

Guidelines for selecting and coding alternating genitives for NSF project
on
'The development of syntactic alternations'

Draft version, last updated October 2013

Part I. Selecting alternating genitive constructions

The easiest way to define interchangeable contexts for genitive constructions ('choice contexts') is by identifying those contexts which we know do not allow alternation at all ('categorical contexts') and consider the remaining data as 'choice contexts'. In so doing we should keep an open mind to the possibility that some alternations are possible although our own intuition tells us they are not. For this reason we should proceed from whatever well-established and well motivated general principles there are that *categorically* ban one construction rather than starting to make up ad hoc criteria based on our own intuition. Also, for a study focussing on the dynamics of the genitive alternation we should keep in mind that what is categorical today may have been variable in the past, and vice versa.

It is advisable to add a category 'inclusion' in the database which allows for monitoring questionable data that could be either justifiably included or excluded as well as for marking those cases where the coder feels undecided about coding.

1. Selecting s-genitives

1.1 Extraction of s-genitives

Extract all s-genitives from the corpus. Automatic retrieval will yield many occurrences of non-possessive 's, as in *John's* been ill. They have to be removed first. Search only for s-genitives with lexical possessors (no pronominal ones) and make sure you are searching for both singular possessors (*'s) as well as plural possessors (*s').

When searching for s-genitives automatically, keep in mind that in older stages of English (well until the early 18thc) the apostrophe 's was not in place and the s-genitive could be realized by various orthographical forms (e.g. -es/is/ys).

1.2 Classes of s-genitives to be excluded

From the remaining dataset exclude the following constructions:

1.2.1 Constructions with s-genitives lacking an explicit possessorum.

These include local genitives as in (1), where the genitive indicates a location, or independent genitives as in (2), where the possessorum may be omitted if recoverable from the context (cf. Quirk et al. 1985: 329)

- (1) *Let's meet at John's.*
- (2) *This is not my cat, this is Anne's.*

1.2.2 Double genitives

These are genitive constructions where possession is marked twice, both by an *of*-genitive and an *s*-genitive, as in (3), cf. Quirk et al. (1985: 330).

- (3) *a friend of John's*

1.2.3 Fixed expressions

These will be predominantly names or titles containing an *s*-genitive, as in (4).

- (4) a. Murphy's Law
b. *King's Lynn* (place name)
c. *the Queen's bench*
d. *King's College*.

1.2.4 Classifying (descriptive) *s*-genitives

These are genitives with a different underlying semantics and syntax than 'standard' (= determiner) *s*-genitives and they require extra attention in the analysis. They have got a non-specific possessor and the possessor is (usually) adjacent to the possessum. Only exclude those *s*-genitives which are clearly classifying, i.e. if there is no specific referent in the context – this is most clearly an option in idiomatic expressions (as in 5), otherwise look if the syntax clearly points to a classifying structure (as in 6), where the determiner 'his' clearly belongs to the possessum and not the possessor:

- (5) *One of the other traits was the presence of a widow's peak.*
(6) *He wants to get rid of his [smoker's cough].*

Most instances of classifying *s*-genitives will probably be ambiguous between a classifying and a determiner reading, as in the example in (7), where the possessor 'solicitor' could either refer to a specific solicitor with (> determiner reading: 'the office of a (particular) solicitor') or where the possessor is unspecific and the interpretation is 'an office typical of solicitors' (> classifying reading) as the context allows both interpretations.

- (7) He went to *a solicitor's office*.

Include these cases but mark them separately in the dataset as potentially classifying (for example by marking them as '?' in a separate 'inclusion' category). See also Rosenbach (2007) for more details on classifying genitives.

1.2.5 Measure genitives

Measure genitives indicate temporal length (8) or value (9); see also Huddleston & Pullum (2002: 470). They constitute another case of special *s*-genitive construction, which wavers between a classifying and a determiner genitive structure. They often have a temporal possessor (as in (8) and

- (8) a. *one year's salary*
 b. *a moment's time*
- (9) *a dollar's worth of chocolate*

Do not exclude measure genitives but - as ambiguous classifying s-genitives - mark them separately as an unusual genitive construction and flag them by '?' in a separate 'inclusion' category.

