John Benjamins Publishing Company

The genitive alternation in Chinese and German ESL learners

Towards a multifactorial notion of *context* in learner corpus research

Stefan Th. Gries and Stefanie Wulff

University of California, Santa Barbara / University of Florida

This paper exemplifies an approach to learner corpus data that adopts a multifactorial definition of 'context'. We apply a logistic regression to 2,986 attestations of the genitive alternation (*the squirrel's nest* vs. *the nest of the squirrel*) from the Chinese and German sub-sections of the International Corpus of Learner English and the British component of the International Corpus of English that were coded for 12 factors. Importantly, the speakers' L1 was included as a predictor to be able to compare properly the native speakers with the learners as well as the two learner groups with each other. The final regression model predicts all speakers' genitive choices very accurately (> 93%) and suggests that (i) the learners rely heavily on processing-related factors, which can be overridden by semantic constraints, and (ii) learners' choices are differentially modulated by their L1. We close with a discussion of how this context-based, multifactorial approach goes beyond traditional learner corpus research.

Keywords: genitive alternation, learner corpus research, native speaker, logistic regression

1. Introduction

Corpus-based research on the second language acquisition of English is a fast-growing strand in SLA research, thanks to the growing availability of learner corpora such as the International Corpus of Learner English (ICLE) compiled under the leadership of the research team around Sylviane Granger at the Catholic University of Louvain. This development has allowed researchers to examine new phenomena, formulate new hypotheses, and ultimately refine existing theories about

second/foreign language development. One of the central goals of learner corpus research is to "uncover factors of 'foreign-soundingness'" (Granger 1996:43). An early view on foreign-soundingness is articulated by Krzeskowski (1990):

> In either case the learner deviates in plus or minus from a certain statistical norm which characterizes native performance in a particular language. To ascertain such an error [though see below], one has to perform a quantitative contrastive study of texts written by native users of a particular language and by a non-native user of the same language and compare the frequencies of use of the investigated forms.   (Krzeskowski 1990:206, quoted from Granger 1996:45)

Granger herself adopts a more nuanced picture, clarifying – correctly, we think – that over- or underuses make up, or contribute to, the "foreign-soundingness *even in the absence of downright errors*" (Granger 2004:132, our emphasis). Specifically, the analysis of foreign-soundingness involves "bring[ing] out the words, phrases, grammatical items or syntactic structures that are either over- or underused by the learner" (Granger 2002:132).

This influential approach has led to a wealth of results. However, in spite of the successful application of this perspective/method to (native and learner) corpus data, we believe that, methodologically speaking, current learner corpus research largely under-utilizes the potential of corpus data. In this paper, we specifically focus our attention on two aspects in regard to which corpus data are frequently under-utilized: 'linguistic (and extra-linguistic) context' and 'multifactoriality'.

As to the first aspect, a frequent disregard for (details of) context, Granger (1996) herself recognized early on that "[t]he contrastive investigation of raw frequencies [is …] undoubtedly the least sophisticated type of quantitative comparison" (Granger 1996:45). The majority of studies to date, however, compare completely decontextualized frequencies of use of *x* in native language (NL) to frequencies of *x* in interlanguage(s) (IL) and then interpret these frequency differences. While such studies can potentially be revealing, we would like to argue that, often, they are not. For example, one case study of Hasselgård & Johansson (2011) compares relative frequencies of *quite* in the Louvain Corpus of Native English Essays (LOCNESS) and four components of the ICLE and comments, among other things, that the learners overuse *quite*. However, the first part of this case study fails to take context into consideration in two ways: first, since the first part of Hasselgård & Johansson's (2011) case study does not consider the contexts of the uses of *quite*, they in effect consider every word a slot in which *quite* could have been used. This is an unrealistic assumption which, methodologically speaking, entails computing the relative frequencies with the frequency of *quite* in the numerator and the corpus size in words in the denominator. Second, ignoring the context(s) of *quite* also allows for the very real possibility that the learners actually

did not overuse *quite* at all, but used more contexts of (a) type(s) in which even NS would typically also use *quite*, and within these contexts the learners actually use *quite* exactly as NSs would have used it. Linguistically, that means the overuse would be not one of *quite* – it would be one of, say, one or more lexico-syntactic contexts and, within these, there is a perfectly natural use of *quite*; technically, that means the denominator for the relative frequencies of *quite* would be the number of contexts in which *quite* (or a functional near-equivalent) could have been used. Unfortunately, this is a risk that the vast majority of studies involving over-/underuse frequencies come with – to avoid spurious overgeneralizations, learner corpus research needs a more fine-grained contextually-grounded approach in the sense of not just counting uses of *x*, but also inspecting each and every context of *x* in sufficient detail.

The second aspect we want to focus on here, the lack of multifactoriality in most current learner corpus research, is theoretically independent of that of context, yet practically usually coincides with it. The notion of multifactoriality entails questions of (i) the number of factors that co-determine when a form *x* is used (or when *x* is used rather than a functional near-equivalent *y*); (ii) how many of these factors are in fact included in a study; and (iii) whether all these factors are included at the same time. One case combining all these aspects is the first part of Hasselgård & Johansson's (2011) first case study on *quite*, in which no factors are considered at all. In the second part of their case study, however, one linguistic contextual factor is examined, namely the pattern in which *quite* is used (with an adjective, with an adverb, with a predicate, etc.). Nevertheless, there are problems with their approach. Not only does it stand to reason that native and non-native speakers' use of *quite* is not only determined by the grammatical pattern in which *quite* occurs, but by a variety of other factors; moreover, the authors report the co-occurrence frequencies of *quite* in grammatical patterns normalized against the corpus size – it might be more intuitive to normalize against the frequency of the grammatical patterns (or that of competing expressions).

Even when multiple factors are considered, they can be considered in different ways:

(i) monofactorially, i.e. each in isolation: what does factor *A* do with regard to the use of *x* (ignoring the simultaneous effects of factors *B*, *C*, etc.)?

(ii) multifactorially$_1$: what does factor *A* do with regard to the use of *x* taking into consideration the simultaneous effects of factors *B*, *C*, etc. but not their interaction(s)?

(iii) multifactorially$_2$: what does factor *A* do with regard to the use of *x* taking into consideration the simultaneous effects of factors *B*, *C*, etc. and their interaction(s)?

Approach (i) is predominant in current learner corpus research. This stands at odds with established findings on (language) learning from cognitive, psycho-linguistic, and psychology-of-learning perspectives (see Ellis & Ferreira-Junior 2009 in an SLA context and Saffran 2003, Ellis & O'Donnell 2012, or Gries 2012 in general). Moreover, it fails to answer Krzeszowski's (1990: 212) call for "massive statistical research" and, thus, take advantage of statistical methods that are bet-ter tailored to capture the context-constrained and multifactorial nature of lan-guage development. Current models of speech comprehension and production, for instance, emphasize that any given factor impacting speaker choice does not necessarily have the same impact across all contexts. Rather, depending on many contextual aspects of the speech situation, one factor (such as animacy) may favor a particular choice in some contexts and disfavor it in others. An analysis that aims to cover a linguistic phenomenon comprehensively must test many (linguis-tic and extra-linguistic) predictors and their interactions at the same time. In SLA research in particular, this means that interactions of linguistic factors (length, animacy, etc.) must be included alongside the L1 of the speakers in order to deter-mine whether the linguistic factors have the same effect for NSs and NNSs alike.

In this paper, we showcase how fine-grained and precise results can be ob-tained once context is given more serious and multidimensional consideration and the statistical analysis is correspondingly more refined. We present a contras-tive multifactorial regression analysis of two ILs (Chinese and German) and NL (English) that includes various linguistic and extra-linguistic factors as well as their interactions. In doing so, we hope to provide an illustration of the useful-ness of such methodologically complex approaches; see Tono (2004) for an early example in this spirit and Deshors & Gries (forthcoming) and Gries & Deshors (forthcoming) for more in-depth discussion. Specifically, we focus on the genitive alternation as in Example (1):

(1)   a.   (*s*-genitive) The squirrel's$_{POSSESSOR}$ nest$_{POSSESSED}$
       b.   (*of*-genitive) The nest$_{POSSESSED}$ of the squirrel$_{POSSESSOR}$

While the genitive alternation has been studied extensively in NL (see below) and while the non-nativelike tendency to overuse the *s*-genitive has been referenced as one marker of fossilization (e.g. Olsen 1999), there is, to our knowledge, no corpus-based study to date on the genitive alternation in IL. Therefore, we here combine our interest to argue in favor of a methodologically complex approach to learner corpus research with a first exploration of the genitive alternation in Chinese and German IL.

