Expression of possession in English: The significance of the right edge

5

Expression of possession in English: The significance of the right edge *

KERSTI BÖRJARS[#], DAVID DENISON[#], GRZEGORZ KRAJEWSKI[#] AND ALAN SCOTT[†]

[#]The University of Manchester, [†]The University of Nottingham

In this paper we investigate one aspect of the factors that govern the choice in expression of possession in English between the possessive *'s* (POSS-S), and the *of*-construction (POSS-OF). We are particularly interested in the categorization of the POSS-S as a clitic or an affix. Since the key evidence in favour of its clitic status is the fact that it can occur at the right edge even when there is postmodification of the possessor as in *the man in the car's wallet*, this is the main focus of our paper. Weight is known to be a factor in the choice between the two constructions. However, the regression analysis we carry out on data from the spoken part of the British National Corpus shows that weight in the form of postmodification of the possessor is a factor in the choice independent of weight in general and it significantly

reduces the odds of POSS-S being chosen. We attribute this to STRUCTURAL PERSISTENCE, a notion extended from Hopper (1991). The approach we take leads us to re-evaluate the categories of affix and clitic and the historical conclusions that have tended to be drawn on the basis of the data. We argue that a model of grammar which includes information about probability looks set to account for the data most appropriately.

5.1 Introduction

The expression of possession in English has attracted a lot of attention in the literature. That which we refer to as 'possession' here is of course a very broad relation, including core possession, but also a wide range of relations such as 'author of', *Strindberg's plays*, 'depiction of/by', *Munch's self-portrait*, or 'leader of', *Cameron's coalition government* (for an extended list see Payne & Huddleston 2002: 474). Though many of our examples do not express core possession, we will follow many others in referring to the construction as the POSSESSIVE and to the two parts as POSSESSOR and POSSESSUM.

In English, as in other Germanic languages, there are a number of alternative ways of expressing possession, and much of the literature on the subject is concerned with what determines this choice. The most commonly

contrasted alternatives are the *s*-possessive, *Obama's government*, and the *of*-possessive, *the government of Obama*. It is these two that we will contrast in this paper, referring to them as POSS-S and POSS-OF, respectively. Other alternatives are the compound possessive, *the Obama government* (see Koptjevskaja-Tamm, this volume), and the double possessive, *a government of Obama's* (see Payne, this volume), but we will not have anything to say about these constructions here.

Most work on the choice of possessor expression in English has focused on the choice itself, aiming to understand what factors influence the choice. A major study by Rosenbach (2002) is concerned with how conceptual factors such as animacy, topicality and the nature of the possessive relation influence the choice, and how these factors interact. Rosenbach's study involved native speakers choosing between two ways of expressing possessives; we return to the results in section 5.2. Other studies involve detailed statistical analyses of corpora. A relatively early example is Leech, Francis & Xu (1994). They are concerned with what they describe as gradience of semantic category membership and use the possessive choice as an example. The results, which they interpret as evidence of gradience (though not morphosyntactic gradience), we would view instead in terms of the interaction of conflicting constraints, but they show that the three factors (i) semantic class of possessor, (ii) text type and (iii) semantic relation between possessor and possessum do influence the choice, and that their

relative influence can be ranked in that order. Impressive statistical corpus work by Hinrichs & Szmrecsanyi (2007) analyses similar properties, but they take a range of other factors, including structural ones, into account. They are particularly interested in how the demonstrable increase since the early 20th century in the use of POSS-S and the concomitant decrease in POSS-OF use has come about. It has been claimed in the literature to be at least partly due to an increased use of POSS-S with inanimate possessors. However, Hinrichs & Szmrecsanyi (2007: 438) show that this is not a factor, but that it 'may well reduce, at least partly, to an increasingly powerful tendency to code thematic NPs with the *s*-genitive, as well as to an epiphenomenon effect of an increasing overall lexical density of journalistic prose – a factor which would always have favored the *s*-genitive.'

It is well-known that phonological factors influence the choice of possessive construction. In particular, if the possessor ends in a sibilant, POSS-S tends not to be used, in order to avoid the clash with *-s*. Shih, Grafmiller, Futrell & Bresnan (To appear) point to a less well-known phonological effect. They carry out a thorough analysis of an annotated spoken corpus to examine the influence of rhythm on the choice between the two constructions. They find that it does play a role, but that it is a relatively small effect compared to other factors such as for instance animacy.

In the morpho-syntactic literature, on the other hand, the interest in POSS-S has specifically focussed on the theoretical status of *'s*. It is referred to in the literature as a clitic by most text books (for instance Katamba 1993, but also in the theoretical literature, most recently as a 'special clitic', Anderson 2005: 423–4), an enclitic postposition (Quirk, Greenbaum, Leech & Svartvik 1985: 328) or a suffix (Biber, Johansson, Leech, Conrad & Finegan 1999: 292).[1] A strand of theoretical work has also developed in which 'phrasal affix' and 'edge affix' are used (Zwicky 1987, Miller & Halpern 1993, Payne 2009). In early work in this direction, the two terms were used more or less synonymously (Zwicky 1987). Either term was used to indicate that even though POSS-S positioned with respect to a phrase, the attachment to its host word did not quite show the characteristics of a clitic but was more affix-like. In this literature, an edge affix was a type of phrasal affix. However, more recently phrasal affix is contrasted with edge affix (see for instance Anderson, this volume). Under this distinction, both are exponents of a feature which belongs to a phrase, but in one case the marker is introduced by a rule that operates on a phrase – postlexically – and in the other it is introduced at word level – lexically. In this approach, a special clitic is a phrasal affix (Anderson, this volume).

The traditional affix-clitic classification has been given further importance in the historical literature. If the modern *'s* is a clitic, given that it developed from an affix, its diachronic development can be seen as an

example of degrammaticalisation; a striking result given the traditional assumption of unidirectionality of grammaticalisation (see for instance Janda 2001, Joseph 2001, Newmeyer 2001 and references there). Crucial to the debate around the morpho-syntactic status of POSS-S have been examples where the marker does not appear on the head noun but at the right edge, on a noun phrase containing postmodification. At least since Jespersen (1909), this construction has been referred to as the GROUP GENITIVE, but the term PHRASAL GENITIVE is also used (e.g. Payne & Huddleston 2002). For a number of reasons, we will follow tradition here and use the term group genitive.[2] There is no doubt that possessor constructions such as *the man over there's expression* are grammatical, though most grammars point to limitations in its use, for instance that it is avoided in written language (Quirk, Greenbaum, Leech & Svartvik 1985: 1344–5) or is restricted largely to collocations (Biber, Johansson, Leech, Conrad & Finegan 1999: 298). In previous work (Denison, Scott & Börjars 2010), we looked at the actual behaviour of POSS-S and found that even in spoken language, the construction is actually avoided by native speakers.

Our interest in this article is then to explore the role of structural factors in determining the choice between POSS-S and POSS-OF, with a view to understanding whether categories such as affix and clitic are appropriate for an accurate description of POSS-S. Though previous statistical analyses have taken some structural factors into account – usually in the form of

weight – we are not aware of an account that uses the data to draw conclusions about morpho-syntactic categories.

We base our discussion on the database we constructed of all instances of possessive NPs containing POSS-S or POSS-OF in the spoken part of the British National Corpus. Our database contains 43,151 British English possessive NPs, reduced to 41,738 when descriptive genitives (*women's magazines*) are stripped out.[3] As we will see, certain other kinds of example had to be removed in the process of carrying out a detailed statistical analysis, further reducing the size of the dataset to 40,354 tokens.

We will start in section 5.2 by reviewing factors that are known to influence the choice between POSS-S and POSS-OF to see whether the results for a spoken corpus show any striking differences from those found in written data. In section 5.3, we use a regression analysis to explore the impact of (mainly) structural factors on the choice of possessive expression in our corpus. In section 5.4, we focus on the crucial examples involving a postmodified possessor; we analyse our own data in this category, and we consider accounts of their relatively infrequent use. Finally, in section 5.5, we discuss some possible historical and theoretical conclusions that can be drawn on the basis of the data.

