WINNER OF THE PRESIDENTS'
2001 POST-DOCTORAL PRIZE

PRESIDENTS' POST-DOCTORAL PRIZE

The Presidents' Post-Doctoral Prize is awarded annually to the lecture judged to make the greatest contribution to linguistic knowledge by an author who has completed a doctorate within the preceding ten years but who does not yet have faculty tenure. The judging panel consists of the current LACUS President and Vice President along with all past presidents in attendance at the meeting.

EVIDENCE IN LINGUISTICS: THREE APPROACHES TO GENITIVES IN ENGLISH

Stefan Th. Gries
Southern Denmark University

over the past several decades[1], syntactic research has been dominated by generative linguistics, the main research technique of which are well-formedness judgements of sentences[2]. The resulting methodological difficulties and points of criticism have been amply documented in, e.g., Cowart (1997) and Schütze (1996). However, given the large amount of harsh criticism that is frequently directed at naïvely gathered acceptability judgements (usually accompanied by commitments to corpus analyses; cf. many references cited in Schütze 1996 and contemporary cognitive-linguistic and/or functionalist studies), it is surprising to see that there is only a relatively small number of studies that explicitly compare the ways in which different methodologies yield different (kinds of) data. Since my point is mainly methodological in nature, I decided to investigate a phenomenon that has already been thoroughly studied, namely what I will, for ease of exposition, refer to as the English genitive alternation.

(1) a. the speech of the President
    b. $NP_{Possessed}$ of $NP_{Possessor}$ (= *of*-genitive)
(2) a. the President 's speech
    b. $NP_{Possessor}$'s $NP_{Possessed}$[3] (= *s*-genitive)

Many variables influencing native speaker's choices of constructions have been identified (cf., e.g., Altenberg 1982; Leech, Francis, and Xu 1994; and especially Stefanowitsch 1997 for overviews); for practical purposes, I will concentrate on three only, namely:

- the syllabic lengths of $NP_{Possessor}$ and $NP_{Possessed}$ (cf., e.g., Poutsma 1914) such that short NPs tend to precede long NPs (to be represented as short » long);
- the animacy of the two NPs' referents (cf., e.g., Poutsma 1914; Jespersen 1949; Hawkins 1981);
- the (discourse-)givenness of the referents of the two NPs such that NPs encoding given referents tend to precede NPs encoding new referents (to be represented as given » new; cf., e.g., Altenberg 1980, 1982; Standwell 1982).

The different kinds of data to be discussed are:

(i) intuitions from informed linguists representing the generative approach, where it often seems that the only informant is the investigating linguist himself;

	Spoken data	Written data	Row totals
***of*-genitives**	75	75	150
***s*-genitives**	75	75	150
Column totals	150	150	300

***Table 1**. Composition of the sample of corpus data.*

(ii) corpus data (both spoken and written);
(iii) acceptability judgements from linguistically naïve native speakers.

First, I will show how the results of the different kinds of data relate to one another. Second, I will show how, counter to popular reasoning, syntactic research benefits from the investigation of both *carefully elicited* judgements and *balanced* corpus data. Note once more that the focus is not on finding out something (new) about the genitive constructions—the analysis of the genitive, whatever results it may yield, is merely a means of making a methodological point.

1. METHODS.

1.1. INFORMED LINGUISTS. In order to obtain informally-gathered intuition data from informed linguists, I presented several linguists with the variables' proposed effects and some example sentences and asked them, on the basis of their intuitions as linguists and native speakers, to formulate generalizations

- concerning the power of the variables in determining the choice of construction;
- concerning the (frequency) distribution of the particular features under investigation and the existence of genitive types that are defined by significant co-occurrences of particular variables' values.

1.2. CORPUS DATA. Using MonoConc Pro 2.0, the pseudo-random sample of genitive constructions given in Table 1 was drawn from the British National Corpus (BNC, first edition). Each instance of a genitive was coded with respect to the above variables, that is the syllabic lengths of the two NPs, the (degrees of) animacy of the referents of the two NPs, and the discourse-givenness of the referents of the two NPs[4]. On the resulting data, I carried out a multifactorial ANCOVA in order to (i) estimate each variable's impact on the choice of the genitive and (ii) investigate the expected two-way interactions of variables and the genitive construction (cf. section 2.2.1)[5]. Also, I determined the most significant clusters of variables describing typical genitives (cf. section 2.2.2).