1.1   Previous research on the genitive alternation in NL

In corpus-based SLA research, alternations have generally received only limited attention (for a few exceptions, see Gries & Wulff 2005, 2009; Callies & Szczesniak 2008; Martinez-Garcia & Wulff 2012). In contrast, as far as native English data are concerned, alternations are among the most intensively researched phenomena. Space does not permit a comprehensive discussion of the entire body of previous research on the genitive alternation; instead, we deliberately focus our discussion on those studies that investigated variables impacting the genitive alternation that are also potentially relevant in the current context; for a comprehensive overview of determinants of the genitive alternation put forward in previous research, see Rosenbach (2002). Accordingly, the subsequent discussion is restricted to research adopting a synchronic perspective (for discussion of the diachronic development of the two genitive forms, see, for example, Wolk et al. 2013), and since the IL data examined here exclusively come from academic writing, we will not concern ourselves with the highly register-, genre-, and dialect-dependent nature of the distribution of the two genitive forms (for that, see Szmrecsanyi & Hinrichs' (2008) excellent analysis tying together diachronic, geographical, and genre-related variables governing genitives in native English).

1.1.1   *Morphosyntactic and semantic determinants*
A variety of morphosyntactic and semantic determinants have been demonstrated to determine the choice of construction. One relevant property is 'animacy': animate possessors prefer the *s*-genitive while inanimate possessors prefer the *of*-genitive (Altenberg 1982, Leech et al. 1994, Biber et al. 1999). For example, *John's book* is preferred over *the book of John* because *John* is a human possessor and thus arguably more compatible with the *s*-genitive's prototypical semantics (possession, see below), while *the outcome of the study* is preferred to *the study's outcome*, where *study* is an inanimate possessor. As Rosenbach (2005) shows, this effect of animacy cannot be reduced to a correlate of end-weight preferences as suggested by Hawkins (1994: 424), but affects speakers' choices independently.

Another relevant feature of both NPs is their 'specificity': where applicable, the preferred alternation variant will be the one in which (the more) specific referent (typically morphologically marked with a definite article) precedes the less specific or non-specific referent (Rosenbach 2002). For example, *Jason's research interests* is more acceptable than *research interests of Jason* because *research interests* is not specific and, therefore, less specific than *Jason*; however, when a definite article is added to *research interests*, the *of*-genitive variant *the research interests of Jason* is just as acceptable as the *s*-genitive variant.

Rosenbach (2002) and Stefanowitsch (2003) demonstrate that the choice of construction is also impacted not just by morphosyntactic features that have broad semantic correlates, but furthermore constrained by the different meanings and functions that are associated with the two genitive variants. Stefanowitsch (2003) presents a detailed account of the different 'semantic relations' that are encodable by the two genitives; Table 1 provides an overview.

Table 1. Semantic relations encodable by the *s*-genitive and the *of*-genitive (with slight modifications from Stefanowitsch 2003:421)

Semantic relation	*of*-genitive	*s*-genitive
possessed-possessor	*the shoes of Kate *the train of John the budget of the university the assets of our company	Kate's shoes John's train the university's budget our company's assets
interpersonal relations	*the Emily of Diane *the girl of Kate the son of my neighbor	your Emily Kate's girl my neighbor's son
component-whole	the eyes of the baby the legs of the table the surface of the earth	the baby's eyes the table's legs the earth's surface
attribute-holder	?the coldness of Kate the beauty of the desert the design of the car	Kate's coldness the desert's beauty the car's design
participant-event	the investigation of the fire department	the fire department's investigation
time-event	??the olympics of last year ??the lecture of yesterday	last year's olympics yesterday's lecture
thing-constituent material	a dress of silk a sense of isolation	*the/a silk's dress *isolation's sense
subcategory-category	a dark kind of wood this sort of person	*the wood's dark kind (his sort)
subpart-whole (quantity)	50% of the population a big chunk of the company a glass of water a bowl of oranges	*the population's 50% *the company's big chunk *the water's glass *the oranges' bowl
depiction-depicted	a picture of Lisa a picture of the table the footage of the riot	(Lisa's picture) ??the table's picture *the riot's footage

Adopting the theoretical perspective of construction grammar, Stefanowitsch (2003) argues that the two variants are in fact semantic-role constructions that only partially overlap in terms of their semantic potential: "the *s*-genitive assigns

the roles POSSESSEE and POSSESSOR to its head and modifier respectively, and the *of*-genitive assigns roles that I have called, for want of a better term, ENTITY and INTRINSIC ENTITY" (Stefanowitsch 2003: 20).

1.1.2   *Processing- and parsing-related determinants*

Recent research has highlighted the significant role of processing- and parsing-related determinants of the genitive alternation. Several studies have considered the genitive alternation as primarily determined by English speakers' preference for *end-weight* (Behaghel 1909, 1910), which arguably increases processability (Kreyer 2003) and parsing efficiency (Hawkins 1994). In accord with this view, heavier or more complex possessor NPs are predicted to prefer the *of*-genitive since it places the possessor in phrase-final position; conversely, heavier possesseds should prefer the *s*-genitive (Quirk et al. 1985: 1282, Biber et al. 1999: 304). That is, *the nest of the squirrel that was hiding all the nuts below the tree* is preferred over *the squirrel that was hiding all the nuts below the tree's nest*.

An alternative view on the same example would be to ascribe the preference to speakers' sensitivity to the underlying 'syntactic branching' (as opposed to simply the weight of the resulting surface structure): postmodified possessed NPs prefer the *s*-genitive, while postmodified possessor NPs prefer the *of*-genitive. To give an example of the latter, *the nest of Jackie, who is a squirrel* sounds considerably better than *Jackie, who is a squirrel,'s nest*.

Speakers' genitive choices have also been related to 'priming effects' (alternatively referred to as persistence phenomena), that is, the well-observed general tendency for speakers to reuse material that they have used or heard in the more or less immediate context. Szmrecsanyi (2006), for example, provides evidence that precedence of either construction in spoken and written discourse increases the odds of that construction being used again at the next given opportunity. At the same time, however, priming effects are apparently heavily constrained in the most local environment by what Rohdenburg (2003) calls the '*horror aequi* effect': formally identical structures in immediate adjacency are strongly dispreferred. For example, *Steffi's brother's dog* is avoided, rendering *the dog of Steffi's brother* the preferred genitive choice.

Finally, various earlier studies have argued that another factor impacting the choice of genitive construction is the NP referents' (relative degree of) 'givenness' (alternatively referred to as 'thematicity', 'topicality', 'identifiability', or, in a more psycholinguistically inspired parlance, 'activation status'; see Rosenbach 2002). In line with the well-attested principle of given-before-new in English, the preferred construction should be the one in which the referent of the NP that is established (more) in discourse precedes the referent that is less established or newly introduced to the discourse (Quirk et al. 1985, Osselton 1988, Biber et al.

1999). For example, in a text about a researcher, a mid-passage *this researcher's newest study* should be preferred over *a new study of this researcher*. More recent corpus-based research, however, suggests that, as far as the genitive alternation is concerned, the influence of givenness "is extremely tenuous, making it look like a stylistic principle rather than a fundamental cognitive mechanism" (Stefanowitsch 2003: 21). Gries (2002) and Hinrichs & Szmrecsanyi (2007) arrive at basically the same conclusion that information-packaging constraints only seem to be manifest in highly selective contexts, and that the systematic patterning of the two variants is better accounted for with recourse to other factors such as animacy and extra-linguistic constraints.

1.1.3 *Phonological determinants*

Couper-Kuhlen (1986: 60) suggests that there is a potentially universal principle called 'rhythmic alternation', according to which strong and weak syllables alternate. When applied to the genitive alternation, we would expect *the <u>students</u>' <u>voices</u>* to be preferred over *the <u>voices</u> of the <u>students</u>* since the former has strictly alternating stressed (underlined) and unstressed syllables, while the latter contains three unstressed syllables in sequence (*-ces of the*). While the universality of this principle is still hotly debated, English, as a prototypical stress-timed language, has been found to exhibit rhythmic alternation effects in at least some contexts. There are two reasons why it is interesting to examine this effect in more depth. First, there seems to be only a single study that explores the potential impact of rhythmic alternation on genitive choices: Shih et al. (in press: 23–24) conclude that "though its role may be small, rhythmicity still participates in the decision between genitive construction alternatives". However, they concede that their operationalization may well have been sub-optimal – they only considered the alternation at the boundary between possessor and possessed NP – and that a "more accurate measure of rhythm might be a more global one with wider scope, testing whether the rhythmic regularity is maintained throughout the genitive construction," (Shih et al. in press: 23–24) which is what we will do here. Second, as far as we know, there is not a single published study that explores this effect in learner language, which is interesting given its presumed universality.

The next potentially relevant determinant is closely tied to a morphological feature, namely 'number marking' on the possessor NP: compared to singular possessor NPs, plurals prefer the *of*-genitive; correspondingly, *the nest of the squirrels* would be preferred to *the squirrels' nest*. This patterning is most likely due to the fact that regular plurals end in an [s], resulting in a close succession of two [s] sounds, which in a way reminiscent of *horror aequi*, is a dispreferred

articulatory sequence (see Altenberg 1982, Shih et al. in press). Unsurprisingly, therefore, irregular plurals are associated with higher shares of the *s*-genitive than regular plurals (cf. Altenberg 1982, Plank 1985): *the mice's cheese* is preferred over *the cheese of the mice*.