5.2 Regression analysis of non-structural factors

A number of factors are known to play a role in the choice between POSS-S and POSS-OF: information structural, semantic, morphological, phonological and structural ones. Rosenbach (2002) studied the influence of three information structural and semantic properties: animacy, topicality and type of possessive relationship. Rosenbach's work is based on experiments in which native speakers are presented with written texts where possessor phrases have been replaced by a choice between two options; POSS-S and POSS-OF. The subjects are asked to choose between the two options. Her results confirm a number of generally held assumptions; the more animate a possessor is, the more likely it is that POSS-S is used; if the possessor is topical, POSS-S is also more likely to be chosen than POSS-OF; if the relation between the possessor and the possessum is one of core possession, inalienable possession for instance, then POSS-S is the preferred expression (for categories of types of possession, see Rosenbach 2002: 120–3). Rosenbach's methodology allows her to compare the relative impact of the factors, and she finds that when there is a conflict between them, for instance when a possessor is animate but not topical, animacy is the most influential factor, ahead of topicality, which in turn influences the choice of expression more than the nature of the possessive relation.

Altenberg (1982: 58–9) showed that in his sample of 17th-century English, regular plurals are more likely to occur as POSS-OF than irregular

plurals, the most obvious explanation for which is a tendency to avoid POSS-S with sibilants (for modern data see Hinrichs & Szmrecsanyi 2007: 452–3). More recent research has shown a more subtle interaction between number and the choice between POSS-S and POSS-OF, however. Zwicky (1987) showed that the realisation of POSS-S after a sibilant may depend on whether that sibilant is part of a plural marker or not and hence that the actual morphology may influence the behaviour of POSS-S). Speakers would say *the terrace's tiling* or *at Thomas's* (both ending in [sɪz]), but not *the cats's* [kætsɪz] *favourite places*. Hudson (this volume) suggests an explanation of this fact; in his approach, the /s/, /z/ or /ɪz/ of POSS-S actually represent the same unit as the plural /s/, /z/ or /ɪz/, the same 'morph' in his terminology, and the sequence of identical morphs leads to merger of the two. Hence the /s/ in *cats'* is the morphological merger of plural and POSS-S. Anderson (this volume), on the other hand, argues in favour of a phonological explanation for the data. It has also been observed that plurality can make the use of POSS-S less likely even when the sibilant realising the feature is not adjacent to POSS-S. Kruisinga (1931) was the first to point out that examples such as *the kings of Sweden's* are avoided, even though the plural marker and POSS-S are separated by the postmodifier. Plank (1985) shows that this may in fact be independent of whether the plural marking is regular or not; that is, POSS-OF is more likely to be chosen even when the plural marking does not involve a sibilant. Jahr-Sørheim

(1980) found that in her data, irregular plural possessors were more likely to occur with POSS-S than regular ones. However, irregular plurals were less likely than singular possessors to occur with POSS-S, and the conclusion on the basis of her data is that there is a morphological as well as a phonological effect with respect to the number of the possessor. Payne (2009) provides an account of the distribution of POSS-S with plurals when the POSS-S is not adjacent to the head noun carrying the plural marking.

In this section we consider the effect of animacy and number of possessor and possessum and we also consider topicality. We use definiteness as a proxy for topicality. There is evidence that this is an oversimplification, in particular with respect to so-called 'first-mention definites' (see for instance Fraurud 1990, Poesio & Vieira 1998). However, one could argue that some of these could appropriately be seen as having some level of topicality for our purposes, for instance when there is what we may call an associative relation with a previous referent, as in (1) and (2).

(1) We had walked for ages when we finally found *a restaurant we both liked*. As we entered, *the waiter* greeted us enthusiastically.

(2) I've just bought *a new computer*. I'm not that happy with *the keyboard* actually, so I'm thinking of returning it.

For our purposes, coding possessor and possessum for definiteness is a reasonable simplification, but for more detailed work, a more subtle form of coding would be required. We will discuss the impact of the topicality of the possessor on the choice of expression. The topicality of possessum is not included in the discussion, since if you take definiteness as a proxy for topicality, this is a knockout context: a possessum in the POSS-S construction is always definite.

In our original coding, measure possessives were included under the heading POSS-S, though with a flag to indicate their special status. Payne & Huddleston combine measure genitives with descriptive genitives under the heading ATTRIBUTIVE GENITIVES, and the following crucial observation is made: 'Because they are modifiers and not determiners, measure genitives do not confer definiteness on the NP' (2002: 470). That means that the distribution of measure possessives will be significantly different from other possessives. Furthermore, measure possessives often lack a POSS-OF alternate.[4] Accordingly we have removed measure possessives from our dataset (n = 613). We coded nine possible values of the variable animacy, but after initial analysis of the data, this was collapsed down to six: human, animal, time, place, body part and inanimate. The value 'collective human' was merged with 'human'. Number had three possible codings, allowing for underspecified or missing NPs or unclear examples. There were some unclear and unclassified examples also for animacy and topicality, and these

were removed. Pronoun heads of possessor or possessum, including determiner genitives, were not included in the database at all; the only exception was the independent genitive of double genitives like *a fan of his, no business of yours*, but they were not included in the regression analysis.

We analysed our data using logistic regression analysis. Regression analysis is a useful statistical tool. It can both model (predict) a value of one variable (response) based on values (levels) of other variables (predictors) and test the significance of the effect of an individual predictor (or an interaction of predictors) on the response variable. Logistic regression is a special case used when the response is dichotomous rather than numerical (in our case, POSS-S vs. POSS-OF): instead of the actual value of the response, it models the natural logarithm of odds in favour of one of the two values (log-odds; in our case, log-odds of POSS-S).

Turning now to the influence of animacy on the choice of possessive expression, using a simple model with possessor animacy as a predictor, we will briefly explain the method. Table 1 presents frequencies of POSS-S and POSS-OF counted from our dataset for each of the six values of possessor animacy, together with the odds of POSS-S occurring (i.e. frequency of POSS-S divided by frequency of POSS-OF). The simple data in Table 1 are manipulated by a computer program for the purposes of a regression analysis. For a variable such as animacy with six discrete values, one value is arbitrarily chosen and used as a reference against which the other five are

compared. Here it is ‘animal’ which is the reference value, and therefore the so-called intercept in Table 2 represents the logarithm of odds of POSS-S for an animal possessor (the log of 0.563 = -0.575). A ‘coefficient’ (*B* in column 2) is computed for each of the five remaining values of animacy by taking the appropriate log odds from Table 1 and shifting it by the amount of the intercept. Thus, for example, when animacy is ‘human’, the odds ratio is 0.864 (see Table 1). The log of 0.864 = -0.146 and the coefficient *B* (Table 2) = -0.146 - (-0.575) = 0.429. Similar calculations will give the remaining figures in the *B* column. A detailed understanding of this calculation is not required in order to interpret Table 2, however. A positive coefficient means increased odds compared to the reference level, i.e. a greater likelihood of POSS-S compared with animal possessors, while a negative coefficient means decreased odds, and zero means no difference from the reference level. The bigger the absolute number, the greater the difference from the reference level.

@@ Börjars et al. Table 1 about here

@@ Börjars et al. Table 2 about here

For each coefficient, Table 2 also lists its standard error (*SE*), which may be thought of as a measure of its “stability”. Keeping in mind that our database is just a random sample of all possible instances of the English

possessive, one would not expect the coefficients to be exactly the same if calculated from a different sample. On the other hand, if our conclusions concerning the relationship between possessor animacy and type of possessive are to be extended beyond this particular database, coefficients calculated from a different (but equivalent) database should not differ too much. In the hypothetical long run, coefficients calculated from the majority of random samples will fall less than one *SE* below or above the actual coefficient (*B*±*SE*), so the smaller the standard error the sounder the results.