1.3. ACCEPTABILITY JUDGEMENTS. Given the six variables (three for $NP_{Possessor}$ and three for $NP_{Possessed}$) to be analysed, I developed a factorial token set (using Cowart's 1997: 48f. terminology), as shown in Table 2. Thus, for a fully factorial set, 2×4×4×3×3=288

Variable	Levels
Genitive	*of* vs. *s*
Animacy of the $NP_{Possessor}$ ($A_{Possessor}$)	human, animate+non-human, concrete+inanimate, abstract
Animacy of the $NP_{Possessed}$ ($A_{Possessed}$)	human, animate+non-human, concrete+inanimate, abstract
Length of $NP_{Possessor}$ ($L_{Possessor}$) relative to length of $NP_{Possessed}$ ($L_{Possessed}$)	$L_{Possessor} > L_{Possessed} +1$ $L_{Possessor} = L_{Possessed} \pm 1$ $L_{Possessor} < L_{Possessed} -1$

Table 2. *Independent variables manipulated in the questionnaire.*

individual tokens had to be developed. To that end, I combined each of the genitives with each degree of animacy of both $NP_{Possessor}$ and $NP_{Possessed}$ and systematically varied the lengths of the two NPs as well as their referents' degrees of givenness (by means of a sentence preceding the sentence with the genitive to be judged)[6]. In order to increase the likelihood of representative results, various controls were implemented. For example, since the frequency of linguistic elements can distort the results, the frequency of the nouns figuring in the genitives was controlled for by only admitting the 2.5% most frequent words of English (according to the Cobuild electronic dictionary E-Dict). Also, in order not to base the interpretation of the results on a single token set (results might then be due to individual lexical items only), a different though analogously designed token set was developed, yielding a total of 576 experimental items. Then, the list of experimental items was interspersed with 576 filler items of other syntactic constructions with varying degrees of acceptability. The questionnaire was standardised such that each subject received a different set of randomly ordered stimuli and fillers and the required judgement process was explained and exemplified. This included that the scale of grades to be used by the subjects was anchored only at its endpoints (cf. Schütze 1996: 189, n. 12; Cowart 1997:71).

The subjects that participated in this experiment voluntarily were all native speakers of English without training in linguistics and unaware of the exact purpose of the analysis. The resulting acceptability ratings were then analysed using AN(C)OVAS in order to determine how each variable's two-way interaction with the construction influences (or fails to influence) the acceptability ratings.

2. RESULTS AND DISCUSSION.

2.1. INFORMED LINGUISTS.

As to the first question (the degree to which the variables analysed influence the choice of construction), the results are fairly heterogeneous. The following rank-orderings of variables were obtained[7]:

(3) a. $A_{Possessor} > A_{Possessed} > (L_{Possessor} \approx L_{Possessed} \approx D_{Possessed} \approx D_{Possessed})$
b. $D_{Possessor} > A_{Possessor} > L_{Possessor} > D_{Possessed} > (A_{Possessed} \approx L_{Possessed})$
c. $(A_{Possessor} \approx A_{Possessed}) > (D_{Possessor} \approx D_{Possessed}) > (L_{Possessed} \approx L_{Possessed})$
d. $A_{Possessor} > (L_{Possessor} \approx D_{Possessor}) > A_{Possessed} > (L_{Possessed} \approx D_{Possessed})$

Two tendencies can be observed: First, animacy is fairly consistently considered to be among the most important determinants of the constructional choice. Second, $NP_{Possessor}$ is, on average at least, considered to be important and $NP_{Possessed}$ is not. Note also, however, that there is an interaction such that if $NP_{Possessed}$ is important, then it is only in terms of its degree of animacy. On the whole, however, the results are heterogeneous: there is no consistent ranking of variables or NP kinds and we find that variables equated by some linguists are not equated at all by others.

As to the second question (the frequency distribution of features co-occurring [frequently/significantly]), the results were fairly homogeneous. Consider (4) and (5) for the feature clusters (for *of*-genitive and *s*-genitive respectively) claimed to be prominent (blanks indicate that the respective variable was not included in the expected significant type by the informants).

(4)	a.			animate	$NP_{Possessed}$
	b.	long	new		$NP_{Possessed}$
	c.		new	abstract	$NP_{Possessed}$
(5)	a.	short		animate	$NP_{Possessor}$
	b.	short	given	(animate)	$NP_{Possessor}$
	c.	short	given	human	$NP_{Possessor}$
	d.			inanimate	$NP_{Possessed}$

These proposals as to frequency distributions of feature clusters also yield interesting results. First, linguists' estimations concerning the *s*-genitive and the *of*-genitive focussed on $NP_{Possessor}$ and $NP_{Possessed}$ respectively. This is somewhat surprising since both genitives obviously consist of $NP_{Possessor}$ and $NP_{Possessed}$, and I do not know how to explain this unanimous focus on one NP in each construction. Second, possessors in *s*-genitives are in general considered to be short, given, and animate (thus supporting the predictions of given » new and short » long). On the other hand, $NP_{Possessed}$ in *of*-genitives is supposed to be long and new (with disagreement concerning animacy). This, however, ties in with the predictions concerning $NP_{Possessed}$ of the *s*-genitive since a long and new $NP_{Possessed}$ in *of*-genitives violates both short » long and given » new. In other words, giving even such a simple constellation of variables and expected effects to experienced linguists seems to pose computational problems such that the subjects ultimately failed to account for the predicted two-way interaction and produced unexpected and contradictory predictions. Finally, the results of both the variable ranking and the expected feature clusters do coincide to some extent in that both strategies lead us to expect that $NP_{Possessor}$ is more important than $NP_{Possessed}$.