A final phonological factor related to the preceding one is 'segment alternation', the preference for consonant-vowel (CV) syllable structure (also referred to as 'ideal syllable structure'; cf. Vennemann 1988: 13–29 and Schlüter 2003: Section 3.1). While most previous work has considered this feature to operate within the syllables of a word, we will explore here whether this tendency also operates between words. Thus, on the basis of segment alternation alone, *Anna's calf*, for example, should be less likely than the *calf of Anna* because in the former, two consonants "clash" at the word boundary, whereas in the latter, consonants and vowels alternate at both word boundaries.

1.1.4   *Extralinguistic determinants*

Next to the various linguistic determinants outlined above, recent research emphasizes the substantial impact of extra-linguistic factors on the choice of genitive. Several studies have argued that speakers' choice varies between *registers and genres* such that the *s*-genitive is generally more frequent in spoken language and informal genres (Altenberg 1982, Rosenbach 2002). However, Hinrichs & Szmrecsanyi (2007) qualify this general observation by pointing out that the *s*-genitive is also on the rise in written genres that strive for verbal economy, such as journalistic prose (see also Raab-Fischer 1995). Finally, the genitive constructions are also distinctively distributed across different *varieties of English*: for instance, Rosenbach (2002) points out that the *s*-genitive is more frequent in American compared to British English.

2.   Methods

In this section, we will first discuss the corpus data that were studied as well as their annotation for many of the above-mentioned variables (Section 2.1); then, we turn to explaining the specifics of our statistical analysis (Section 2.2).

2.1   Data

To study which of the above variables affect the genitive alternation (and how much so), and in order to identify potential differences (i) between NS and NNS and (ii) between NNS of different L1 backgrounds, we retrieved examples of

of- and *s*-genitives from three different corpora. As for NNS, we retrieved all *of-* and *s*-genitives from both the Chinese and the German parts of the ICLE (International Corpus of Learner English, version 2). Given the large overall number of matches, we decided to randomly sample 1000 candidates for *of-* and *s*-genitives from the full concordances; manual checking of these 2000 cases led to 10 and 6 cases (involving plurals) being discarded from the Chinese and German data respectively. While the corpus files will contain students at different proficiency levels, the random sampling ensures that no systematic/anticonservative bias results from the sampling. In addition, we randomly sampled *of-* and *s*-genitives from all files of the ICE-GB (British component of the International Corpus of English), which resulted in examples being taken from 303 of the 500 files, which arguably constitutes a good representation of the target variety of the learners. The overall distribution of the data is represented in Table 2.

Table 2. Distribution of genitives across the three samples

	of-genitive	*s*-genitive	Totals
NNS: Chinese	872	118	990
NNS: German	892	104	996
NS: English	817	183	1000
Totals	2581	405	2986

These matches were then annotated for a large number of variables; many of these have been argued to play a role in previous work and have been discussed above, some others we added because they turned out to be significant predictors in our own pilot studies. The following is a list of predictors that we included in our analysis:

I.   GENITIVE: *of* vs. *s*;
II.   SEMCLASS: all genitives were annotated for the semantic relation they encoded; we distinguished the following categories: *attribute/holder of attribute* vs. *possession* vs. *event/participant in event* vs. *whole/part of whole* vs. *category/subcategory* vs. *event/time of event* vs. *personal relation* vs. *depiction/depicted* vs. *thing/material of thing*;
III. POSSORNUMBER and POSSEDNUMBER: *singular* vs. *regular plural* (with *s*) vs. *irregular plural* (e.g. *children* or *women*);
IV.   POSSORANIM and POSSEDANIM: *inanimate* vs. *animate* (but not human) vs. *human*;
V.   POSSORSPEC and POSSEDSPEC: *specific* vs. *non-specific*;

VI. POSSORBRANCH and POSSEDBRANCH: *none* vs. *premodified* (e.g. *the natural environment*) vs. *post-modified* (e.g. *their right of choice of smoking*, where the underlined *of* is the genitive analyzed and *right* is post-modified by *of choice*) vs. *pre- and post-modified* (e.g. *the richest source of protein of all veggies*);

VII. POSSORCOMPLEXITY and POSSEDCOMPLEXITY: *simple* (e.g. non-modified nouns) vs. *intermediate* (nouns with adjectival or PP modification) vs. *complex* (nouns with clausal modification);

VIII. HORRORAEQUI: all genitives were annotated with regard to whether they contained additional genitives: *none* (e.g. *the parts of the Saudi desert*) vs. *of* (e.g. *the part of the map of Kent*) vs. *s* (*my neighbour's dung heap's odours*);

IX. LENGTHDIFF: the difference in number of characters of possessor minus possessed;

X. RHYTHALTDIFF: every phrase with a genitive and its other-genitive counterpart was coded for its sequence of *s̲*tressed and *u̲*nstressed syllables (*people's persona̲lities* = suuusuu; *persona̲lities of people̲* = uusuuusu). From these, we computed a value whose absolute size increases with the number of stress clashes (sequences of stressed syllables) and stress lapses (sequences of unstressed syllables) and whose sign (positive or negative) indicates which genitive has the preferred rhythmic alternation pattern;

XI. SEGALTDIFF: every phrase with a genitive and its other-genitive counterpart was coded for its two transitions from the end of one NP to the genitive marker and the genitive marker to the beginning of the next NP such that a CV/VC transition was scored as 0, a C_1C_2 transition (where $C_1 \neq C_2$) was scored as 1, and a C_1C_1 transition was scored as 2 (*Isobel_'s_grief* = 1+1 and *grief_of_Isobel* = 0+0, i.e. the difference is 2). Thus, the higher the absolute value of the difference, the more the sequence violates the segment alternation preference of 'ideal syllable structure', and the sign (positive or negative) was again chosen to represent which genitive has the preferred pattern;

XII. LANG: the L1 of the speakers: *English* vs. *Chinese* vs. *German*.

We then explored the data to avoid potential pitfalls for our multifactorial analysis. As a result of this exploration, we had to discard the variables POSSORCOMPLEXITY and POSSEDCOMPLEXITY because they introduced too high a degree of collinearity into the subsequent multifactorial model. Their impact will be represented by the predictors representing the lengths and the branching directions of the relevant NPs. In addition, the variables HORRORAEQUI and SEMCLASS had levels that were too infrequent to permit any statistical analysis, which is why HORRORAEQUI had to be discarded altogether and SEMCLASS was recoded into just three levels (based on their genitive preferences): *attribute/holder of attribute*

vs. *possession or personal relation or event/time of event* vs. *other*. Similarly, there were very few examples of NPs involving both pre- and postmodification, and these overwhelmingly patterned with the merely post-modified NPs, which is why these two categories were grouped together. Finally, the number of irregular plurals in our sample was too small to allow for meaningful regression modeling (especially for possesseds), which is why the regular and irregular plurals had to be grouped together. The resulting data – altogether 2,986 genitives manually annotated for 14 variables, amounting to 41,804 data points – were then analyzed statistically as described in the following section.

2.2   Statistical method

We analyzed the data with a binary logistic regression (using R 2.15.3, cf. R Development Core Team 2013; for regression/model selection, see Gries 2013). The dependent variable is GENITIVE, all other variables retained after the initial exploration were entered as independent variables. Because of large amounts of collinearity of several predictors, standard stepwise model selection procedures were not possible as they resulted in variance inflation factors (VIF) > 10. Thus, we used a hybrid step-wise approach and, given the absence of previous studies on the genitive alternation in IL, adopted a less conservative significance threshold of $p < 0.1$ to decrease the probability of type-II errors. We started out with a model that included all contextual and linguistic predictors and their interactions with LANG that did not involve empty cells and then successively removed all predictors with $p \geq 0.1$ and one predictor that was responsible for the high VIFs. Then, we proceeded in a stepwise fashion by adding interactions of predictors with LANG to the model if those predictors significantly improved the model and did not cause any VIF to exceed 10. After 17 stepwise additions, no more predictors could be added that would significantly ($p < 0.1$) improve the model without increasing collinearity.

3.   Results

3.1   General overview

The final model reveals a highly significant correlation between the predictors and speakers' genitive choices (likelihood ratio = 1394.7, $df = 33$, $p < 10^{-200}$), a correspondingly strong correlation ($R^2 = 0.68$), and a very high classification accuracy (93.3%, $C = 0.96$). There was no significant overdispersion (residual

deviance = 975.8; residual *df* = 2952). Table 3 provides the significance test for the highest-level predictors in the final model to be discussed below; the first nine are effects that apply to NS as well as both NNS groups (cf. Section 3.2.1); the remaining three are effects that distinguish between different speaker groups (cf. Section 3.2.2).