2. Selecting *of*-genitives

2.1 Extraction of *of*-genitives

As with s-genitives only extract lexical (non-pronominal) possessors. What complicates the selection of alternating *of*-genitives is the presence of many *of*-constructions which are not possessive in nature. As a first rule, only select constructions which clearly connect 2 NPs, excluding, for example, constructions where the *of* is part of a complex predicate, as e.g. in "... took advantage of the occasion...".

2.2 Classes of *of*-genitives/constructions to be excluded

2.2.1 *Of*-genitive constructions not headed by definite article

As a rule only search for strings of *of*-genitive constructions beginning with the definite article *the*: [*the X [of Y]*], as in (10).

- (10) a. *the population of the United States*
 b. *the top of the barrel*
 c. *the value of voting*
 d. *the character of the child*

In determiner s-genitives the possessor is in determiner position rendering the whole genitive construction definite. A corresponding *of*-genitive construction must therefore be definite and begin with the definite article as constructions with s-genitives cannot have another determiner but the s-genitive. That is, examples as in (11) are to be excluded.

- (11) a. *his idea of the common European home*
 b. *their definition of security*
 c. *a peace of trust*
 d. *40 years of friendship and alliance*
 e. *pictures of Mr Arafat*
 f. *another day of chaos*
 g. *such a display of flags*
 h. *this first meeting of the assembly*

- i. *no immediate recognition of the new government*

Note: The generalization above excludes the following constructions:

- a) **of-genitive constructions with an initial proper name** (e.g. *Anette Rosenbach of Tanagra, President Giscard d'Estaing of France*). They do allow alternation with s-genitives, however, (> *Tanagra's Anette Rosenbach, France's President Giscard d'Estaing*) and there's nothing in the semantics and syntax of s-genitive constructions which would exclude them.
- b) **Clearly semantically definite possessums in certain type of datasets**, which may omit the definite article in otherwise clearly definite constructions. These cases are to be marked by '?' and specified in the column 'inclusion' but in general are included in the analysis. In particular, the following cases may be found in the data:
 - Headlines in press articles: e.g. *Suicide of the Austrian ambassador at Paris*;
 - Constructions such as *CEO of x* or *father of x*, where the possessive construction designates a title or designation and where there is a unique referent. Excluded under this definition are constructions with possessum nouns that do not have a unique referent, e.g. *manager of the place* or *aunt of Margaret*.
 - Spoken language: In spoken languages we sometimes find the definite article left out in expressions that typically have it, e.g. *end of the month*.
 - Earlier English: The English article system becomes largely grammaticalized by the end of Early Modern English but until the 18thc we still occasionally find determinerless constructions that are definite in the texts, e.g. *death of Mr H C Richards MP*. Only include cases here that are clearly semantically definite.

2.2.2 Appositive and partitive of-genitives

From this dataset further exclude cases of

- a) appositive genitives, where possessor and possessum are co-referential, e.g. *the city of Rome*
- b) partitive constructions, e.g. *some of my students*
(NB: part/whole constructions however are to be included, e.g. *the top of the stairs, the frame of the chair*)

2.2.3 Of-genitive is a modifier

If the of-genitive is a modifier qualifying the possessum rather than a complement, it does not alternate with a corresponding s-genitive but with an adjective, see (12)

- (12) a. *a man of honour* (> *an honourable man*, not: *honour's man*)
b. *a ring of gold* (> *a golden ring*; not: *gold's ring*)

As most such constructions are headed by an indefinite article they should be excluded from the start from the dataset (see 2.2.1 above).

2.2.4 Fixed and idiomatic expressions

Constructions where the *of*-genitive is part of a name, title or date must also be excluded as such expression do not allow for alternation, see the examples in (13).

- (13) a. *the Cape of Good Hope*
b. *the House of Commons*
c. *the King of England*
d. *the first of April*

Only dates of the type in (13d) are to be excluded as they are conventionalized and do not alternate with a corresponding *s*-genitive (**April's first*) but rather with an appositive construction (*April first ('1 April')*). In contrast, we do include dates of the form 'the first day of April' as they can alternate with an *s*-genitive in principle (*April's first day*).