Standard error might be interesting on its own but, most of all, it is crucial for evaluating whether the coefficient is significantly different from zero, i.e., whether there is an actual difference between a given level and the reference level of the predictor in question. For each coefficient, a statistical test (*z*) is calculated, and its corresponding *p* value tells us how likely the given results would be if the coefficient were different from zero only by chance. If the *p* value is sufficiently small – and in this paper we apply the conventional threshold of .05 – the results can be deemed too unlikely to be explained by pure chance.[5] As can be seen from Table 2, all coefficients are significantly different from zero, which means that all other levels of possessor animacy are significantly different from the reference level (animal). A human possessor significantly increases the odds of POSS-S compared to an animal possessor, whereas all other levels of animacy decrease it. Furthermore, changing the reference level and refitting the

model would prove that all other differences are significant as well. Not surprisingly, our data show that animacy is a predictor and that the categories we have used are appropriate in that they influence the outcome in a statistically significant way. However, as we shall see, comparing these results with the way in which other factors influence the results is of interest.

Turning now to the animacy of the possessum, this is on the face of it less promising as a predictor. A similar analysis with possessum animacy as a predictor (Table 3, Table 4) shows that possessum referring to a body part favours POSS-S most. In fact, in such cases POSS-S is even more likely than POSS-OF (odds greater than 1). Animals and places favour POSS-S significantly less (without, however, a significant difference between them), and the other levels of possessum animacy even less so, with time being the category for which POSS-S is least likely. We will be able to make more sense of this factor in conjunction with possessor animacy.

@@ Börjars et al. Table 3 about here

@@ Börjars et al. Table 4 about here

Analyses involving number of possessor and possessum and topicality of possessor are much simpler, since these predictors have only two levels each (Tables 5-10). It turns out that all three predictors significantly affect

the choice of possessive, with POSS-S being more likely when the possessor is singular rather than plural, definite rather than indefinite, and when the possessum is plural rather than singular.

@@ Börjars et al. Table 5 about here

@@ Börjars et al. Table 6 about here

@@ Börjars et al. Table 7 about here

@@ Börjars et al. Table 8 about here

@@ Börjars et al. Table 9 about here

@@ Börjars et al. Table 10 about here

One disadvantage of performing a separate analysis for each individual predictor is that when testing a potential effect, we do not control for the other predictors. For example, we found a strong preference for POSS-S when the possessum denotes a body part. However, we also found that human possessors significantly increase the odds of POSS-S. Assuming that human possessors and body part possessums often go together, by testing them separately we run the risk of overestimating their effects. A

merit of regression analysis is that we can put all the predictors into the same model and test the effect of each one while keeping the others constant. Note that the different coefficients in Table 11 compared to earlier tables are because Table 11 represents a quite different model.

@@ Börjars et al. Table 11 about here

Table 11 presents coefficients of the model with all three predictors included simultaneously. It should be noticed that the intercept of this model (-1.138) equals the log-odds of POSS-S when both possessor and possessum are animal and plural, and possessor is definite. Each coefficient represents a change of log-odds resulting from a corresponding change of a predictor with other predictors remaining the same. Thus, a human possessor on its own increases the log-odds by 0.668 and a body part possessum on its own increases it by 0.569, so when both possessor is human and possessum is a body part the log-odds increases by 0.668+0.569=1.237. Interestingly, in this more stringent analysis, the effect of a possessum referring to a body part turns out to be less pronounced than initially; in fact, it appears to be only marginally significant, $p < .047$ (our threshold value is .05). The effect we saw with respect to body part possessums is then down to the fact that the possessor in these cases is with overwhelming frequency highly animate. This confirms the results of Rosenbach and others that animacy of the possessor is the dominant factor in the choice of possessive expression.

A comparison of the coefficients from Table 11 with those from Table 2, Table 4, Table 6, Table 8 and Table 10 (their *p* values in particular) reveals that the animacy of the possessum is the only predictor really affected by this more rigorous approach: most differences appear weaker when other factors are controlled. The only exception to this generalisation is when the possessum is human, where there is even more evidence now that this significantly reduces the odds of POSS-S.

5.3 Regression analysis of structural factors

In comparison with the factors discussed so far, the influence of structural factors on the choice between POSS-S and POSS-OF has not been so well studied. Weight of possessor (and possessum) is included in a number of studies, but fewer authors make a distinction as to how the weight is distributed within the phrase. Jucker (1993), in a study of newspaper texts, found that postmodification on the possessor decreases the chance of it being expressed by POSS-S. Kreyer's (2003: 194) dataset – admittedly relatively small (n = 698) and from a written corpus – draws even stronger conclusions: 'Our data, then, show that the *of*-construction ('*the* N2 *of* N1') is compulsory with postmodified modifiers' ['possessors' in our terminology].

In order to investigate whether the structural complexity of the possessor (or the possessum) is a factor in the choice of possessive expression, one would need to establish whether two possessors with the same number of words, but with different structural complexity, behave differently. However, measuring degree of structural complexity is a controversial matter and depends on theoretical assumptions (though see e.g. Hawkins 1994). For instance, different ways of determining structural complexity could provide different answers to the question of whether a possessor with a modified adjective such as *the incredibly stupid dog* is more or less complex than one with two unmodified adjectives, *the stupid dirty dog*. Similarly, how would one compare a possessor post-modified by two PPs with one containing one more complex PP postmodifier, say *the student of chemistry from Bristol* as opposed to *the student from the mayor's estate*? In order to avoid these issues, we will use length as a proxy for structural complexity.[6] Our database is coded for length both in terms of syllables and in terms of words. However, the two length variables were closely correlated with each other and produced very similar results, so that in what follows we count length in words only.[7] For both possessors and possessums, we counted every word, so that determiners are included in the count, except that in POSS-S constructions the possessor phrase – which functions as the determiner of the possessum – was not counted in the possessum, while the *of* in POSS-OF was not counted at all.

One reason that structural complexity or length is assumed to be a factor influencing the choice between POSS-S and POSS-OF is the general tendency for language to prefer long constituents at the end of a phrase. As illustrated in (3), the order between the possessor and the possessum varies between the two constructions.

(3) a. a football player's performance

POSS-S: POSSESSOR < POSSESSUM

b. the performance of a football player

POSS-OF: POSSESSUM < POSSESSOR

On the principle of end weight, we would expect increasing length of the possessum to favour POSS-S, since in this construction, the possessum follows the possessor. Longer possessors, on the other hand, would favour POSS-OF. That is indeed what even a simple inspection of the data reveals, with one noticeable exception where the value for possessum length = 1 (see Figure 1).

@@ Börjars et al. Figure 1 about here

Our explanation for the anomaly with possessum length=1 is as follows. In a POSS-OF construction, the possessum is represented by a standard noun phrase, and a one-word non-pronominal noun phrase in

English is only possible if the head is a proper noun, a non-count common noun or a plural – otherwise a determiner is required. Indeed, 95% of two-word possessums in the POSS-OF construction consist of D + N. In a POSS-OF constructions, we would then expect one-word possessums only with these sub-classes of nouns:

(4) especially for *mothers* of young children. D8Y 422[8]

(5) ask the assembly to authorize *publication* of the reworded statement F85 046

(6) *…., especially for *mother* of young children

In the POSS-S construction on the other hand, the possessor phrase functions as determiner so that the possessum itself is expressed as a nominal constituent of a lower level than noun phrase. This means that possessums which are impossible in the POSS-OF construction, such as singular count nouns, can be used grammatically in the POSS-S construction.