Table 3. Significance tests for the highest-level predictors

Predictor	*LR, df, p*	Predictor	*LR, df, p*
SEGALTDIFF	5.964, 1, 0.015	POSSORNUMBER : POSSEDBRANCH	19.214, 2, < 0.0001
POSSORNUMBER : POSSEDSPEC	6.957, 1, 0.008	POSSORNUMBER : POSSORSPEC	2.985, 1, 0.084
POSSORSPEC : POSSEDSPEC	3.167, 1, 0.075	POSSORNUMBER : LENGTHDIFF	6.542, 1, 0.011
POSSEDBRANCH : LENGTHDIFF	15.551, 2, < 0.001	POSSORBRANCH : LENGTHDIFF	6.725, 2, 0.035
POSSORANIM : LENGTHDIFF	7.777, 1, 0.005	SEMCLASS : LANG	8.253, 4, 0.083
POSSORSPEC : LANG	26.73, 2, < 0.0001	POSSEDSPEC : LANG	83.747, 2, < 0.0001

3.2   Specific results

In this section, we present graphs that show each significant predictor's effect on the probability of the *s*-genitive predicted by the regression model. In all graphs, we indicate the predicted probabilities of the *s*-genitive on the *x*-axis and their uncertainty with black filled circles and confidence intervals/bands. Where appropriate, we show the results of an interaction from complementary perspectives in two panels, which considerably facilitates the identification of different patterns, and sometimes we add predictions for a comparison category in gray. In such cases, we also use letters to highlight different predictor levels; these letters are underlined in the prose describing the graph. Finally, for the numeric predictor LENGTHDIFF, we show which values were observed in the data with rugs – very short horizontal lines – directly on the *y*-axis.

3.2.1   *Effects that apply to all (NS and NNS) speakers alike*
The first effect is a main effect, i.e. an effect that is not qualified by any other predictors and that therefore applies across all data. Figure 1 shows that all speakers' genitive choices do exhibit a significant but weak tendency towards ideal syllable structure.
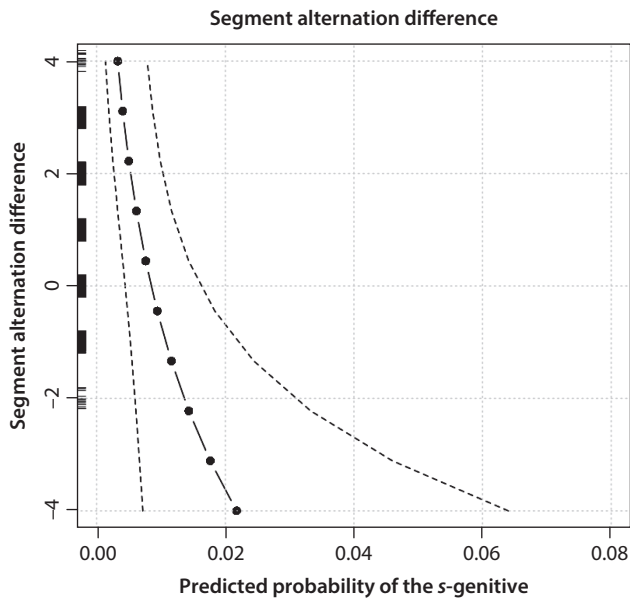
Segment alternation difference

Figure 1. The main effect SEGALTDIFF

Figure 2 shows that with plural possessors, branching of the possessed does not affect genitive choices much – *s*-genitives are dispreferred. However, with singular possessors, *s*-genitives become likely, particularly so with post-modified possesseds.
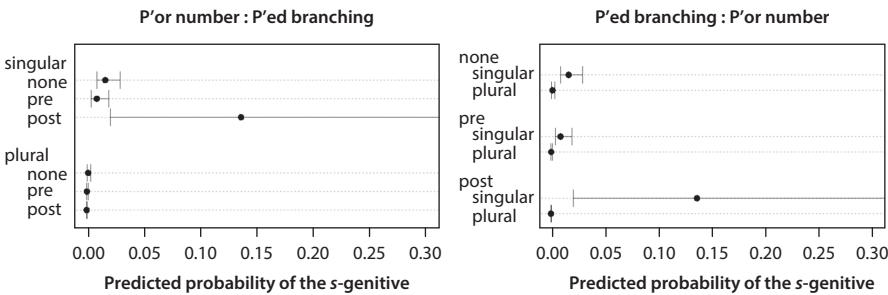

Figure 2. The interaction POSSORNUMBER : POSSEDBRANCH

Figure 3 reinforces that plural possessors disprefer *s*-genitives (see in particular the left panel), and it shows that specific possesseds also disprefer *s*-genitives (see in particular the right panel). However, the two predictors interact: non-specific possesseds increase the likelihood of *s*-genitives much more with singular than with plural possessors.
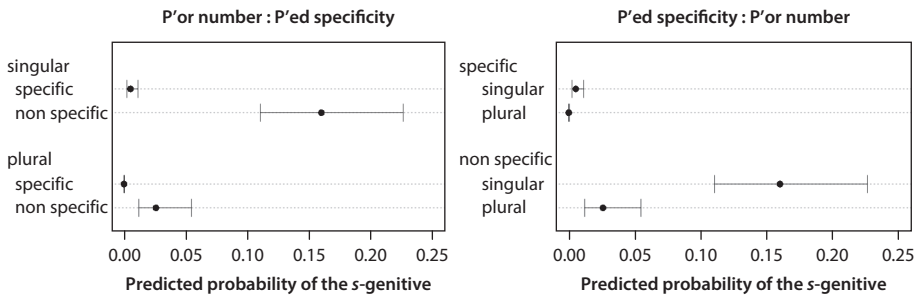
Figure 3. The interaction PossorNumber : PossedSpec

Figure 4 again shows that plural possessors disprefer *s*-genitives (see in particular the left panel). However, the picture is more complicated: recall from Figure 3 that with possess*ed*s, non-specificity makes *s*-genitives more likely; here it is the opposite: with possess*or*s, specificity makes *s*-genitives more likely (and again especially with singulars). Conversely, singular always makes *s*-genitives more likely, but more so with non-specific possesseds and specific possessors.
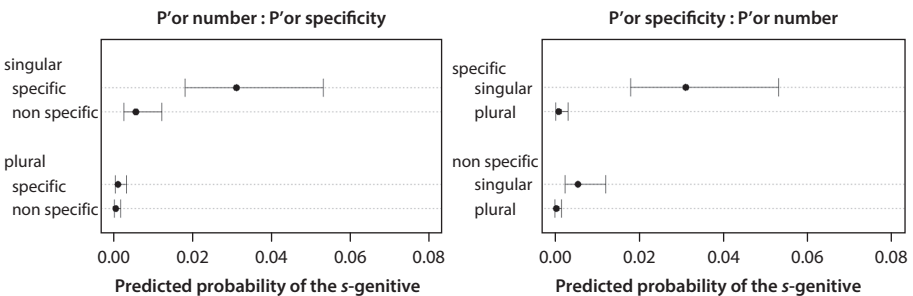

Figure 4. The interaction PossorNumber : PossorSpec

Given Figure 3 and Figure 4, it is unsurprising that both types of specificity interact, as is shown in Figure 5. Non-specific possesseds increase the likelihood of *s*-genitives, as do specific possessors, and when these two features coincide, the effect is particularly strong.
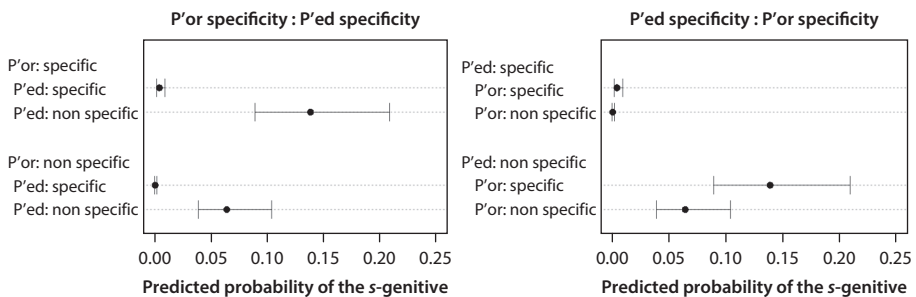

Figure 5. The interaction PossorSpec : PossedSpec

The next interactions are interesting because they reveal that a factor that virtually every analysis of the genitive alternation would have to mention – LENGTHDIFF – applies across all three speaker groups, but not across the entire range of the data. Figure 6 reveals that LENGTHDIFF has the expected effect of generally preferring a short-before-long arrangement of the two NPs. However, this is much more pronounced with s̲ingular possessors, as is highlighted in the left panel (with plu̲ral possessors added in gray for ease of comparison). With plural possessors, the avoidance of *s*-genitives is so strong that *s*-genitives only become more likely with considerably larger values of LENGTHDIFF, as we indicate with black plural and gray singular predictions in the right panel.
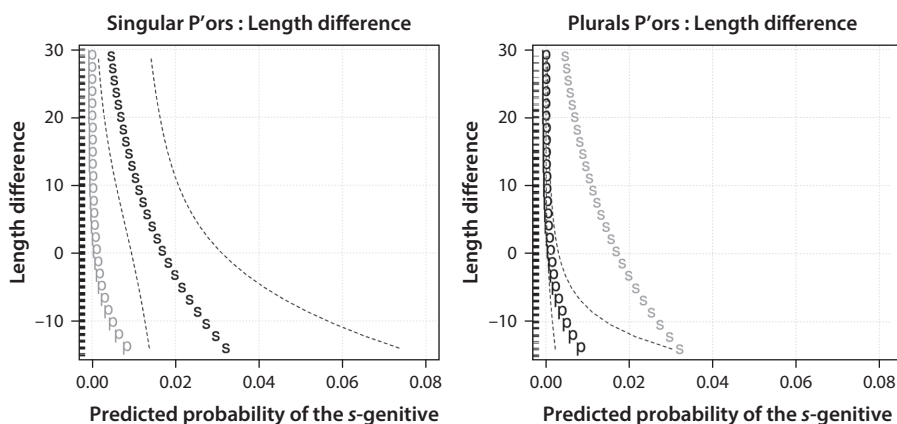