Highly problematic are IDIOMATIC EXPRESSIONS (e.g. *in the middle of nowhere*, *the tricks of the trade*). As a rule of thumb, include the expression if it is not fixed completely. A google search may give you an idea whether an expression may alternate in principle although you have yourself strong intuitions against an alternation and thus help you decide (for contemporary data only).

Part II. Coding alternating genitive constructions

1. General lay-out of database

It is good practice to have the full text forms of the possessor, possessor, and complete genitive construction as three separate variables (columns) in the database. This will allow easier coding of alternative predictors (e.g. weight measures), checking for errors, and checking for the values of vaguely named variables. In addition, each genitive construction must be indexed for the text it occurred in (as a text identifier), speaker identity (where applicable), and year of the text (for historical corpora), and of course the outcome variable of construction type produced.

2. Factors and coding levels

All operationalizations given below are to be considered minimum requirements to ensure cross-corpus coding consistency.

2.1 Animacy

The animacy of the possessor is coded, distinguishing the following values:

Code levels	Category	comments	examples
A	Human and animal	Only higher animals (not included e.g. 'fish'); including dead persons unless focus is on body	<i>King, murderer, God, people, horse, bishop, Dutch, Dr Keating, the deceased</i>
C	Collective	organisations (incl. newspapers when focus on organization rather than issue)	<i>the French army (incl. regiments/troops), the Times, Parliament, church, nation, administration, power(s)</i>
		No organisation but potentially variable concord (<i>it/they</i>), animate context	<i>Family, multitude, majority, crew, enemy, convoy, public, expedition, delegation,</i>
T	Temporal	Nouns or adverbs with time reference	<i>February, day, next month, New Year, this evening, moment, yesterday</i>
L	Locative	Places, including 'state'/'empire' (not referable by 'they')	<i>Russia, World, water, town, county, Rome, this kingdom, state, St. John's College, the seas,</i>
I	Other inanimates	Other inanimates – concrete and abstract, all gerunds/infinitives	<i>taxes, justice, vessel, battle, building, refusal, things, war</i>

2.2 Definiteness

The (formal semantic) definiteness of the possessor is coded, categories as follows:

Code levels	Category	comments	examples
Definite	Definite possessors	anything headed by a definite determiner (including definite quantifiers, Lyons 1999), proper nouns/names, s-genitive constructions	<i>the journals, those waters, its most important provisions, our forefathers, one family's murder</i>
Definite-pn	Proper names	Proper nouns and names, excluding clearly common-noun uses	<i>including titles, e.g. the King of England, the Lord Chancellor, his/her Majesty's/Sovereign's, his Grace, and inanimate names (Rhine, February), Government/ President /Court etc. as titles when capitalized in text</i>

Indefinite	Indefinite possessors	anything headed by an indefinite determiner and bare (determinerless) NPs (incl. bare gerunds and indef. quantifiers)	<i>the solidarity of students, the name of a minority, the armies of other nations, the danger of any future troubles, the advantage of being young</i>
-------------------	-----------------------	--	---

Comments:

1. Proper nouns/names: It is sometimes difficult to decide when an expression is being used as a common noun or as a title. For modern texts we can use capitalization as an operationalization, as it indicates that something is being conceived as a name rather than an entity (for example, we can refer to the office of a 'president' or to the person 'the President'). For the ONZE corpus I suggest to follow the intuition of the transcribers (if they perceived of anything as a name they'll probably have capitalized it), in doubt check in dictionaries or on google whether a certain expression is used as a name.

2. Quantifiers: Basically follow the list of Gregory Garretson's manual (HCTC/CSLI Link project):

- **Definite:** *all, all such, both, each, either, every, most, neither*
- **Indefinite:** *all starting with 'a'* (e.g. *a bit of, a few...*), *another, any, enough, few, half, less, little, lots of, many, more, much, no (such/more), none, one, plenty of, several, some, twice*

3. Gerunds: Coded as indefinite unless headed by a definite determiner (e.g. *the voice of my calling*)

2.3 Semantic relation (Genitive functions)

We are using a binary distinction between prototypical possessive relations and non-prototypical ones, following Koptjevskaja-Tamm (2001, 2002) and Rosenbach (2002)¹. For the prototypical relations start out with the fine-grained categories of kin terms, body parts, legal ownership and part/whole relations and collapse them later on into the broader category of 'prototypical' relations.