(7) we lived in my aunt's *house* / **house* of my aunt D90 109

(8) waters passing through the earth's *crust* / **crust* of the earth at hydrothermal vents F8E 094

In that case, then, there is no real choice between POSS-S and POSS-OF: it is a knockout context, hence the anomalous result for length 1 in Figure 1.

Because of the anomaly of one-word possessums, we excluded all such datapoints in our further examination of the data (n = 8994). We also excluded datapoints for which possessum length was greater than 15 and those for which possessor length was greater than 10 since they were already invariably POSS-S or POSS-OF respectively (n = 384). All datapoints for which possessum or possessor length was 0 (n = 566, n = 1, respectively) were also removed: absence of expressed possessum is relatively common in elliptical constructions, as in (9). The only example coded as lacking a possessor is (10), which involves co-ordination of two possessors with a possessum which could be described as having undergone right-node raising.

(9) its [*sic*] about to me hidden desire of some sort, either *the artist's* [∅] or hers F71 198

(10) that would be *the terminating point of* [∅] or the end of the procedure as such GYV 332

We then modelled the type of possessive using logistic regression, with the length in words of possessum and the length in words of possessor as predictors.

@@ Börjars et al. Table 12 about here

Our predictors in this model were numerical rather than categorical, so there is only one coefficient per predictor and the interpretation of the coefficients is as follows: if the length of possessum increases by one word, the log-odds of POSS-S increases by 2.497, and if the length of possessor increases by one word, the log-odds of POSS-S decreases by 1.578.[9] The results of the analysis were as predicted: the longer the possessum, the greater the odds of POSS-S, with the opposite effect for the length of possessor. However, we also included the interaction between these two predictors in the model and, as can be seen in Table 12, the interaction is significant as well. It means that the actual effect of one predictor depends

on the value of the other: as the length of possessor increases, the *effect* of the length of possessum increases as well (e.g., if the length of possessor is 1, the actual effect of possessum length is 2.497 + 1*0.352 = 2.849 and if the length of possessor is 2, it reaches 2.497 + 2*0.352 = 3.201). Analogically, as the length of possessum increases, the (negative) effect of the length of possessor decreases (remaining, however, significant for all datapoints). What this means is that of the two length variables, possessum length is the more potent both in itself and in combination with possessor length; as possessum length grows, it reduces the effect of possessor length. We have no explanation for this interaction between possessor and possessum length, but we note that Wolk, Bresnan, Rosenbach & Szmrecsanyi (2011: 19), in their study of choice in possessive constructions, find 'fairly complex, nonlinear relationships' between constituent lengths and genitive choice. They assume that these effects may be due to the fact that they modelled possessum and possessor weight separately and that this 'may not do full justice to the possibly very complex interplay between relative and absolute weights'.

We have referred to the effect here as one of length or weight, but length and weight are known to interact with information structural notions such as given and new. New elements tend to be longer and given elements shorter, with pronouns as the limiting instance (or agreement marking or zero in languages that allow arguments not to be realised syntactically).

There are different views on whether weight is an independent factor in the choice between POSS-S and POSS-OF, or whether its apparent effect is actually due to information structural considerations, or possibly the other way around. Consider examples (11) to (13).

(11) [An anthropologist describes how a professional ascetic sings a "burlesque dirge" in front of a house.] *It was the house of a very much alive moneylender*.

(http://www.battleofideas.org.uk/index.php/2010/battles/3613/)

(12) The Square tavern: it is a construction dating from the 18th century. It was *the house of a Spanish wealthy farmer*

(http://www.hoteldelrijo.com/hotel-del-rijo-main-attractions.htm)

(13) Unlike Roy Kaspar's little black Mercedes with its wood-grain panel and tape deck and leather seats and convertible top, Mr. Bloodworth's dusty gray Chevrolet was all business. […] It was *the car of a dedicated, working gumshoe*.

Dick Lochte, *Sleeping dog* (2000:23, www.poisonedpenpress.com /30-page-pdfs/9781890208516-30pp.pdf)

In these examples we have a human and hence highly animate possessor, but in each case POSS-OF is used in preference to the expected POSS-S construction. The possessor in each case introduces a new referent and is fairly heavy. There are then two potential explanations for the use of POSS-

OF: it could be in order to place the new information after the given (*house* and *car*), or it could be to place a heavy constituent at the end of the phrase. In previous work (Börjars, Denison & Scott 2009), we concluded on the basis of a *Goldvarb* analysis of our database that weight is a more influential factor than information structure in determining the choice between POSS-S and POSS-OF. This fits in with other, more detailed studies. Hinrichs & Szmrecsanyi (2007: 460–1), for instance, find that 'givenness of the possessor head' is not significant in their regression analysis. However, if end weight-related factors and factors related to thematicity of possessor are removed, then givenness does become significant. They take this to indicate that information structural factors are epiphenomenal to factors such as weight. This chimes with Hawkins (1994: 238–42), who argues that information structural effects are actually epiphenomenal to end weight. Arnold, Wasow, Losongco & Ginstrom (2000), on the basis of a study of the complements of verbs, argue that both weight and discourse status have an effect on word order. Interestingly, they show that this is not always driven by hearer-based processing constraints as often claimed but is also motivated by considerations related to the speaker, such as planning and production.

5.4 Postmodification of the possessor

5.4.1 *Distribution in our data*

Since our main research questions relate to the appropriateness of using morpho-syntactic categories such as affix and clitic to describe the behaviour of POSS-S, we will be particularly interested in factors that have been used to argue for the status of *'s* with respect to these categories. The crucial constructions here are those where POSS-S does not appear on the head of the possessor, i.e. those where the possessor contains a postmodifier. These are factors that are important also for our purposes.

In order to explore the issues relating to clitic status, new variables were brought into play, therefore: length of premodifying sequence in words and length of postmodifying sequence in words, in both cases applying to the possessor NP. We wanted to check whether in fact it made a difference *where* the length of the possessor was located, before or after the head. Since for 26% of datapoints there was no premodification and for as many as 93% there was no postmodification (length=0), the first approach was to treat each as a binary variable: absence vs. presence of pre-/ postmodification. Table 13 presents coefficients of the model with these two variables and the interaction between them as predictors. As can be seen from Table 13 and Figure 2 (with log-odds transformed into probabilities for easier interpretation), the presence of premodification has a much weaker effect than the presence of postmodification on the odds of POSS-S. Given the significant interaction between the two, the effect of the presence of premodification disappears completely when postmodification is present

(-0.848+1.301=0.453, which is non-significant), but the effect of the latter survives (though somewhat weaker) when premodification is present (-2.477+1.301=-1.176).

@@ Börjars et al. Table 13 about here

@@ Börjars et al. Figure 2 about here

As a second step, we wanted to test whether the effect of postmodification remains stronger when we take into account the actual length, that is to say whether a three-word postmodification influences the odds more than a two-word one. However, the data turned out to be too sparse to allow any conclusive analyses: only for two values of premodification length, Table 14, and for three values of postmodification length, Table 15, was there some variation in the type of possessive; even restoring the length=0 datapoints, which we had excluded because they were so disproportionally frequent, would not help much. The phenomenon under investigation is so rare that a subtler investigation would require a much larger corpus. An alternative approach to teasing out the more subtle distinctions would be an experimental approach involving some form of gradient grammaticality judgement, a magnitude estimation test for instance

(see for instance Bard, Robertson & Sorace 1996) or so-called 100-split task (see Bresnan 2007).

@@ Börjars et al. Table 14 about here

@@ Börjars et al. Table 15 about here

In this section we have shown so far that the odds of POSS-S increases with the length of possessum and decreases with the length of possessor. Furthermore, within the latter, the presence of postmodification decreases the odds more than the presence of premodification. In previous work (Denison, Scott & Börjars 2010), we have discussed an additional option for expressing possessives in English, the so-called SPLIT POSSESSIVE (or SPLIT-POSS) exemplified in (14) and (15).[10]

(14) you must put something in *a person*'s mouth *that has epilepsy*

F8C 105

(15) We don't know *the gentleman*'s name *with the tape recorder*.