Figure 6. The interaction POSSORNUMBER : LENGTHDIFF

Figure 7 shows another interaction involving LENGTHDIFF, namely with the branching kind/direction of the possessor. On the whole, LENGTHDIFF has the expected effect, but its strength varies considerably. The left panel shows the effect of LENGTHDIFF for u̲nmodified possessors; the middle panel shows the effect of LENGTHDIFF with premodified possessors (i.e. on the l̲eft side of the noun); and the right panel shows LENGTHDIFF's effect with postmodified possessors (i.e. on the r̲ight side of the noun). As we can see in the middle and right panels, LENGTH-DIFF has less of an effect when the possessor is pre- or postmodified. Note in particular that while it may seem as if the strength of LENGTHDIFF with postmodified possessors at some point overtakes that of unmodified ones (beginning at LENGTHDIFF = −10, which means that the possessed NP is 10 or more characters longer than the possessor NP), the rugs on the *y*-axis in the right panel indicate that there are only very few such instances, so this rather underpopulated part of the curve must not be overinterpreted.
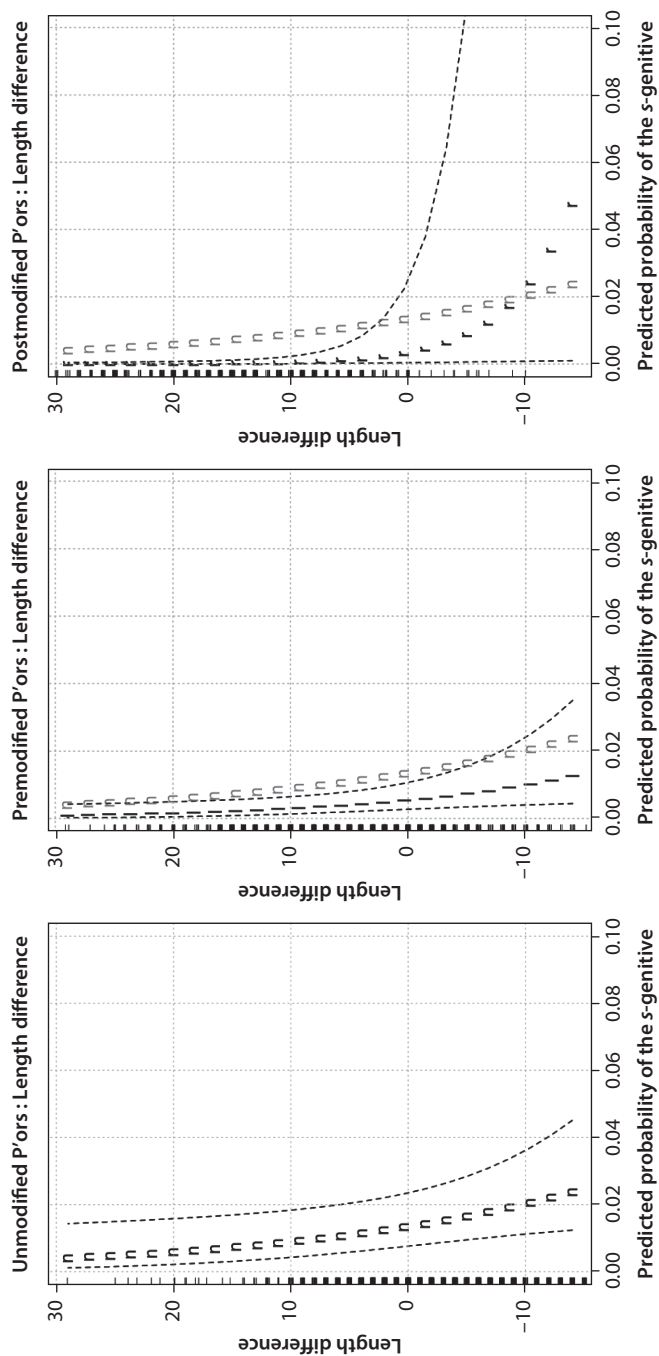
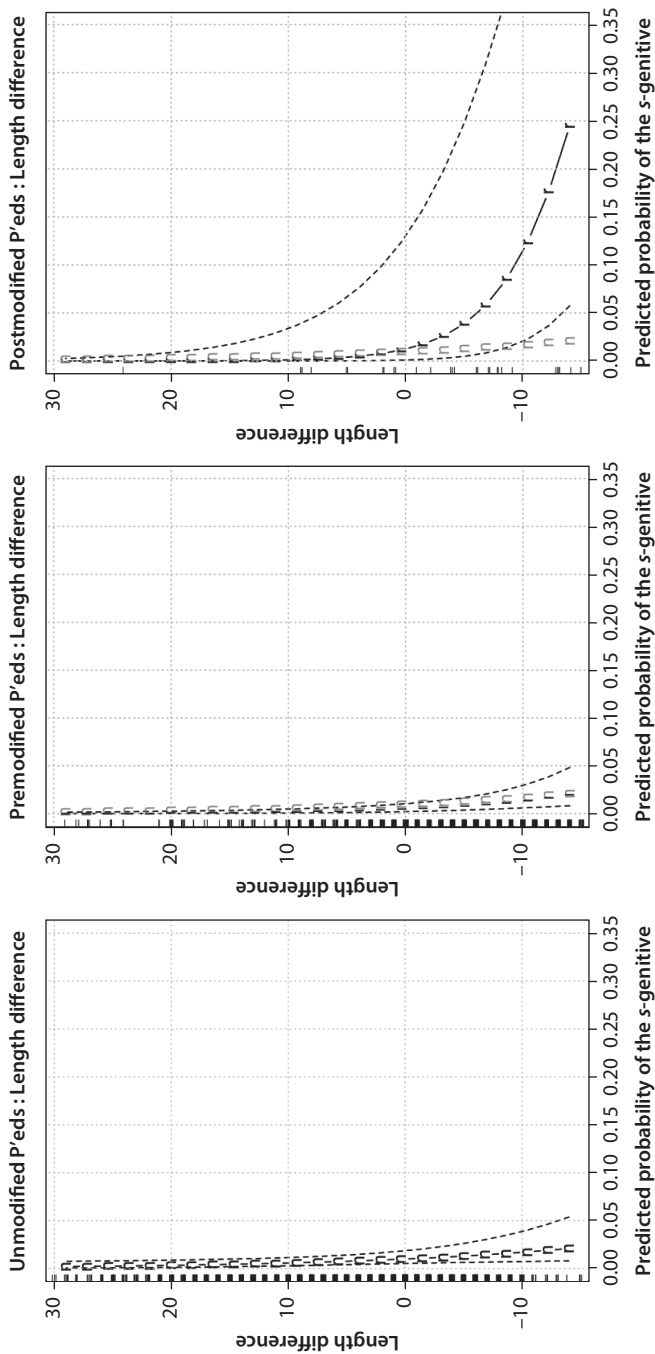**Figure 7.** The interaction POSSORBRANCH : LENGTHDIFF

Figure 8. The interaction POSSEDBRANCH : LENGTHDIFF

Figure 8 shows the corresponding interaction of LENGTHDIFF with the branching kind/direction of the possessed. Again, LENGTHDIFF has the expected effect everywhere, but it is quite weak with u<u>n</u>- and pre-/<u>l</u>eft-modified possesseds and stronger with post/<u>r</u>ight-modified possesseds. Again, one has to note the small number of postmodified possesseds, however.

Figure 9 is the final interaction applying to all speaker groups. The effect of LENGTHDIFF is also qualified by the animacy of the possessor: LENGTHDIFF has a strong effect if the possessor is <u>a</u>nimate, otherwise, i.e. if it is <u>i</u>nanimate, it does not.
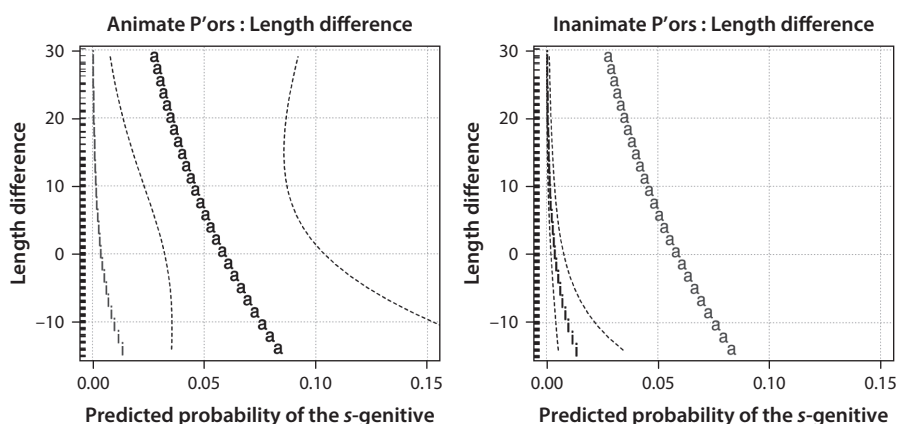

Figure 9. The interaction POSSORANIM : LENGTHDIFF

After having discussed the effects that hold for all speakers, we now turn to cases where the NS differ from the NNS, or where the two NNS groups differ from each other.