¹ In her typological work Koptjevskaja-Tamm (2001, 2002) distinguishes between prototypical and non-prototypical possessive relations within the languages of Europe, subsuming legal ownership, kin terms and body parts under the category of 'prototypical' relations and regarding all other relations as 'non-prototypical'. To tease apart the correlation between animacy and possessive relation, we further include part/whole relations of concrete physical entities in the category of 'prototypical' relations, following Rosenbach (2002), who has found significant effects of concrete part/whole relations as opposed to non-part/whole inanimate possessive relations in her experimental study.

Code levels	categories	comments	examples
prototypical	Kin terms	Including all types of family relations	<i>my aunt's daughter, the nephew of the duke</i>
	Body parts	Excluding metaphorical uses (e.g. the jaws of death, <i>in the hands of the federal district courts</i>)	<i>the girl's eyes, the patient's head, the legs of a sports enthusiast</i>
	legal ownership	anything possessed by someone in legal terms	<i>the man's car, Mr Ian Smith's cattle ranch, the Cham's palace</i>
	Part/whole	Applying to concrete physical inanimate possessors only	<i>the chair's frame, the hotel's lobby, the soles of his boots, the hull of the vessel, the top of the stairs</i>
Non-prototypical	any other relation	including derived nominals (subjective/objective genitives) including part/whole relations of non-concrete physical objects/entities	<i>the legal status of slavery, the result of a conference, the captain's name, the branch's performance, Vietnam's southern border, the civil branches of the government, the latter part of August</i>

2.4 Syntactic weight/ length

Identify and code for the possessor phrase and the possessum separately in the database. This will allow for considering / comparing different measures of weight/length such as number of orthographic words (usually) or number of characters (see Wolk et al, forthc.). Per default proceed from orthographic words. Count exactly the number of words for each possessor and possessum (in the *of*-genitive the initial article is not to be counted). In hyphenated words each part counts. In compounds and complex names count each orthographic word. Abbreviations (e.g. VCR, IQ) are counted as one word. Contractions, as in *I've* or *you're*, *he's* or *don't* are counted as separate words (i.e. *I've* = 2 words)

(NB: In the Stanford dataset extracted from the Switchboard corpus of American English and the Brown corpus (Grafmiller, to appear) any possessor of length greater than 10 words was coded as a length of '11' to avoid 'undue influence from extremely long possessors'; see Grafmiller (to appear). If we code the exact number of words we can later on still identify any possessors longer than 11 words and level them out.)

examples	Possessor	Por weight/length	Possessum	Pum weight/length
<i>the power of convincing proofs and arguments</i>	<i>convincing proofs and arguments</i>	4	<i>the power</i>	2
<i>Queen Anne's county goal</i>	<i>Queen Anne</i>	2	<i>county goal</i>	2
<i>the president-elect's principal appointments</i>	<i>the president-elect</i>	3	<i>principal appointments</i>	2
<i>Lord Salisbury's soninlaw</i>	<i>Lord Salisbury</i>	2	<i>soninlaw</i>	1
<i>Lord Salisbury's son-in-law</i>	<i>Lord Salisbury</i>	2	<i>son-in-law</i>	3
<i>the son-in-law of Lord Salisbury</i>	<i>Lord Salisbury</i>	2	<i>the son-in-law</i>	4

In a second step remove the initial definite article in the *of*-genitive from the count as it doesn't have a correspondence in the *s*-genitive, see also Altenberg (1982: 79), Rosenbach (2005: 623) and Hinrichs & Szmrecsanyi (2007:453-454), Wolk et al. (forthc.) for the same procedure.² See the example of *Lord Salisbury's son-in-law* versus *the son-in-law of Lord Salisbury*. When strictly counting the orthographic words the possessum is 3 words in the *s*-genitive but 4 words in the *of*-genitive.

If the possessor and the possessum and their respective weights/lengths are coded separately in the way described above, it should be no problem to calculate also their relative weight/length as a further measurement.