$A_{Possessed}$ / $A_{Possessor}$	abstract		concrete		human		Row totals		
	of	s	of	s	of	s	of	s	total
abstract	80 (+)	37 (-)	9	8	3	2	92	47	139
concrete	22 (+)	0 (-)	20 (+)	1 (-)	0	0	42	1	43
animate + human	9 (---)	58 (+++)	1 (---)	35 (+++)	6	9	16	102	118
Column totals	111	95	30	44	9	11	150	150	300
	206		74		20				

***Table 3**. Genitives relative to animacy in the corpus data (as a (3 × 3) × 2 table)*[9].

2.2. CORPUS DATA

2.2.1. VARIABLE STRENGTHS. As a first step, before we look at the individual variables' effects, let us look at whether the variables singled out for attention correlate with the choice of genitive constructions in the data in any way worth mentioning. Without belabouring statistical technicalities, the overall correlation is fairly high and highly significant, showing that the variables included in the analysis indeed contribute strongly to the alternation[8].

Let us first look at the impact of animacy on the choice of genitive constructions. Consider Table 3, which provides the frequencies of each genitive construction depending on $A_{Possessor}$ and $A_{Possessed}$. The distribution of constructions is, as can be easily seen, different from chance (R_{mult}=.64; $F_{7, 292}$=29.3; p<.001). The cells responsible for this effect contain plusses/minuses (depending on whether the observed frequency is higher/lower than the expected one), the numbers of plusses/minuses indicate the significance level of the cells' deviations from the expected frequencies as determined by a configural frequency analysis (cf. Krauth 1993).

On the level of row and column totals, two results are immediately obvious: first, animate/human $NP_{Possessed}$s are rare and the more human an NP's referent is, the less likely it is to occur as $NP_{Possessed}$. That humans are rarely $NP_{Possessed}$ is, on the one hand, not surprising, given how we conceptualise possession (cf. Taylor 1995:202ff., 1996). On the other hand, it is interesting to note in passing that 206 out of 300 genitive constructions (nearly equally *of*- and *s*-genitives) have an abstract entity as $NP_{Possessed}$ rather than a concrete object (as would be expected from such prototype-based approaches to possession and genitives in English). No similarly clear bias, however, can be observed for $NP_{Possessor}$: animate and human possessors occur often (though abstract possessors are most frequent) and concrete possessors occur only rarely.

Let us finally turn to significant individual (pairs of) cells and, thus, two distinct usage patterns of genitive constructions. On the basis of the data, two significant patterns of genitive usage can be identified.

(6) $NP_{abstract}$ of $NP_{abstract}$ / $NP_{concrete}$
(7) $NP_{animate/human}$'s $NP_{abstract/concrete}$

On the one hand, the *of*-genitive is significantly preferred when both NPs are abstract, when both NPs are concrete, and when the $NP_{Possessor}$ is concrete and $NP_{Possessed}$ is abstract. On the other hand, the *s*-genitive is preferred when $NP_{Possessor}$ is animate/human. These patterns are so strong that, once we know what $NP_{Possessor}$ looks like, we can predict (92+42+102) 236 out of 300 (78.7%) genitive constructions correctly. $NP_{Possessed}$, however, does not play a prominent role when it comes to deciding on a construction.

Space limitations do not permit detailed inspection of the corpus data with respect to the types just mentioned (cf., however, section 2.2.2 below), but a brief comment will serve to indicate the ways an analysis could be continued rewardingly. One such possibility is the analysis of semantic relations between the two NPs involved: The pattern in (6) admits a variety of semantic relations between $NP_{Possessor}$ and $NP_{Possessed}$ such as attribute/holder of attribute, part/whole, etc.; (cf. Stefanowitsch (1997) for an illuminating inventory of relations and their distribution) whereas the semantic relations of the pattern in (7) are most often that of possessor/possession, agent/action and attribute/holder of attribute. That is, even a cursory glance at real data shows the implausibility of assuming that the two constructions are synonymous or used interchangeably; this implies that, at least on the basis of our data, there is no need to derive one construction from the other in any way whatsoever.