3.2.2   *Effects that do not apply to all speakers alike*

The first L1/speaker-specific effect involves SEMCLASS. The overall preferences of the semantic classes for *of-/s*-genitives are the same across the three speaker groups: *possessors*, *personal relations*, and *event/time of event* prefer *s*-genitives most, followed by *attribute/holder of attribute*, followed by *other*, which has a marked preference for *of*-genitives. However, Figure 10 shows that the degree to which these preferences are realized differ across the three groups. Specifically, the Chinese NNS exhibit the above differences, but they are so weak as to be not significant. The English NS, on the other hand, distinguish the semantic classes a bit more: the category with possessives and the category with others do not overlap, but the differences are still small. The German NNS, finally, exhibit larger

differences, larger preferences for *s*-genitives, and more diversity than both other speaker groups. This is quite apparent from the left panel, where the Germans' results always predict the *s*-genitive significantly more strongly.
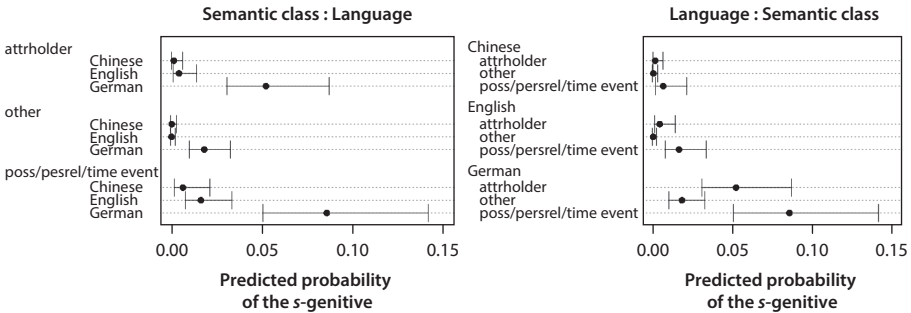

Figure 10.   The interaction SEMCLASS : LANG

Figure 11 shows that the above is not the only way in which the Germans differ from the other speakers. All speakers exhibit a non-significant preference of *s*-genitives with specific possessors, but the Germans have a significantly higher overall preference for *s*-genitives than the other speakers, while the Chinese and the English speakers do not differ from each other.
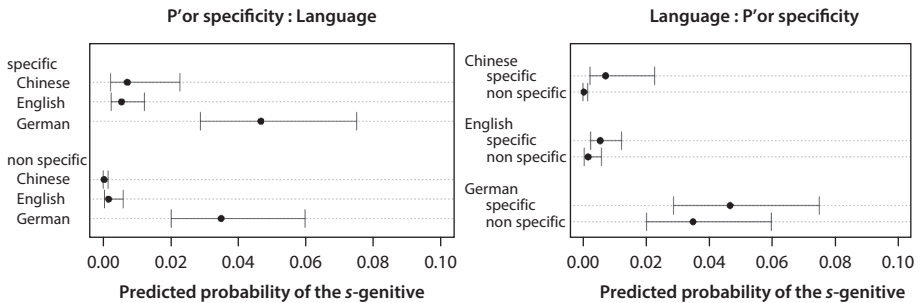
Figure 11.   The interaction POSSORSPEC : LANG

Figure 12 shows that the role of specificity is again more complex, however. The right panel indicates that, as above, all speakers prefer *s*-genitives with non-specific possessseds, but this interaction now shows how the speakers differ: the effect is smallest for the German NNS, the English NS exhibit an intermediately strong effect, and for the Chinese NNS, the specificity of the possessed makes the by far largest difference.

Figure 12. The interaction Possed Spec : Lang

In the next section, we will discuss these results and their implications in more detail.

4. Discussion and concluding remarks

Following the organization of Section 3, we will first discuss the findings that apply to all speakers, then we turn to those findings that vary as a function of L1 background.
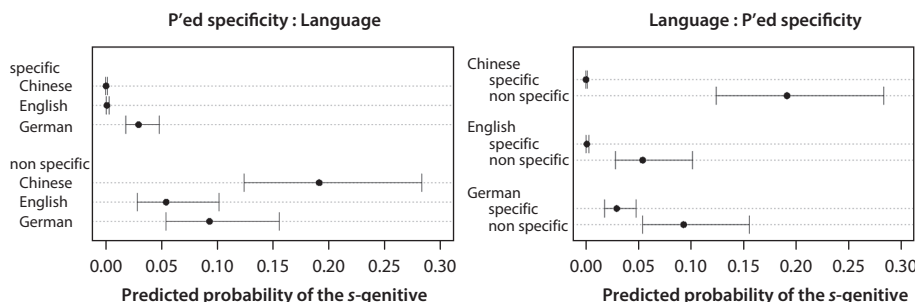
4.1 Commonalities of native language and interlanguage

Firstly, segment alternation was found to impact all speakers' choice of genitive such that speakers tend to choose that variant which results in the strictest possible alternation of consonants and vowels. This is an extremely interesting finding for two reasons. First, as far as we know, no previous study has explored the role of segment alternation for genitives – neither in NL nor in IL. The fact that this tendency surfaces as a main effect for all speakers therefore suggests that even IL is apparently sensitive to segment alternation constraints to some degree; this stands in accord with Schlüter's (2003) work on historical and present-day NS English. Second and more generally, one might have considered the inclusion of this variable irrelevant given that our corpus data consist of written language. However, the significant result and the fact that it cannot be explained away easily with reference to some confound(s) shows that even the written data reflect this phonological tendency, which provides further evidence for the claim that even written production can inform the analysis of spoken/phonological processes (cf. Myslín & Gries 2010 for additional examples). Finally, our result opens up

follow-up questions related to the presumed universality of this principle. To give but one example: if segment alternation was not universal in language X, would intermediate or advanced learners of English whose L1 is X still exhibit such an effect just because, as they become more advanced, they use more and more nativelike multi-word units, which in turn only arose, among other things, because English NS capitalized on the units' ideal syllable structure? Hopefully, future work will look into this.

Next to this global effect of segment alternation, the results of our analysis confirm the usefulness of the kind of multifactorial approach of previous work on NS data (e.g. Hinrichs & Szmrecsanyi 2007) to learner corpus data. Phonological predictors (PossorNumber and LengthDiff), syntactic predictors (the branching of both NPs), and semantic predictors (the specificity of both NPs) are intertwined in a complex network of interactions, not all of which are straightforward to interpret.

In agreement with previous NS research, we too find effects of PossorNumber for the NNSs, and the nature of PossorNumber in turn reinforces the effect of segment alternation. While PossorNumber interacts with four other predictors, the overall directionality of its effect is the same for both NS and NNS, and as expected from NS studies: plural possessors disprefer *s*-genitives. But what about the interactions? The interaction of PossorNumber and PossedBranch branching is interesting as it reveals how this combination of predictors can lead to unexpected effects: singular possessors already prefer *s*-genitives, but when the possessed is not un-/pre-, but post-modified and thus longer, then this preference for *s*-genitives becomes much stronger. In other words, all speakers strongly react to this conspiracy of phonology and, ultimately, short-before-long.

The way in which PossorSpec and PossedSpec affect NS and NNS genitives is more complex but can also be accounted for. The bottom line is this: *s*-genitives are preferred when the possessor is specific and/or the possessed is not; this tendency is particularly strong when paired with the articulatory effect we already mentioned above, possessors in the singular. This may seem perplexing given that specificity is mostly marked by definite and indefinite articles in English, and their L2 acquisition is one of the most notoriously difficult aspects of the English grammar for most learners of English (cf. Master 1997, Butler 2002, Miller 2005). However, the specificity preferences are compatible with, although not completely identical to, another widely-attested constituent ordering preference – given-before-new – and may be recast in terms of difficulty of establishing reference. Adopting only the language of a hearer-centered perspective, if given precedes new, then the earlier given material will aid comprehension in providing a reference point for incoming new information. With regard to specificity: when the possessor in an *s*-genitive is specific, then it will often, though not always,

be given, and likewise serve as a good anchor against which the referent of the non-specific possessed can be identified or construed. Thus, what is at work here is clearly a processing preference that is, on the one hand, reflected in a morpho-syntactic (determiner) choice and, on the other hand, potentially strongly amplified by congruent articulatory patterning/planning. It is worth bearing in mind again that both NL and IL exhibit this processing principle.