In the spoken datasets hesitation marks or fillers (*uhm*), hesitation repeats (*the...the*) and discourse markers (e.g. *you know*, *sort of*, *like*) have been excluded from the word count as they are strictly speaking not part of the syntactic frame of the possessor or the possessum, although of course they may well affect genitive choice. In the case of discourse markers it can sometimes be difficult to decide whether an item is a true discourse marker or part of the possessor/possessum. In doubt, we regard an item to be a discourse marker if it follows a pause or hesitation, as e.g. in *the ah sort of elements of language*.

² Note that including the definite article in the *of*-genitive would introduce a circularity into the model. The model is meant to predict the outcome variable (construction choice) given the input variables, and it is not legitimate to define one of the input variables in terms of the outcome variable/construction choice.

2.5 Persistence

It is advisable to provide for different measures of persistence within a dataset when compiling a new genitive dataset. For existing datasets either one of the following two options should be used:

We code for the type of genitive construction (*s*-genitive/*of*-genitive) which was used either

- a) the last time there was a genitive choice, and/or
- b) within a certain space or time frame.

Persistence is to be coded for individual conversations (within and across turns). The first genitive in each conversation or text file is automatically to be coded as 'none'. To allow later post-editing for individual corpora it is desirable to include a substantial amount of preceding context (with time-stamps, if available). It is important to consider only genitives in genitive choice contexts, excluding any genitive occurrences that have been excluded from the analysis (see Part I above). At the moment the space differs among corpora (for option b) and needs to be levelled out at a later stage, so make sure the number of words or the time can be tracked.

2.6 Final sibilancy

Code for presence (yes/1) or absence (no/0) of a final sibilant of the possessor. As final sibilant consider

- [s]
- [z]
- [ʃ]
- [tʃ]
- [ʒ]
- [dʒ]

To identify the final sibilancy of the possessor use the phonological segment annotation the CMU dictionary of the variety of your corpus data, following Shih et al. (forthc.), Grafmiller (forthc.) and Wolk et al. (forthc.).

References

- Altenberg, Bengt (1982). *The Genitive v. the of-construction: A study of Syntactic Variation in 17th Century English*. Malmö: CWK Gleerup.
- Grafmiller, Jason (forthc.). "Variation in English genitives across modality and genre". To appear in *English Language and Linguistics*.

- Hinrichs, Lars & Benedikt Szmrecsányi (2007). "Recent changes in the function and frequency of standard English genitive constructions: a multivariate analysis of tagged corpora." *English Language and Linguistics* 11(3): 437 – 474.
- Huddleston, Rodney & Geoffrey K. Pullum (2002). *The Cambridge Grammar of the English Language*. Cambridge: CUP.
- Koptjevskaja-Tamm, Maria. 2001. "Adnominal possession". *Language Typology and Language Universals*, volume 2 of *Handbooks of Linguistics and Communiction Science* ed. by M. Haspelmath, E König, W Oesterreicher & W. Raible, 960–70., Berlin & New York: Mouton de Gruyter..
- Koptjevskaja-Tamm, Maria. 2002. "Adnominal possession in European languages: form and function". *Sprachtypologie und Universalienforschung* 55.141–172.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik (eds) (1985). *A Comprehensive Grammar of the English Language*. London: London.
- Rosenbach, Anette (2002). *Genitive Variation in English*. Berlin: Mouton de Gruyter.
- Rosenbach, Anette (2005). "Animacy versus weight as determinants of grammatical variation in English. *Language* 81(3): 613-644.
- Shih, Stephanie, Jason Grafmiller, Richard Futrell & Joan Bresnan (forthc.). "Rhythm's role in genitive construction choice in spoken English." In: Ralf Vogel & Ruben van de Vijver (eds). *Rhythm in Phonetics, Grammar and Cognition*. Berlin: Mouton de Gruyter.
- Szmrecsányi, Benedikt (2006). *Morphosyntactic Persistence in Spoken English: A corpus study of the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Berlin: Mouton de Gruyter.
- Wolk, Christoph, Joan Bresnan, Anette Rosenbach & Benedikt Szmrecsányi (forthc.). "Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change." To appear in *Diachronica*.