The next variable to be investigated is concerned with the syllabic lengths of the two NPs involved in the genitive. According to previous studies, we would expect to find a two-way interaction between the NP ($NP_{Possessor}$ vs. $NP_{Possessed}$) and the genitive construction (*of*-genitive vs. *s*-genitive) such that short » long. A 2-way ANOVA, however, shows that the overall correlation between the kinds of NP and genitives is significant ($F_{3, 596}$=2.97; p=.031), but not in ways we would expect:

- there is a significant main effect such that the two genitives differ with respect to the average lengths of the NPs: *of*-genitives are formed out of longer NPs than *s*-genitives ($F_{1, 596}$=7.14; p=.008);
- the predicted two-way interaction is insignificant ($F_{1, 596}$=.7; p=.405) and the observed tendency is even in the opposite direction of what syntactic-weight approaches would predict; cf. the left part of Figure 1.

That is to say, approaches to the genitive placing a strong emphasis on heaviness of constituents are not supported by the data, a result I found somewhat astonishing. But before we jump to conclusions too hastily, recall that many analyses of corpus data are based on written data only – the present corpus, however, is balanced with respect to the medium so we can easily filter out this effect. Consider Table 4, where (within each medium and across all examples) for each construction the average lengths of $NP_{Possessor}$ and $NP_{Possessed}$ are compared.

A 3-way ANOVA including the medium (spoken vs. written) yielded two results worth further discussion. First, the analysis revealed that the NPs in the written part of the corpus are on average significantly longer ($F_{1, 592}$=11.96, p<.001). Second, and

	oral		written		total	
	of-genitive	*s*-genitive	*of*-genitive	*s*-genitive	*of*-genitive	*s*-genitive
$NP_{Possessor}$	3.7	3.1	4	3.9	3.8	3.5
$NP_{Possessed}$	3.1	3	4.5	3.2	3.8	3.1

***Table 4**. Average NP syllabic lengths of the genitives in the corpus data.*

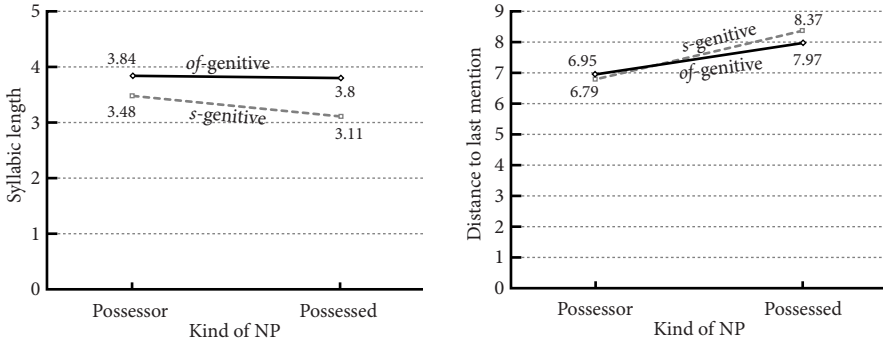

***Figure 1**. Interaction plots: (Genitive × NP) for lengths (left) and (Genitive × NP) for DTLM (right).*

more interestingly, there is a significant 3-way interaction ($F_{1, 592}$=4.48, p<.034) such that:

- for the written data, the two-way interaction is even more in the unexpected direction;
- for the oral data, the two-way interaction is nearly as expected: with *of*-genitives, $NP_{Possessor}$ is longer than $NP_{Possessed}$ – with *s*-genitives, there is practically no difference.

That is to say, we must be careful not to leave aside medium-specific differences: the results for written and oral data diverge so strongly that the unexpected overall results may hide the expected results of the oral data, if the medium is not accounted for carefully. This is an important lesson to learn for corpus-based analyses of syntactic phenomena, especially when one tries to account for syntactic phenomena in terms of processing restrictions or similar variables where medium differences can be decisive.

Finally, let us deal with the discourse givenness of the two NPs and their effect on the choice of genitive. Consider Table 5 (overleaf), where the average values of the distance to last mention (DTLM) in clauses are given.

Again, previous studies lead us to expect a two-way interaction between NP ($NP_{Possessor}$ vs. $NP_{Possessed}$) and the genitive construction (*of*-genitive vs. *s*-genitive) such

	oral		written		total	
	of-genitive	*s*-genitive	*of*-genitive	*s*-genitive	*of*-genitive	*s*-genitive
$NP_{Possessor}$	5.9	7.3	8	6.3	6.9	6.8
$NP_{Possessed}$	7.2	9.1	8.7	7.6	8	8.4

***Table 5**. Average* DTLM *scores of the NPs in the genitives of the corpus data*

that given » new. Again, however, the 3-way ANOVA (this time including the medium right from the start), though highly significant ($F_{7, 592}$=7.7; p<.001), shows that:

- the two genitives differ with respect to the average givenness of their NPs such that the average $NP_{Possessor}$ is more given than the average $NP_{Possessed}$ ($F_{1, 596}$=20.57; p<.001);
- the predicted two-way interaction is not significant ($F_{1, 596}$=.95; p=.329 ns), but the observed tendency is indeed in the predicted direction; cf. the right part of Figure 1.