FM7 8

Quirk et al. (1985: 1282) claim that such constructions are ungrammatical. Payne & Huddleston (2002: 479 fn 65) provide the following example with possessum ellipsis, *I could feel the hair stand up on the back of my neck like*

a dog's that's going to get into a fight, and comment 'Examples of the kind cited are not acceptable and frequent enough to qualify as grammatical.' We will return to a discussion of graded grammaticality and frequency in section 5.5. In Denison, Scott & Börjars (2010) we argue against any attempt at ruling these out of bounds as ungrammatical, part of a more general process of extraposition or production errors. Since the split possessive can be seen as a strategy to avoid standard POSS-S constructions when the possessor contains postmodification, it is relevant to consider whether there is evidence that a split possessive is more likely the longer the postmodification. The relevant data from our database are found in Table 16. It shows a clear relationship between the presence of a split and the length of postmodification.

@@ Börjars et al. Table 16 about here

5.4.2 *Previous accounts*

The fact that the weight of the possessor and the possessum may influence the choice between POSS-S and POSS-OF has been noted in the literature as referenced above, as has the fact that the explanation for this may be sought in a syntactic end weight principle or that it may be due to organisational principles associated with information structure which prefer given information – usually expressed by a short, light constituent – to precede

new information – usually represented by longer phrases. The outcome of our study, using definiteness as an imperfect proxy for topicality, chimes with previous studies of possessives: that weight is an independent factor and has more of an effect than information structural status.

Jespersen already made the point that heavy possessors make POSS-S a less likely choice. Interestingly, all four examples he uses to make this point include postmodification (1987 [1933]: 105), which we have shown is a factor independently of weight.[11] A number of accounts of and explanations for the avoidance of POSS-S with postmodified possessors have been offered in the literature. One is that if the postmodification ends in a noun, there could be ambiguity as to which noun should be interpreted as the possessor. Quirk et al. (1985: 1345) give the following two examples and state that whereas type (16) is used (though mainly in spoken language), (17) is avoided since there is ambiguity with respect to the possessor of the ears.

(16) the man with the car's ears

(17) the man with the cat's ears

Kreyer (2003) offers a processing-based explanation for the avoidance of POSS-S with possessors with postmodification. He posits the PROXIMITY PRINCIPLE (2003: 179), which states that 'related constituents should be in the proximity of one another'. By the proximity principle, any modification

should be as close to its head as possible. Within a possessive construction, this can give rise to conflicting pressures. In Kreyer's analysis, the possessor modifies the possessum and hence the head of the possessor should be in proximity to the head of the possessum. Similarly, any modifier of the possessor or the possessum should stand in proximity to its head. This then means that any postmodification of the possessor in the POSS-S construction violates the proximity principle in that it increases the distance between the head of the possessor and the head of the possessum. Similarly, postmodification of the possessum in the POSS-OF construction will separate the head of the possessum from the head of the possessor, as illustrated in (18) and (19). With respect to postmodified possessors, this leads to a preference for (20a) over (20b).

(18) POSS-S: POSSESSOR HEAD +*POSTMOD*'S POSSESSUM

(19) POSS-OF:POSSESSUM HEAD +*POSTMOD* *of* POSSESSOR

(20) a. the *car* [of the *man* that is talking to you]

b. [the *man* that is talking to you's] *car*

Kreyer does of course recognize that constructions such as (20b) are used, and he sees this as possessor-internal proximity being prioritized above proximity between possessor and possessum. There is no explanation for

why the proximity principle is sometimes contravened in this way, indeed possessor-internal proximity seems to be identical in the two examples in (20). It is not clear why the grammar (or the speaker) should choose (20b), since both proximity requirements are satisfied in (20a); the possessor and the possessum are closer to each other in (20a) than in (20b) and the possessor postmodification is as close to the head it modifies in (20a) as in (20b).

The principle may also lead us to expect postmodification of the possessum to strongly favour POSS-S, but as we saw, this effect is not as strong as the effect of postmodification on the possessor. This is also what Kreyer finds. With respect to premodification, the proximity principle does not seem to have much influence. Kreyer considers premodification in terms of difference in the amount of premodification between the possessor and possessum, hence it is essentially a matter of comparative weight.[12] Like us, he finds that premodification has a less strong influence on the choice between POSS-S and POSS-OF; only when the possessor is heavier than the possessum by 2 or more is there a significant effect. Given the proximity principle, this seems to us a surprising result. After all, premodification also has the effect of reducing the proximity of the possessor and possessum, as in (21).

(21) a. POSS-S: POSSESSOR's *PREMOD*+POSSESSUM HEAD

b. POSS-OF:POSSESSUM *of* *PREMOD*+POSSESSOR HEAD

The proximity principle would then lead us to expect premodified possessors to favour POSS-S as strongly as postmodified possessors prefer POSS-OF. This, however, is not the case, either in Kreyer's data or in ours. In (22) there are perfectly natural examples of an OF construction with a premodified possessor which is human.

(22) a. the general *knowledge* [of the sixth form science *teacher*]

KRH 4047

b. the *productivity* [of the independent middle *peasant*]

KM6 1055

(23) a. [the sixth form science *teacher*'s] general *knowledge*

b. [the independent middle *peasant*'s] *productivity*

The proximity principle would lead us to expect the examples in (23) to be strongly preferred to those in (22), but there is no evidence that this is indeed the case. Indeed, Kreyer (2003: 201) states that '**With regard to postmodification** the data showed that a proximity principle is at work' [our emphasis]. In order for the proximity principle to have any explanatory

power as a principle of processing ease, we then need to know why it does not apply in equal measure to premodification.

5.5 Conclusions

One of the theoretical concerns we have in this paper is with the categories affix and clitic, in particular the way in which they are applied to POSS-S and the repercussions this has for assumptions about its diachronic development. Börjars (2003) argued for the Swedish POSS-S, which behaves in many ways like its English equivalent, that a simple dichotomy between affix and clitic does not allow a proper description of either its current properties or its historical development. We would argue on the basis of the evidence presented here and elsewhere (Denison, Scott & Börjars 2010, Börjars & Vincent 2011) that the same can be said for the English POSS-S.

The six criteria posited by Zwicky & Pullum (1983: 503–504) are generally used to distinguish between affixes and clitics. The criteria as applied in the subsequent literature have frequently implied that there is a clear dichotomy between two distinct categories. In synchronic descriptions, elements that show inconsistent behaviour with respect to these criteria have been deemed to be untypical affixes or untypical clitics. As mentioned in section 5.1, a more subtle distinction has been made in the literature; elements which occur on the edge, i.e. position like clitics, can attach either postlexically or lexically. This can be described as an analysis involving

variation in two dimensions (and it is an interesting issue whether there are also “postlexical head affixes”). We would extend this line of reasoning and see the clitic-affix distinction as a multi-dimensional distinction, with affix and clitic representing relatively common clusterings of properties, but with every expectation that some elements may be characterised by a set of properties which distinguishes them from both categories. When looked at in this light, the Present Day English POSS-S is neither an affix nor a clitic, in the sense that the properties that characterise it do not cluster neatly at either the clitic or the affix end of the spectrum. This in turn means that claims of degrammaticalisation must also be considered in each dimension independently.