The next few interactions all involve another also highly processing-related factor. The straightforward overall effect is that we observe a short-before-long effect, but one that is qualified in a variety of ways. For instance, the effect of PossorNumber – the avoidance of *s*-genitives with plurals – can mute Length-Diff. In other words, when the possessor is plural, speakers dislike *s*-genitives so much that the length difference between possessor/possessed has to become quite large to "entice" them to use an *s*-genitive nonetheless. Similarly, any modification of the possessor dampens the effect of length whereas postmodification of the possessed amplifies it (unsurprisingly). Finally, the animacy of the possessor can also subdue effects of LengthDiff because length has a much stronger effect with animate possessors. Put differently, when the possessor is inanimate, it does not qualify easily for the *s*-genitive, which is biased towards animate/human possessors to denote its prototypical scenario of possession. Thus, if the semantics do not fit, even a predictor as powerful as length cannot coerce an *s*-genitive.

In combination, these interactions strongly suggest that NS and NNS choices are first and foremost influenced by processing demands; furthermore, the results suggest that at least with regard to written L2 production, intermediate/advanced learners' production data already reflect target-like processing preferences at an astonishing level of complexity.

4.2   Differences of native language and interlanguage

Let us now turn to those findings that highlighted differences between the three speaker groups. Since the focus of the present paper is largely methodological, our discussion here will be more speculative; however, we hope that our fine-grained definition of context and its application within regression methods will stimulate more detailed studies of the influence of transfer on IL.

Our results show that all speakers use the *s*-genitive less than the *of*-genitive and all speakers use the *s*-genitive most with semantic relations of possession, interpersonal relation, and time-event. However, they differ in the strength of their preferences. On the whole, we find that the Chinese learners' choices of genitives are not very different from those of English NS, which may be due to the fact that Mandarin Chinese and English share at least a few aspects that should be relevant

to Chinese learners of English: both have rather analytic morphologies and rely on word order as well as particles to convey syntactic information; genitive-type relations, which in English can be expressed with *of*, would usually be expressed with the modification marker 的 (*de*), which maybe explains why in our data the genitive choices of the Chinese learners of English are closer to the NS choices than those of the German learners. As two separate MuPDAR analyses show (see Section 4.3 below), the patterns in the English NS choices predict > 91.4% of the Chinese NNSs' choices correctly, but "only" 86.3% of the German NNSs' choices. The only exception for the overall close match of the Chinese NNS to the English NS is that, with non-specific possesseds, Chinese NNS overuse *s*-genitives (esp. when the possessed is also animate and/or the possessor is inanimate and/or specific); at present, we have no explanation for this patterning.

On the other hand, the German NNS differ considerably from English NS, but even this difference mostly boils down to the German learners doing what the other speakers do, just with a much stronger tendency to use *s*-genitives – the only exception to this pattern is the just-mentioned Chinese learners' extreme preference for *s*-genitives with non-specific possesseds. The following is some speculation on the German learners' fondness of *s*-genitives.

German NNS have to deal with a complex mapping of German options to express a possessor-possessed relationship onto English. For instance, what would be *the player's shoes* in English could be expressed with any of the options shown in Example (2), which means that both orders of possessor and possessed are attested (as in English), that there is a seemingly straightforward equivalent to the English *s*-genitive in German (Example (2a)) (and a stylistically very marked alternative with the dative (Example (2b))), but that there are also two competing constructions in which the possessed precedes the possessor: one that is related to (2a) in that it also uses the genitive case Example (2c) and one that again uses the dative but is more similar to the English *of*-genitive (Example (2d)).

(2) a. des   Spielers   Schuhe   (possessor possessed)
         the-GEN   player's-GEN   shoes-NOM

     b. dem   Spieler   seine   Schuhe   (possessor possessed)
         the-DAT   player-DAT   his-NOM   shoes-NOM

     c. die   Schuhe   des   Spielers   (possessed possessor)
         the-NOM   shoes-NOM   the-GEN   player-GEN

     d. die   Schuhe   vo(n de)m   Spieler   (possessed possessor)
         the-NOM   shoes-NOM   of (the-DAT)   player-DAT

How does German differ from English with regard to these options? Disregarding considerably dialectal variation, the construction in (2a) is strongly preferred with pronominal and proper-name possessors, but strongly dispreferred with lexical NPs, and this is further qualified by semantic relations. For instance, in English expressions of interpersonal relations, both orders of possessor and possessed are possible (e.g. *my neighbor's son* and *the son of my neighbor*) but in German, only the latter order would be widely used (e.g. *meines Nachbarns Sohn* is decidedly more marked than *der Sohn meines Nachbarn* or *der Sohn vo*(*n meine*) *m Nachbarn*). Other relations, such as time-event, again allow both orders in English (e.g. *last week's lecture* and *the lecture of last week*) but not in German, where this would only be expressed as *der Vortrag der letzten Woche* (the-NOM lecture-NOM the-GEN last-GEN week-GEN) or with a *von*+DAT construction.

While this discussion has of course only scratched the surface of the fairly complex German system, it seems as if there are several potential and interrelated reasons for why the German NNS overgeneralize: first, English allows *s*-genitives for possession (similar to German) but also for many other relations such as interpersonal relations or time-event relations (less similar to German), which may suggest to the NNS a greater degree of permissibility of English *s*-genitives than actually exists. Second, the use of the possessor-possessed order with the English *s*-genitive is short and morphologically quite simple whereas the same order in German is longer and can be difficult even for German NSs, as many queries on language forums on the web suggest. For example, a German NNS wanting to express "the budget of the university" in English has a choice: he could write either *the university's budget* or *the budget of the university*. German, in contrast, is more restrictive: the possessed-possessor order *das Budget (von) der Universität* (i.e. a genitive or a *von*-dative) is unproblematic, but the corresponding possessor-possessed order is not at all. Consequently, even German NS might just go with the compound *das Universitätsbudget*. Thus, if some semantic relation is to be expressed in English, a German NNS often has a simple and economical option in English that he does not have in German and may go with it (too often).

4.3   Concluding remarks

We hope to have shown several things. First, we hope it has become clear how much corpus analyses have to offer once learner corpus linguists take rich context and multifactoriality seriously. Let us compare our approach to the still traditional kind of approach which restricts itself to mere over- and underuse counts, maybe augmented by one additional contextual feature, maybe (also) augmented with a chi-squared test (see Section 1 above). In the case of the genitive, adopting

this former approach would mean reporting Table 2 from above, maybe adding that the distribution is significant ($\chi^2 = 29.7$, $df = 2$, $p < 0.0001$) but only a weak effect (Cramer's $V = 0.1$), and pointing out the main effect of German NNS over-using *s*-genitives. Similarly, including one contextual feature might mean discussing, say, whether the effect of PossorNumber on genitive choices differs between the NS in the upper panel and the NNS in the lower panel of Table 4.

Table 4. Distribution of genitives with singular/plural possessors (English vs. German)

English NS	*of*-genitive	*s*-genitive	Totals
PossorNumber: singular	690	96	786
PossorNumber: plural	202	8	210
Totals	892	104	996
German NNS	***of*-genitive**	***s*-genitive**	**Totals**
PossorNumber: singular	571	168	739
PossorNumber: plural	246	15	261
Totals	817	183	1000

This approach would adopt an incredibly impoverished definition of context by ignoring all other factors that we know affect genitives in English (and presumably in other languages). As we argued, while this is still the standard approach, it is severely deficient. In contrast, the regression approach adopted here offers much more detailed insight, and its findings can help us develop a more precise qualitative understanding of how native and learner English differ, and how different IL Englishes differ from each other.

Second, we also hope to have painted an interesting picture of the factors that govern NL and IL and in particular how complex the linguistic system of the intermediate learners already is. All three speaker groups exhibited complex interactions that reflect an intricate system of different factors amplifying or muting each other: articulatory factors and various processing-related factors (length, branching, and specificity) were shown to "battle it out", and possessor animacy, a factor closely related to constructional semantics, was shown to be able to over-power even length, a factor that strongly determines many other alternations. The complexity of these results and their granularity can inform qualitative interpretation and more theoretical analysis (e.g. within Granger's CIA framework), theoretical development, and interdisciplinary approaches much more than the mere cross-tabulation of the traditional kinds exemplified in Table 2/Table 4. Given the prevailing tension between qualitative and quantitative approaches, the above results are also useful to point towards the potential limits of mere qualitative analysis: it seems difficult to envisage how a qualitative analysis would be able to

come to grips with the simultaneous effects of more than ten linguistic and extra-linguistic predictors. We certainly subscribe to the idea that quantitative analysis needs qualitative/theoretical interpretation – however, qualitative analysis needs quantitative analysis just as much. In fact, we would go so far as to say that quantitative analysis is qualitative analysis made testable, falsifiable, and predictive. In this spirit, the present study hopefully also shows how the occasional gap between these two kinds of research can be bridged.

Future study, especially experimental work, would be ideal to explore some of the ways in which speakers resolve the conflicting pressures they are under. In addition and as mentioned above, larger samples would be a natural next step, as would be longitudinal data and spoken data. However, the relevance played by articulatory/phonological factors in this study suggests that the often knee-jerk call for "more spoken data" may not always be justified.