While the second result is easy to account for (since it is, though non-significant, at least in the correct direction) I find it difficult to account for the first one. An explanation might be that we simply speak about possessors more often since, as we have seen above, they tend to be human. If we speak about them more often, then of course the distance between the different occasions on which we refer to them are closer to one another, resulting in the observed main effect of DTLM. It remains to be seen to what extent the analysis of the acceptability judgements can shed light on this issue.

2.2.2. TYPES OF GENITIVES AS DETERMINED BY SIGNIFICANT FREQUENCY. While the previous section has investigated each variable on its own, let us now look at the genitive types defined by significant feature clusters of all variables simultaneously. While the overall number of significant types (as determined by a hierarchical configural frequency analysis) is too large to be discussed in detail, the most important types for *of*-genitives and *s*-genitives are given in (8) and (9) respectively (the interval variables [length and DTLM] were dichotomised on the basis of their arithmetic mean within each register).

(8) a. $NP_{concrete}$ new short of $NP_{concrete}$ new
b. $NP_{abstract}$ new of $NP_{human/animate}$ short
(9) $NP_{human/animate}$ short 's $NP_{concrete}$ short

On the whole, the types already obtained by the analysis of $A_{Possessor}$ and $A_{Possessed}$ alone are supported—given the above corpus results on length and givenness, it is not surprising to see that the identifiable types do not unanimously support the expected tendencies (short » long and given » new).

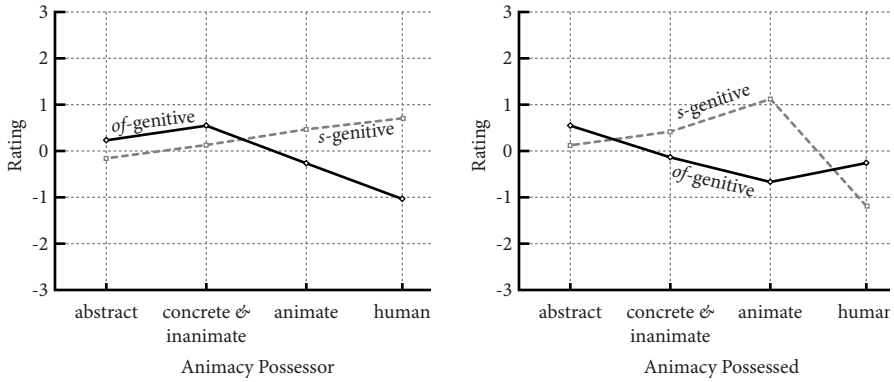

***Figure 2**. Interaction plots: Genitive × $A_{Possessor}$ (left) and Genitive × $A_{Possessed}$ (right).*

2.3. ACCEPTABILITY JUDGEMENTS. As we have seen, the corpus analysis yielded fairly heterogeneous results such that some previous accounts were supported whereas some tendencies one might have taken for granted were not. Let us now turn to the results of the survey and find out to what extent the results fit together. First, again, the overall correlation between all variables and the choice of construction is highly significant[10]. We will now proceed in the same order as for the corpus data and start with $A_{Possessor}$.

While no significant main effect can be found, for the two-way interaction between the genitive and $A_{Possessor}$, we obtain a clear and significant pattern ($F_{3, 499}$=7,15; p<.001): animate and human possessors are preferred in the *s*-genitive whereas abstract as well as concrete and inanimate possessors are preferred in the *of*-genitive. These findings are virtually identical to and, thus, strongly support the results obtained in and interpretations derived from the corpus analysis (cf. the left part of Figure 2)[11].

A somewhat different picture emerges from the analogous analysis of $A_{Possessed}$. First, there is a significant main effect showing that the more human $NP_{Possessed}$ is, the less acceptable are both constructions ($F_{3, 498}$=6.18; p<.001), an effect we are already familiar with from the corpus data. More importantly, however, is the (significant) two-way interaction ($F_{3, 498}$=9.09; p<.001) between the genitive constructions and $A_{Possessed}$. In the corpus analysis, $NP_{Possessed}$ does not differentiate between the two constructions. The results of the acceptability judgements support these results for abstract and concrete $NP_{Possessed}$, which again obtain virtually identical ratings in both constructions. However, animate $NP_{Possessed}$ are preferred in *s*-genitives, whereas human ones are preferred in *of*-genitives. This is interesting in two respects: first, it shows that there is a strikingly high general coincidence of corpus and judgement data. Second, it shows that, where the corpus data have not provided relevant information (recall no cases of animate $NP_{Possessed}$ were found), the judgement data help us to describe the constructional preferences in such cases (cf. the right part of Figure 2)[12]. (Space does not permit the discussion of the marginally significant interaction Genitive × $A_{Possessor}$ × $A_{Possessed}$.)

