	9 (11%)	82
Total	116	20	136

Table 2c. Semantics of the subject in progressives (NZ science)

	animate	inanimate	total
1870s	2	6	8
1900s	16	7	23
1930s	1	3	4
1960s	1	7	8
Total	20	23 (53.5%)	43

Table 2d. Semantics of the subject in progressives (NZ private letters)

	animate	inanimate	total
1840s	63	25 (28.4%)	88
1860s	31	14 (31.1%)	45
1920s	90	16 (15.1%)	106
Total	184	55	240

Authors' addresses

Marianne Hundt
Englisches Seminar
Plattenstrasse 47
CH-8032 Zürich

m.hundt@es.uzh.ch

Benedikt Szmrecsanyi
Freiburg Institute for Advanced Studies
Albertstr. 19
D-79104 Freiburg im Breisgau

bszm@frias.uni-freiburg.de