Arguments for the clitic status of POSS-S frequently involves a comparison with earlier stages of the language (see for instance Campbell 2001, Janda 2001, Norde 2001 and references there). The logic runs that the element in Old English from which POSS-S has developed was clearly an affix, or an inflection, and the Present Day English POSS-S behaves differently in a number of respects; if a dichotomy is assumed, this means that it must now be a clitic. On the assumption that there is multi-dimensional variation, on the other hand, each aspect of its behaviour needs to be contrasted. The most common characteristics referred to when OE (*e*)*s* and PDE POSS-S are contrasted are (i) (*e*)*s* was one exponent in a paradigm, *’s* has only one form; (ii) GEN in Old English was an agreement feature, *’s* is

marked once only; (iii) (*e*)*s* occurred on the head, *'s* occurs on the right edge of the phrase. Vincent & Börjars (2010) in the context of the directionality of grammaticalisation have argued that none of these changes can be described as degrammaticalisation. The change in (i) could in fact be seen as an example of grammaticalisation in the light of Lehmann's (1995) criterion of paradigmatisation. Vincent & Börjars 2010 also argue that (ii), the loss of agreement marking, cannot be seen as evidence for or against unidirectionality. Neither (i) or (ii) impinges on the issue of whether the description of POSS-S as a clitic is appropriate; this rests on (iii), and given the assumption that POSS-S is lexically attached, this is now just a matter of placement with respect to a phrase. Furthermore, the evidence we have presented in this paper shows that even the matter of placement is not as clear-cut an issue as has generally been assumed. It is evident that speakers avoid realising POSS-S on the right edge when the right edge is not also the head. This goes beyond any general process of extraposition. In ICE-GB, another English corpus, but one which is POS-tagged and parsed and hence easily searchable on structural criteria, 14.8% of all noun phrases have postmodification, whereas in our corpus, the proportion of possessors in the POSS-S construction with postmodification is about 2.2%, or if the examples consisting of a head and *else* are discounted, the proportion is 1% (see also Denison, Scott & Börjars 2010: 555–6). It is then clear that there is a special

interaction between postmodification and possessors in the POSS-S construction.

In section 5.4.2 above, we have looked at Kreyer's (2003) processing-based account in terms of a principle of proximity. Though we have no doubt that proximity will play a role in shaping grammar, we rejected Kreyer's explanation as lacking generality. We shall argue instead for an extension of Hopper's (1991: 28–30) principle of PERSISTENCE.[13] Hopper argues with respect to semantic change that a grammaticalised element may retain some evidence of its original lexical meaning. We would argue that in the same way, some earlier structural properties can persist in a grammaticalised element (see Breban 2009, van Bogaert 2011 for other examples of structural extension of persistence). The idea is that the head placement of the Old English genitive (*e*)*s* persists to some extent in POSS-S, even though it has developed into an edge-based once-only marking element. This means then that there are two constraints on the placement of POSS-S, which can only both be satisfied when the possessor noun phrase is head-final. This explains the low rate of properly postmodified possessors in the POSS-S construction. When a speaker is faced with a possessor which is not head-final, an alternative strategy tends to be followed: either the POSS-OF construction is used even if other factors would militate against its use, or the possessor noun phrase is altered in such a way that it becomes head-final; this is where the split construction is used.

Given this conclusion about the placement of POSS-S, what do we say about its status as an affix or a clitic? We assume that the properties which are generally used to characterise the two categories are relatively independent properties, where we think of one end of the spectrum as the 'clitic end' and the other as the 'affix end'. For one thing, form and function need to be distinguished. For instance, the fact that an element can be argued to be part of a case system does not say anything about whether it should be described as an affix or a clitic; Korean, for instance, has an element that is best described as a case marker functionally, but it is found on the right edge and its attachment is syntactic in nature (see for instance Blake 1994: 11–12). The definiteness marker occurs only once in both Danish and Bulgarian, but in Danish it always occurs on the head, whereas in Bulgarian it occurs in the second position in the phrase. Some properties are probably not so independent; edge placement is probably generally correlated with low degree of boundedness or integration. If an element always occurs on the head, it is likely that some morphological integration occurs over time, whereas if it is edge-placed, it will attach phonologically to whatever element happens to be adjacent in that particular phrase regardless of category. This means it will have a different phonological host each time, and morphological integration is less likely to occur. However, this would only be a tendency; POSS-S, even when edge-placed, shows some degree of integration, hence the lexical attachment posited in the literature.

There are then some elements where each characteristic points to the "clitic end" of the spectrum and some where all evidence points towards the "affix end". However, we suspect that there are as many where the evidence is ambiguous, and rather than end up with descriptions such as 'affix-like clitics' or 'clitic-like affixes' it is better to recognise that affix and clitic are idealised, "pure" categories and that the behaviour of most bound elements will be messier than that. Compare this with Hudson's analysis (this volume), in which it is assumed that *'s* can be both a clitic and an affix, but in any one token it is either one or the other; a clitic in the case of group genitives and an affix in other environments (see also Miller & Halpern 1993). This contrasts with our view, in which the head and the edge criteria apply to all instances of POSS-S. The conflict that results gives rise to the alternative avoidance strategies, such as the use of POSS-OF or SPLIT-POSS. An explanation for the avoidance of the group genitive is not so obvious in Hudson's account.

A consequence of what we have said here is that we are moving towards a view of grammaticality as a non-categorical property. It is not our intention here to enter too deeply into a discussion of grammar, usage and the role of probability in the mental grammar (see for instance Newmeyer 2003 and responses, such as Clark 2005 and Gahl & Garnsey 2006), but some comments are in order. The issue at hand is whether the mental grammar has only categorical knowledge of sentences and phrases as being

either grammatical or ungrammatical, with usage being a separate, extra-grammatical component, or whether the mental grammar contains some information about the probability of a phrase occurring in a particular environment. In the former approach a phrase can be completely grammatical, but hardly ever used. In this model the low usage would not be accounted for in the grammar but would be a separate usage phenomenon. Similarly, a construction could be deemed ungrammatical but still occur, even though it is very rare (see Payne & Huddleston's (2002: 479 fn 65) comment on split possessives referred to in section 5.3 above). Our results make us inclined towards the alternative view, also held by Hudson (this volume) and O'Connor, Maling & Skarabela (this volume). In models that include information about probability, such as stochastic models of the mental grammar, the native speaker is assumed to have knowledge of statistical preferences as part of their grammar (see for instance Boersma & Hayes 2001). The influence of constituent weight on word order is an example frequently cited in the literature taking this view of grammar. In English and many other languages, the preferred word order may be influenced by the length of the constituents involved. We have seen effects of this in the possessive alternations we have studied in this paper, but it also affects clausal word order. In a stochastic model, these preferences are assumed to form part of the native speaker's knowledge of grammar. One argument in favour of it actually being part of grammar is the fact that the

same weight constraint is categorical in some languages but represents preferences in others. There is a wealth of literature on this: an early example is Givón (1979), some more recent ones are Bresnan, Dingare & Manning (2001), Bresnan & Nikitina (2009), Bresnan & Ford (2010) and Wasow (2002). O'Connor, Maling & Skarabela (this volume) show that tendencies relating to the distribution of POSS-S in English parallel categorical grammaticality in other languages.

This literature shows that graded grammaticality judgements, or maybe better statistical grammaticality judgements, frequently result from an interaction between competing constraints, for instance a conflict between a default position for, say, an object and the desire to have heavy constituents at the end of a clause. With respect to the possessive we have seen that the animacy of the possessor and the weight of the possessor may point towards different choices of possessive construction. The competing constraints may also result from historical change. Clark (2004) considers changes to word order and the behaviour of types of subject in early English which at some stage showed the variability typical of competing constraints. It is well-recognised that a historical change will involve a stage of variability, that is, a change from the use of construction A to the use of construction B will involve a period in which both constructions are used. Clark's approach is to assume that a speaker's mental grammar contains a ranking of constraints. The strength of a constraint is modified in the light of

the speaker using or hearing a construction. Historical change is then represented as the gradual adjustment over time to the strength of the constraints. We would argue that this would be an appropriate analysis of the uncertainty around the use of possessive construction; there are two constraints relating to the placement of POSS-S; ATTACH TO HEAD and ATTACH AT RIGHT EDGE. When the possessor contains postmodification, both constraints cannot be satisfied and the speaker may employ an alternative strategy, such as the use of POSS-OF even when other factors would militate against this choice. Speakers will vary, however, with respect to the relative strength of the two constraints, so that those with weaker ATTACH TO HEAD and stronger ATTACH AT RIGHT EDGE may use the group genitive more liberally. Given what we have just said about Clark's analysis, this may then lead to the assumption that competing constraints will always be evidence of a language in flux and the prediction that one of the two constraints, usually the originally stronger one, will weaken to the point where its effect can hardly be noticed. This would seem to be the wrong conclusion, since there is evidence of the effect of ATTACH TO RIGHT EDGE from the second half of the Middle English period, but it has still not taken over from ATTACH TO HEAD, as witness the reluctance to use POSS-S when the possessor is modified. However, in the model developed by Clark 'there is no pressure for grammatical systems that exhibit multiple options to disappear over time.' (2004: 257). We would argue then that the tendency

to avoid the so-called group genitive results from a stable close competition between the two constraints that dictate the placement of *’s*.