Let us now turn to the lengths of NPs and degrees of givenness of the NPs' referents. In sum, the results of the questionnaire study again support those of the corpus analysis: while there is a tendency in the direction of short » long, the two-way interaction between the length of the two NPs and the construction clearly fails to reach standard levels of significance ($F_{2, 500}$=1.53; p=.72). Thus length does not seem to play a role in the choice of construction. With distance to last mention the situation is slightly different: the interaction between the distances to last mention and the construction is significant ($F_{2, 500}$=3.28; p=.039), such that there is a tendency of given possessors to be preferred in the *s*-genitive. This is, however, only a tendency, as post hoc tests (Scheffé) reveal no significant differences between the six arithmetic means (as opposed those of $A_{Possessor}$ and $A_{Possessed}$). However, we still need to test whether the main effect noted above (the average $NP_{Possessor}$ is less given than the average $NP_{Possessed}$) has been verified experimentally. In accordance with the corpus data, there is in fact an non-significant tendency in this direction ($F_{2, 500}$=1.81; p=.17): *s*-genitives are preferred when $NP_{Possessor}$ is more given than $NP_{Possessed}$. Thus, while the two kinds of results are as yet inconclusive, the *a posteriori* hypothesis I proposed above could at least be explanatorily adequate.

3. CONCLUSION.

3.1. INTERIM SUMMARY. The intuitions of informed linguists did not convey a unified picture: while we find agreement between the importance of variables and NP types, we obtain contradictory results for the frequent/typical clusters to be expected. The results of the corpus analysis are highly heterogeneous in how the results relate to previous approaches or more general predictions, something often found once natural data are analysed. Still, though, the corpus data have proven useful in several respects: variables could be weighted according to their importance for the alternation, it was possible to identify constructional types, and we saw how the neglect of medium differences can influence (not to say, distort) the results. On the whole, the corpus data correspond to the experimental acceptability judgement data. For most of the variables, virtually complete overlap between the kinds of results was found and, in the case of $A_{Possessed}$, the judgement data even add precision to the corpus findings.

Let us now turn to the more central question, that of how these results relate to the linguists' intuitions? On the positive side, we find that the informants' expectation as to the relevance of the variables was, though far from unanimous, accurate, at least to some degree: $A_{Possessor}$ is indeed the strongest variable determining the choice of construction. Also, the intuitions that (i) $NP_{Possessor}$ of *s*-genitives would frequently be animate/human and short as well as (ii) $NP_{Possessed}$ of *of*-genitives would frequently be new (counter to discourse-functional predictions!) are borne out by the corpus data. On the (I believe somewhat stronger) negative side, however, we find that, on the whole, the linguists failed to predict:

- the complete overall irrelevance of length and givenness to the choice of construction that was found in both the corpus data and the acceptability

judgements (recall the comparisons of means in the ANOVAs) as well as the relative irrelevance of $A_{Possessed}$ to the choice of construction;
- the relevance of the difference between animate and human $NP_{Possessed}$ found in the judgement data;
- the fact emerging from the corpus data that $NP_{Possessed}$ tends to be abstract in both constructions[13].

These results, I submit, strongly support the claim that informed linguists' intuitions on (syntactic) phenomena are inadequate. Obviously, such intuitions can serve as a good, easy-to-obtain and (at times) even accurate starting point of the analysis, but the analyst must be willing to (i) discard every single working hypothesis in the light of evidence to the contrary and (ii) integrate the more fine-grained information of corpus data and methodologically sensible questionnaire studies into his account. The following section addresses this issue in slightly more detail.

3.2. CONCLUSION (AND A GUIDELINE). Given the course of the analysis, I believe the following conclusions are warranted. On the one hand, individual intuitive data may, but need not, provide valuable insights into a phenomenon. Given the overwhelming empirical evidence pointing to potential threats to the objectivity, validity and reliability of intuition data thus obtained, however, I believe that empirically more sensible strategies are required. On the other hand, simply abandoning acceptability judgements in general seems premature, to say the least, since, once gathered in scientifically appropriate ways, they strongly coincide with or even improve on the often desired alternative of corpus data. (For a completely different study where equally refined judgement data are compared to corpus findings with similar results, cf. Gries ms.) Note especially that this coincidence of results has been found for cases where variables have turned out to be important *and* cases where variables turned out to be unimportant.

In sum, on the basis of the above results and the conclusions that can be drawn from the empirical process as such, I suggest the following strategy (of methodologically different but converging evidence) to incorporate all the above methods in a single methodology for a thorough analysis of syntactic phenomena. This strategy does not totally abandon naïvely collected judgement data, but rather treats them as a heuristic exploratory device, the implications of which are subjected to a wide array of methodologically more reliable strategies.