A final important follow-up would be to fully apply Gries & Deshors' (forthcoming) MuPDAR approach, which would allow us to explore in much more detail what exactly it is that makes the NNS make non-native choices. This approach would compare each genitive choice of a NNS to the genitive choice a NS would make under identical circumstances, i.e. in identical contexts. In other words, MuPDAR provides an answer to the following tripartite-part question that should be at the heart of much learner corpus research: "given the situation that the NNS is in, what would a NS do, how much (if at all) do NNS deviate from what NS do, and what are the factors that result in the NNS making non-native choices?" This approach is particularly promising since it returns regression results separately for NS and NNS and allows for an even cleaner comparison of speakers' choices than the present regression model.

Any of these follow-ups can help us develop a more detailed understanding of how IL evolves as a result of the articulatory, processing, and communicative forces learners react and adjust to; we hope this study has taken a small step towards this goal and inspires others to do the same.

References

Altenberg, B. 1982. *The Genitive vs. the of-Construction: A Study of Syntactic Variation in 17th Century English*. Malmö: CWK Gleerup.

Behaghel, O. 1909/1910. "Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern". *Indogermanische Forschungen*, 25, 110–142.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Butler, Y. G. 2002. "Second language learners' theories on the use of English articles: An analysis of the metalinguistic knowledge used by Japanese students in acquiring the English article system". *Studies in Second Language Acquisition*, 24 (3), 451–480.

Callies, M. & Szczesniak, K. 2008. "Argument realization, information status and syntactic weight – A learner-corpus study of the dative alternation". In P. Grommes & M. Walter (Eds.), *Fortgeschrittene Lernervarietäten. Korpuslinguistik und Zweit-spracherwerbsforschung*. Berlin/New York: Niemeyer, 165–188.

Couper-Kuhlen, E. 1986. *An Introduction to English Prosody*. London / Tübingen: Edward Arnold and Niemeyer.

Deshors, S. C. & Gries, St. Th. forthcoming. "A case for the multifactorial assessment of learner language: The uses of *may* and *can* in French-English interlanguage". In D. Glynn & J. Robinson (Eds.), *Polysemy and Synonymy: Corpus Methods and Applications in Cognitive Linguistics*. Amsterdam: John Benjamins.

Ellis, N. C. & O'Donnell, M. B. 2012. "Statistical construction learning: Does a Zipfian problem space ensure robust language learning?" In P. Rebuschat & J. N. Williams (Eds.), *Statistical Learning and Language Acquisition*. Berlin: de Gruyter Mouton, 265–304.

Ellis, N. C. & Ferreira-Junior, F. 2009. "Constructions and their acquisition: Islands and the distinctiveness of their occupancy". *Annual Review of Cognitive Linguistics*, 7, 187–220.

Granger, S. 1996. "From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora." In K. Aijmer, B. Altenberg & M. Johansson (Eds.), *Languages in Contrast. Text-based Cross-linguistic Studies*. Lund: Lund University Press, 37–51.

Granger, S. 2002. "A bird's eye view of learner corpus research". In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching*. Amsterdam: John Benjamins, 3–33.

Granger, S. 2004. "Computer learner corpus research: Current status and future prospects". In U. Connor & T. Upton (Eds.), *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam: Rodopi, 123–145.

Gries, St. Th. 2002. "Evidence in linguistics: Three approaches to genitives in English". In R. M. Brend, W. J. Sullivan & A. R. Lommel (Eds.), *LACUS Forum XXVIII: What Constitutes Evidence in Linguistics*. Fullerton, CA: LACUS, 17–31.

Gries, St. Th. & Wulff, S. 2005. "Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora". *Annual Review of Cognitive Linguistics*, 3, 182–200.

Gries, St. Th. & Wulff, S. 2009. "Psycholinguistic and corpus linguistic evidence for L2 constructions". *Annual Review of Cognitive Linguistics*, 7, 163–186.

Gries, St. Th. 2012. "Frequencies, probabilities, association measures in usage-/exemplar-based linguistics: Some necessary clarifications". *Studies in Language*, 36 (3), 477–510.

Gries, St. Th. 2013. *Statistics for Linguistics with R*. 2nd rev. and ext. ed. Berlin/New York: De Gruyter Mouton.

Gries, St. Th. & Deshors, S. C. forthcoming. "Using regressions to explore deviations between corpus data and a standard/target: Two suggestions". *Corpora*.

Hawkins, J. A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Hasselgård, H. & Johansson, K. A. 2011. "Learner corpora and contrastive interlanguage analysis." In In F. Meunier, S. De Cock, G. Gilquin & M. Paquot (Eds.), *A Taste for Corpora. In Honour of Sylviane Granger*. Amsterdam: John Benjamins, 33–62.

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 and Linguistics*, 11 (3), 437–474.

Kreyer, R. 2003. "Genitive and *of*-construction in modern written English: Processability and human involvement". *International Journal of Corpus Linguistics*, 8 (2), 169–207.

Leech, G. N., Francis, B. & Xu, X. 1994. "The use of computer corpora in the textual demonstrability of gradience in linguistic categories". In C. Fuchs & B. Victorri (Eds.), *Continuity in Linguistic Semantics*. Amsterdam: John Benjamins, 57–76.

Martinez-Garcia, M. T. & Wulff, S. 2012. "Not wrong, yet not quite right: Spanish ESL students' use of gerundial and infinitival complementation". *International Journal of Applied Linguistics*, 22 (2), 225–244.

Master, P. 1997. "The English article system: Acquisition, function, and pedagogy". *System*, 25 (2), 215–232.

Miller, J. L. 2005. "Most of ESL students have trouble with the articles". *International Education Journal*, 5 (5), 80–88.

Myslín, M. & Gries, St. Th. 2010. "k dizez?: A corpus study of Spanish Internet Orthography". *Literary and Linguistic Computing*, 25 (1), 85–105.

Olsen, S. 1999. "Errors and compensatory strategies: A study of grammar and vocabulary in texts written by Norwegian learners of English". *System*, 27 (2), 197–205.

Osselton, N. E. 1988. "Thematic genitives". In G. Nixon & J. Honey (Eds.), *An Historic Tongue: Studies in English Linguistics in Memory of Barbara Strang*. London: Routledge, 138–144.

Plank, F. 1985. "The interpretation and development of form alternations conditioned across word boundaries: The case of *wife's*, *wives*, and *wives'*". In R. Eaton, O. Fischer, W. F. Koopman & F. van der Leek (Eds.), *Papers from the 4th International Conference on Historical Linguistics*. Amsterdam: John Benjamins, 205–233.

Quirk, R., Greenbaum, S., Leech, G. N. & Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London/New York: Longman.

R Development Core Team. 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: www.R-project.org (accessed June 2013).

Raab-Fischer, R. 1995. "Löst der Genitiv die *of*-Phrase ab? Eine korpusgestützte Studie zum Sprachwandel im heutigen Englisch". *Zeitschrift für Anglistik und Amerikanistik*, 43 (2), 123–132.

Rohdenburg, G. 2003. "Cognitive complexity and horror aequi as factors determining the use of interrogative clause linkers in English." In G. Rohdenburg & B. Mondorf (Eds.), *Determinants of Grammatical Variation in English*. Berlin/New York: Mouton de Gruyter, 205–250.

Rosenbach, A. 2002. *Genitive Variation in English: Conceptual Factors in Synchronic and Diachronic Studies*. Berlin/New York: Mouton de Gruyter.

Rosenbach, A. 2005. "Animacy versus weight as determinants of grammatical variation in English". *Language*, 81 (3), 613–644.

Saffran, J. R. 2003. "Statistical language learning: Mechanisms and constraints". *Current Directions in Psychological Science*, 12 (4), 110–114.

Schlüter, J. 2003. "Phonological determinants of grammatical variation in English: Chomsky's worst possible case". In G. Rohdenburg & B. Mondorf (Eds.), *Determinants of Grammatical Variation in English*. Berlin/New York: Mouton de Gruyter, 69–118.

Shih, S., Grafmiller, J., Futrell, R. & Bresnan, J. in press. "Rhythm's role in genitive construction choice in spoken English".

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

Szmrecsanyi, B. 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, B. & Hinrichs, L. 2008. "Probabilistic determinants of genitive variation in spoken and written English". In T. Nevalainen, I. Taavitsainen, P. Pahta & M. Korhonen (Eds.), *The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present*. Amsterdam: John Benjamins, 291–309.

Tono, Y. 2004. "Multiple comparisons of IL, L1 and TL corpora: The case of L2 acquisition of verb subcategorization patterns by Japanese learners of English". In G. Aston, S. Bernardini & D. Stewart (Eds.), *Corpora and Language Learners*. Amsterdam: John Benjamins, 45–66.

Vennemann, T. 1988. *Preference Laws for Syllable Structure and the Explanation of Sound Change. With Special Reference to German, Germanic, Italian and Latin*. Berlin/New York: Mouton de Gruyter.

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 (2), 1–32.

Authors' addresses

Stefan Th. Gries
Department of Linguistics
University of California
Santa Barbara, CA 93106-3100
USA

stgries@linguistics.ucsb.edu

Stefanie Wulff
Linguistics Department
University of Florida
4131 Turlington Hall
Gainesville, FL 32611-5454
USA

swulff@ufl.edu