TABLE 1 FREQUENCIES OF POSSESSIVES FOR DIFFERENT LEVELS OF POSSESSOR ANIMACY

Type of possessive	Possessor animacy					
	human	animal	time	place	body part	inanimate
POSS-OF	7907	199	1989	4265	362	16893
POSS-S	6832	112	729	878	20	156
odds of POSS-S	0.864	0.563	0.367	0.206	0.055	0.009

TABLE 2 COEFFICIENTS OF THE MODEL WITH POSSESSOR ANIMACY AS A PREDICTOR

	B	*SE*	*z*	*p*
(intercept)	-0.575	0.118	-4.870	<.001
possessor animacy = body part	-2.321	0.258	-8.990	<.001
possessor animacy = human	0.429	0.119	3.590	<.001
possessor animacy = inanimate	-4.110	0.143	-28.760	<.001
possessor animacy = place	-1.006	0.124	-8.120	<.001
possessor animacy = time	-0.429	0.126	-3.410	<.001

TABLE 3 FREQUENCIES OF POSSESSIVES FOR DIFFERENT LEVELS OF POSSESSUM ANIMACY

Type of possessive	Possessum animacy					
	body part	animal	place	human	inanimate	time
POSS-OF	296	50	1418	4097	25004	757
POSS-S	358	26	729	1252	6245	117
odds of POSS-S	1.209	0.520	0.514	0.306	0.250	0.155

TABLE 4 COEFFICIENTS OF THE MODEL WITH POSSESSUM ANIMACY AS A PREDICTOR

	B	*SE*	*z*	*p*
(intercept)	-0.654	0.242	-2.700	<.008
possessum animacy = body part	0.844	0.254	3.320	<.002
possessum animacy = human	-0.532	0.244	-2.180	<.03
possessum animacy = inanimate	-0.733	0.242	-3.030	<.004
possessum animacy = place	-0.011	0.246	-0.050	<.964
possessum animacy = time	-1.213	0.261	-4.640	<.001

TABLE 5 FREQUENCIES OF POSSESSIVES WITH RESPECT TO THE NUMBER OF POSSESSOR

Type of possessive	Possessor number	
	singular	plural
POSS-OF	24819	6798
POSS-S	7292	1310
odds of POSS-S	0.294	0.193

TABLE 6 COEFFICIENTS OF THE MODEL WITH POSSESSOR NUMBER AS A PREDICTOR

	B	*SE*	*z*	*p*
(intercept)	-1.647	0.030	-54.570	<.001
possessor number = singular	0.422	0.033	12.790	<.001

TABLE 7 FREQUENCIES OF POSSESSIVES WITH RESPECT TO THE NUMBER OF POSSESSUM

Type of possessive	Possessor number	
	singular	plural
POSS-OF	25129	6489
POSS-S	6604	2127
odds of POSS-S	0.263	0.328

TABLE 8 COEFFICIENTS OF THE MODEL WITH POSSESSUM NUMBER AS A PREDICTOR

	B	*SE*	*z*	*p*
(intercept)	-1.115	0.025	-44.640	<.001
possessum number = singular	-0.221	0.029	-7.740	<.001

TABLE 9 FREQUENCIES OF POSSESSIVES WITH RESPECT TO THE TOPICALITY OF POSSESSOR

Type of possessive	Possessor topicality	
	definite	indefinite
POSS-OF	21393	10228
POSS-S	7089	1638
odds of POSS-S	0.331	0.160

TABLE 10 COEFFICIENTS OF THE MODEL WITH POSSESSOR TOPICALITY AS A PREDICTOR

	B	*SE*	*z*	*p*
(intercept)	-1.105	0.014	-80.600	<.001
possessor topicality = indefinite	-0.727	0.030	-24.290	<.001

TABLE 11 COEFFICIENTS OF THE MODEL WITH TOPICALITY, ANIMACY AND NUMBER (OF BOTH POSSESSOR AND POSSESSUM)

	B	*SE*	*z*	*p*
(intercept)	-1.138	0.298	-3.820	<.001
possessor animacy = body part	-2.392	0.263	-9.090	<.001
possessor animacy = human	0.668	0.130	5.150	<.001
possessor animacy = inanimate	-4.032	0.152	-26.500	<.001
possessor animacy = place	-1.010	0.136	-7.440	<.001
possessor animacy = time	-0.396	0.137	-2.890	<.005
possessum animacy = body part	0.569	0.286	1.990	<.047
possessum animacy = human	-0.855	0.274	-3.120	<.003
possessum animacy = inanimate	-0.066	0.273	-0.240	<.808
possessum animacy = place	0.240	0.277	0.870	<.388
possessum animacy = time	-0.727	0.294	-2.470	<.014
possessor topicality = indefinite	-0.194	0.039	-5.000	<.001
possessor number = singular	0.877	0.041	21.470	<.001
possessum number = singular	-0.126	0.034	-3.670	<.001

FIGURE 1 OBSERVED PROPORTION OF POSS-S AS A FUNCTION OF LENGTH OF (A) POSSESSUM, (B) POSSESSOR

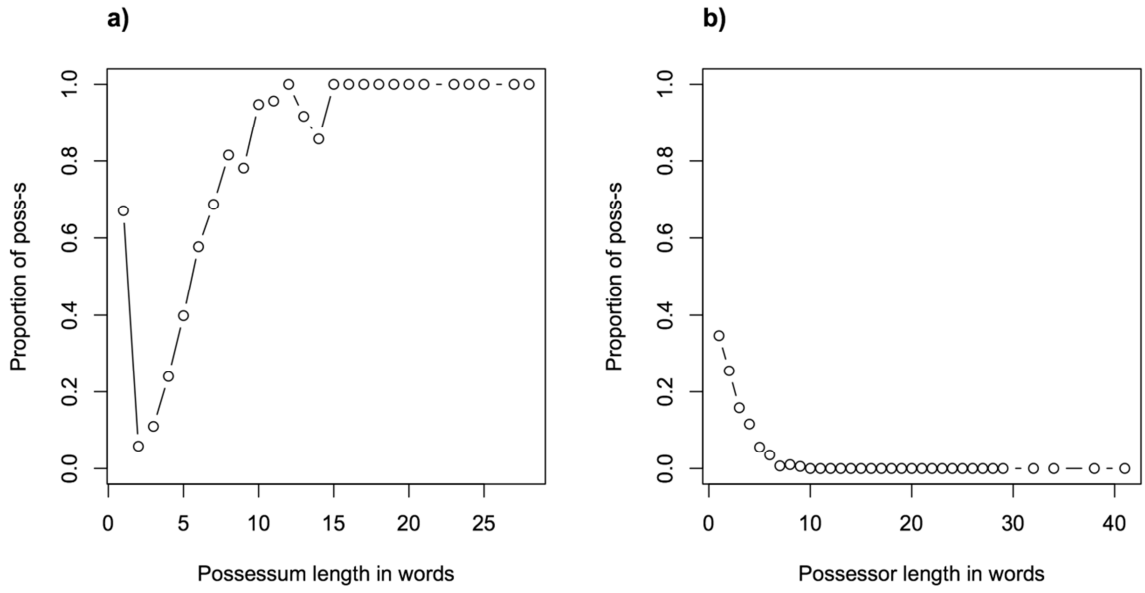

TABLE 12 COEFFICIENTS OF THE MODEL WITH LENGTH OF POSSESSUM AND POSSESSOR

	B	*SE*	*z*	*p*
(Intercept)	-3.842	0.091	-41.97	<.001
Possessum length	2.497	0.090	27.65	<.001
Possessor length	-1.578	0.127	-12.41	<.001
Possessum length x Possessor length	0.352	0.117	3.01	<.003