(i) Collect ideas of what variables influence the phenomenon under investigation on the basis of relevant literature as well as introspective data (including people's intuitions) and formulate hypotheses;
(ii) obtain carefully-balanced corpus data (recall the effect of the medium) relevant to the phenomenon under investigation in order to (a) perform exploratory data analysis and (b) gather evidence bearing on one's hypotheses;

(iii) depending on the results of step (ii), conduct methodologically sound experiments (i.e. conforming to standards outlined in Cowart 1997 and Schütze 1996) on those aspects of the phenomenon for which (a) no corpus data could be obtained and/or (b) one's hypotheses were not supported[14];
(iv) repeat steps (ii) and (iii) until you obtain mutually confirming results or identify additional factors.

One important question remains, however: what do we do when the different strategies (e.g. corpus data and judgements) do not yield converging evidence? That is, if there is no single a priori hypothesis in support of a particular interpretation or if each of the two different results can be explained with reference to two mutually exclusive hypotheses, then which of the results (and hypotheses) should be preferred and on what grounds?

The ultimate answer to this question is probably contingent on a variety of factors (such as personal taste, preference for methods of data collection and evaluation, the willingness to admit that the contradictory results cannot be reconciled at present). I would advocate accepting the hypotheses whose supporting results have been obtained most naturally. In other words, if results from corpus data contradict results from acceptability judgements and both could be explained equally well but differently, I would always tend to accept the hypothesis supported by the corpus data: the production of linguistic utterances/texts that happen to end up in a corpus occurred under completely natural circumstances and is, thus, less likely to be subject to experimental bias than questionnaire data (and many other experimental designs). Moreover, I would in general consider corpus data to be more precise in the sense that factors such as register, prescriptive attitudes and medium can be filtered out, whereas we can never be sure to what extent they influence subjects' reactions in experimental settings (even if subjects are advised not to let such factors influence their reactions). Nevertheless, I hope (i) to have shown how, counter to common criticism, careful experimentation by means of acceptability judgement data can support our analysis of linguistic phenomena and (ii) that these findings stimulate further research of this kind.

[1] I thank Hans Boas (University of Texas at Austin), Verena Gries (Unilever Germany), Barbara Lohse (University of Southern California) and Debra Ziegeler (University of Manchester) for their help in obtaining judgement data (by forwarding questionnaires) to be discussed in what follows. Also, my thanks go to Constanze Bühner of Southern Denmark University for helping me encode the corpus data and all colleagues participating in my experiment, even though they might have guessed that the results should show the inadequacy of linguists' intuitions. Finally, I am indebted to Heike Wagner (University of Hamburg) and the Institut for Fagsprog, Kommunikation og Informationsvidenskab at SDU for providing computer equipment and assistant funding respectively. Without the kind assistance of all of these people, the huge amount of data necessary for this study could not have been obtained in time.

Finally, let me note that some of the judgement results have slightly changed since the time of the presentation in Montréal. This is due to the fact that additional questionnaire

data reached me only after my return. However, in all cases but one (where results have undergone a slight change), the results have not changed at all in the light of these additional data.

2 In general, two kinds of well-formedness judgements are distinguished, namely grammaticality judgements and acceptability judgements (i.e. judgements concerned with competence and performance respectively). My study is concerned with acceptability judgements only. However, I believe that both kinds of judgements are difficult to distinguish on a principled basis since, e.g., different versions of generative grammar do not always agree on what factor is a matter of competence or performance. For instance, the introduction of semantic concepts such as theta roles into generative grammar enables generative grammarians to claim that particular semantic phenomena (i.e. phenomena outside of the grammar) can suddenly be explained in grammatical terms.

3 I will use the expressions $NP_{Possessor}$ and $NP_{Possessed}$ throughout the remainder of the paper for expository reasons although in many cases it is not (prototypical) possession that is denoted.

4 The degree of animacy of the NPs' referents was measured using the following scale: human > animate and non-human > concrete and inanimate > abstract. The discourse-givenness of the NPs' referents was measured using the distance to last mention (DTLM) of the referent in the preceding ten clauses. For the purposes of this analysis, expressions qualified as clauses when they contained a noun phrase or a clause as a grammatical subject together with a finite verb; when they were participial or gerundival clauses (e.g., the non-italicised part in *The new rules forbid more than one to put up a sign,* a rule usually ignored); or when a new conversational turn started. However, in order not to be too overly restrictive and proceed with too little context, the following cases were not counted as clauses even if they met one or more of the above-mentioned criteria: question tags; discourse markers such as *you know*, *as it were*, *I mean*; cleft sentences and false starts.

5 We need to analyse interactions rather than main effects because of the different orders of NP types ($NP_{Possessor}$ vs. $NP_{Possessed}$) in the constructions. For example, the preference *short » long* means that possessors should be short and long in the case of *s*-genitives and *of*-genitives respectively, a paradigm case of a two-way interaction.

6 It is well-known that there are also semantic restrictions on the use of the two different genitives. While these semantic variables are not focused upon in the present study, one still needs to take them into account so as not to bias the results systematically. In order to avoid such a skewing in the data, wherever possible I preferred semantic relations between the two NPs that, according to previous corpus-based analyses (Stefanowitsch 1997, to appear), are known to occur in both genitives; such examples include possessor/possessed, component/whole, attribute/holder of attribute, location/thing at location and family relations.

7 In the representations of variable strengths in (3), '>' and '≈' mean 'is more important than' and 'is equally important as'; parentheses are used to support the grouping of similarly influential variables visually.

8 R_{mult}=.65; $F_{14, 286}$=15.33; p<.001; the analysis was an ANCOVA (Type VI sums of squares, no constant, sigma-restricted model).

9 Animate and human possessors were subsumed under a single value because there were only very few animate possessors and no animate possessed at all.

10 R_{mult}=.77; $F_{238, 268}$=1.65; p<.001; the analysis was an ANCOVA (Type VI sums of squares, no constant, sigma-restricted model).

11 This behaviour of human and animate $NP_{Possessor}$s provides post hoc support for grouping these classes together.

12 Note also that the acceptability judgements show that, while human and animate $NP_{Possessor}$ behave identically, human and animate $NP_{Possessed}$ do not.

13 Also, the linguists formulated no register-/medium-specific predictions. Admittedly, I did not ask for those, but it is plausible to assume that the heterogeneity of the above results would not have been resolved by asking the linguists to include even more information in their already very heterogeneous intuitions.

14 Needless to say, I do not advocate experiments where acceptability is the only dependent variables. Alternatives involve operation and selection tests (Quirk & Svartvik 1966), reading and reaction time studies, ambiguity tests, paraphrasing and many more.

REFERENCES

Altenberg, Bengt. 1980. Binominal NPs in a thematic perspective: genitive vs. *of*-constructions in 17th century English. In *Papers from the Scandinavian symposium on syntactic variation*, ed. by Sven Jacobson, 149–72. Stockholm: Almqvist & Wiksell.

———. 1982. *The genitive v. the* of*-construction: a study of syntactic variation in the 17th century English*. Lund: Gleerup.

Collins Cobuild e-dict cd-rom. 1999. Harper Collins.

Cowart, Wayne. 1997. *Experimental syntax: applying objective methods to sentence judgements*. Thousand Oaks: Sage Publications.

Gries, Stefan Th. ms. Towards a corpus-based identification of prototypical instances of constructions. IFKI, Southern Denmark University.

Hawkins, Roger. 1981. Towards an account of the possessive constructions: NP's N and the N of NP. *Journal of linguistics* 17:247–69.

Krauth, Joachim. 1993. *Einführung in die Konfigurationsfrequenzanalyse*. Weinheim: Beltz, Psychologie-Verlags-Union.

Jespersen, Otto. 1949. *A modern English grammar on historical principles,* vol. 7, *Syntax*. Copenhagen: Munksgaard.

Leech, Geoffrey, B. Francis & X. Xu. 1994. The use of computer corpora in the textual demonstrability of gradience in linguistic categories. In *Continuity in linguistic semantics*, ed. by Catherine Fuchs & Bernard Victorri, 57–76. Amsterdam, Philadelphia: John Benjamins.

Poutsma, H. 1914. *A grammar of late modern English*, part II, *The parts of speech*. Groningen: Noordhoff.

Quirk, Randolph & Jan Svartvik. 1966. *Investigating linguistic acceptability*. The Hague: Mouton.

Schütze, Carson T. 1996. *The empirical base of linguistics: grammaticality judgements and linguistic methodology*. Chicago: University of Chicago Press.

STANDWELL, G. B. J. 1982. Genitive constructions and functional sentence perspective. *International review of applied linguistics* 20:257–61.

STEFANOWITSCH, ANATOL. 1997. *On the genitive constructions in English: a cognitive linguistic approach*. Unpubl. M.A. thesis, University of Hamburg.

———. to appear. Constructional semantics as a limit to grammatical alternation: The two genitives of English. In *Determinants of grammatical variation in English*, ed. by Günter Rohdenburg & Britta Mondorf. Berlin: Mouton de Gruyter.

TAYLOR, JOHN R. 1995. *Linguistic categorization: prototypes in linguistic theory*. Oxford: Clarendon Press.

———. 1996. *Possessives in English: an exploration in cognitive grammar*. Oxford: Clarendon Press.