TABLE 13 COEFFICIENTS OF THE MODEL WITH PRESENCE OF PREMODIFICATION AND POSTMODIFICATION OF POSSESSOR

	B	*SE*	$\tilde{z}$	*p*
(Intercept)	-1.647	0.031	-52.59	<.001
Premodification present	-0.848	0.041	-20.87	<.001
Postmodification present	-2.477	0.320	-7.73	<.001
Premodification present x Postmodification present	1.301	0.359	3.62	<.001

FIGURE 2 PREDICTED PROBABILITY OF POSS-S AS A FUNCTION OF PRESENCE OF PREMODIFICATION AND POSTMODIFICATION

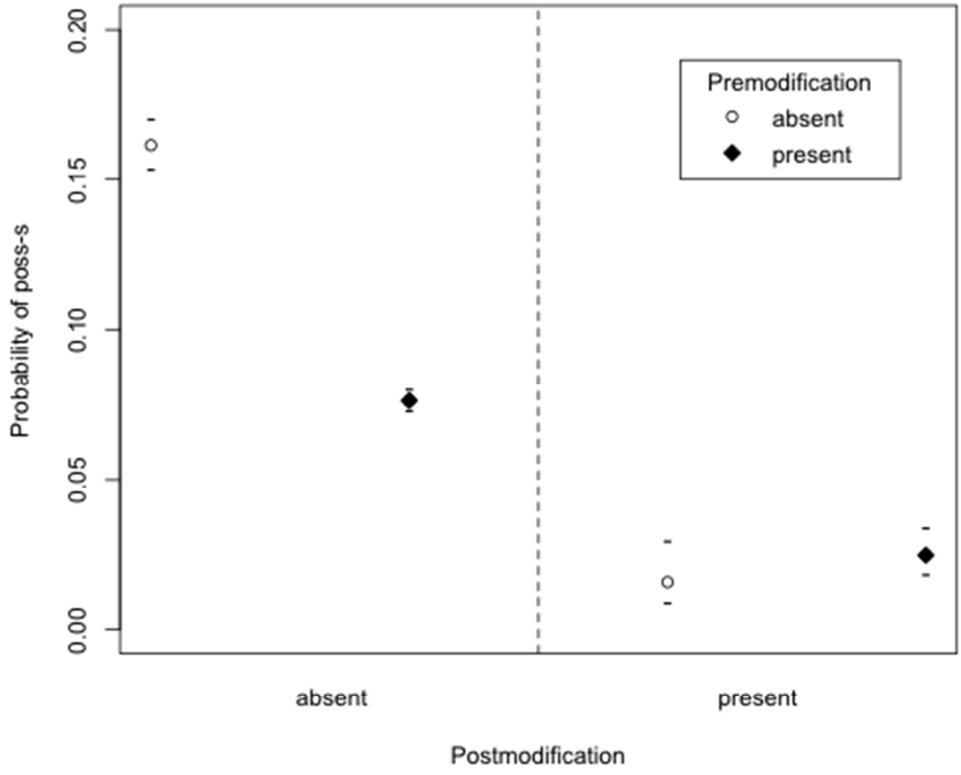

TABLE 14 FREQUENCY OF POSSESSIVES FOR DIFFERENT LENGTHS OF POSSESSOR PREMODIFICATION

Type of possessive	Length of premodification						
	1	2	3	4	5	6	7
POSS-OF	1007	330	116	36	12	1	1
POSS-S	37	3	0	0	0	0	0

TABLE 15 FREQUENCY OF POSSESSIVES FOR DIFFERENT LENGTHS OF POSSESSOR POSTMODIFICATION

Type of possessive	Length of postmodification						
	1[1]	2	3	4	5	6	7
POSS-OF	66	391	489	265	153	96	43
POSS-S	0	33	6	1	0	0	0

[1] The number of instances of POSS-S with postmodification length=1 may appear surprisingly low given how common possessors of the type *someone else's* are. However, the vast majority of these (119 out of 123) involved a possessum of length 0 or 1 and hence were removed at an earlier stage, see discussion of Figure 1.

TABLE 16 PRESENCE OF SPLIT AS A FUNCTION OF POSTMODIFICATION LENGTH

Split	Length of postmodification							
	1	2	3	4	5	8	9	10
absent	78	83	21	1	0	0	0	0
present	0	2	8	1	1	1	1	1

* The work reported here was carried out as part of the project 'Germanic possessive -*s*: an empirical, historical and theoretical study', funded by the Arts and Humanities Research Council. We gratefully acknowledge their support. A very early version of this paper was presented at the project workshop held in Manchester 3-4 April 2009. We are grateful for helpful discussion there and also to Dick Hudson and John Payne, who provided useful comments on the written version. Maciej Baranowski and Inbal Arnon have been very generous in helping us with earlier work on statistical analysis. We would also like to thank Stephanie Dipper for a discussion of aspects of the categorisation. Only we can be held responsible for the final content. Until 30 August 2009, Alan Scott's affiliation was The University of Manchester.

[1] There are also different ways of referring to its function, in particular whether it should be described as a genitive case marker or not.

[2] One reason is the fact that 'phrasal' can be used in different ways, as indicated by the discussion of 'phrasal affix' above. Furthermore, in Payne & Huddleston (2002) phrasal genitive is contrasted with head genitive, and group genitive is then a more neutral term and as such suits our purposes here.

[3] A version of that database also including data from a corpus of spoken Swedish can be accessed online at http://www.llc.manchester.ac.uk/research/projects/germanic-possessive-s/data/.

[4] This is not always the case; compare *a delay of about twenty minutes* (KRT 5372) with *about twenty minutes' delay*.

[5] In our application of regression analysis, SE, *z*, and *p* value for the intercept are irrelevant, but are included in the table since it is part of a standard regression analysis output.

[6] Some authors do not explicitly make the distinction between structural complexity and length but use length to determine what they refer to as structural factors.

[7] Wolk, Bresnan, Rosenbach & Szmrecsanyi (2011: 10) found a similarly close correlation between different ways of measuring length, but found that 'using the individual number of characters provided the best results'. Given the indirect correspondence between orthographic characters and sound in English that is a characteristic of the English spelling system, this is to us an unexpected outcome. We did not compare the length in characters with the results we got for word and syllable.

[8] The codes after some examples are references to the BNC.

[9] The actual interpretation is slightly more complex, since all length-related predictors we discuss in this paper were log transformed before entering them into a regression model. This is a standard procedure to ensure their better compliance with some basic assumptions of regression analysis. It means that in fact the relations modelled are not linear and the actual effects become weaker for higher values of length.

[10] For a discussion of the split possessive in earlier stages of English, see Juvonen (this volume).

[11] O'Connor, Maling & Skarabela (this volume) give one of the examples – *He is the son of the well-known politician whose death was announced the other day…* – but with reference to Jespersen (1954: 143).

[12] Kreyer does not discuss the potential interaction between weight and information structural factors.

[13] Note that this is a different use of the term from the priming-related use in Szmrecsanyi (2006).