



German long and short genitive endings and their spacial distribution

Workshop “Germanic Genitives”, Freie Universität Berlin, 2014.

Simone Ueberwasser M.A.



Die Endung des Genitiv(e)s

Evaluating a corpus of contemporary newspaper texts from Germany, Austria and Switzerland, is there an observable difference in the ratio of the long and the short genitive endings used, depending on the texts' origin?



Outline

Alternating Genitive Endings in German

The corpus

Evaluation of the Data

Influences

Conclusions



Outline

Alternating Genitive Endings in German

The corpus

Evaluation of the Data

Influences

Conclusions



Possible genitive endings

- ▶ No *-(e)s* ending (E.g. *femina: der Frau*, 'of the woman')
- ▶ Only short endings (E.g. *des Friedens*, 'of peace')
- ▶ Only long endings (E.g. *des Sitzes*, 'of the seat')
- ▶ Alternation possible (E.g. *des Vereins/des Vereines*, 'of the association')

- ▶ Compounds and derivations often show a stronger tendency towards the short genitive ending than the simplexes they end on (cf. Szczepaniak, 2010).



Alternation possible: Tendencies

End syllable of lemma	Origin of word		Example
	Germanic	Foreign	
<i>-sch, -tsch, -st</i>	+++	++	<i>Zwist, Busch</i>
End vowel plus 1 or more consonants			
Single consonant		---	<i>Resultat, Profit, Flop</i>
Unstressed	-	---	<i>Vertrag, Fahrrad, Urwald</i>
Stressed end vowels			
Diphthong	--		<i>Bau, Heu, Abbau</i>
Long vowel	---		<i>Stroh, Reh</i>

Following: *Duden Grammatik* (2009: 197-199)



Alternating Genitive Endings in German

Summary

Important influencing factors:

- ▶ The length of the lemma in syllables
- ▶ the phonology of the ending
- ▶ The stress pattern
- ▶ The origin of the word

Compounds do not behave like their basis.



Outline

Alternating Genitive Endings in German

The corpus

Evaluation of the Data

Influences

Conclusions



„Variantengrammatik des Standarddeutschen“

- ▶ Tri-national Project (D, CH, AT)
- ▶ Aim: Handbook of areal linguistic variation
- ▶ Standard German (no dialects)
- ▶ Cf. www.variantengrammatik.net



Data in the corpus

- ▶ Regional parts of 69 online newspapers (e.g. Kölner Stadt-Anzeiger, Lausitzer Rundschau)
- ▶ 643'502'344 tokens
- ▶ Annotation:
 - (a) TreeTagger: PoS; lemma.
 - (b) RFTagger: PoS; morphosyntactic features like case
 - (c) Struktural units (Author, source, data, url, etc.).
 - (d) Syntactic units (sentence, clause etc.).
- ▶ Not (yet?) publicly available



Corpus size

- ▶ **Total: 595'986'953**
- ▶ Germany: 499'112'306
- ▶ Austria: 57'027'064
- ▶ Switzerland: 28'173'761
- ▶ Belgium: 3'745'540
- ▶ Luxembourg: 3'808'477
- ▶ Liechtenstein: 887'957
- ▶ South Tyrol: 3'231'848



Methodology

- ▶ Query lemmas that end in *-(e)s* and are annotated as genitives.
- ▶ Aggregate them by region and genitive ending.
- ▶ Remove small countries (too many hapax legomena, too often idiolectal).
- ▶ Remove lemmas with <10 occurrences per country.
- ▶ Manual clean up (e.g. *Adel* ('gentry') and *Adele* (as a proper name)).



Outline

Alternating Genitive Endings in German

The corpus

Evaluation of the Data

Influences

Conclusions



Outline

Alternating Genitive Endings in German

The corpus

Evaluation of the Data
Switzerland
Austria

Influences

Conclusions



χ^2 -tests of independence

0-hypothesis:

- ▶ The ratio between long and short genitive endings in CH and in AT is the same.
- ▶ The ratio between long and short genitive endings in CH and in DE is the same.

Both 0-hypothesis are wrong with a probability of 95% ($p < 0.05$):

- ▶ With CH having significantly more lemmas with short genitives or
- ▶ With CH having significantly more lemmas with long genitives.

	for DE & AT (n=591)		DE (AT = \emptyset) (n=190)		AT (DE = \emptyset) (n=5)	
		[%]		[%]		[%]
Short	85	14.4	28	14.7	2	40
Long	2	0.3	9	4.7	0	0



More long genitive forms in CH than in DE and AT

- ▶ *Austritt* ('leaving'): Four occurrences of the long genitive form, two of them from the same author.
- ▶ *Parlament*: Maybe an analogy to *Departement* and *Reglement* ('Regulation'), which show a similar distribution but without values for AT.



More long genitive forms in CH than in DE and AT

Digression: *Depart(e)ment* and its phonetic variants

Ammon et al. (2004: 175-176):

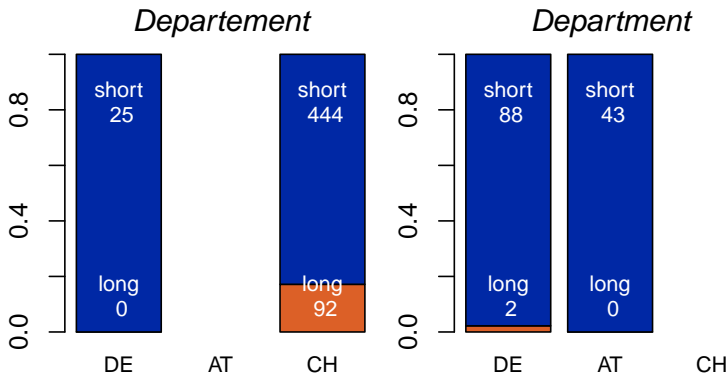
- ▶ DE: /departə'mã:/
- ▶ AT: /depart'mã:/
- ▶ CH: /departə'ment/

The nasal ending in DE and AT disfavor the long genitive ending. *Reglement* shows a similar regional phonetic distribution and also a similar distribution of long and short genitives.



More long genitive forms in CH than in DE and AT

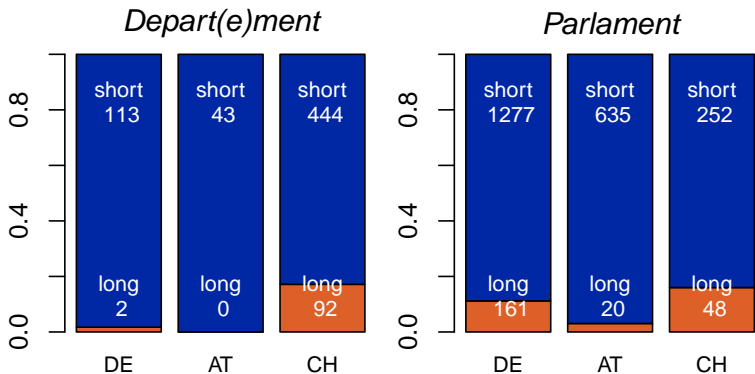
Digression: *Depart(e)ment* and its phonetic variants





More long genitive forms in CH than in DE and AT

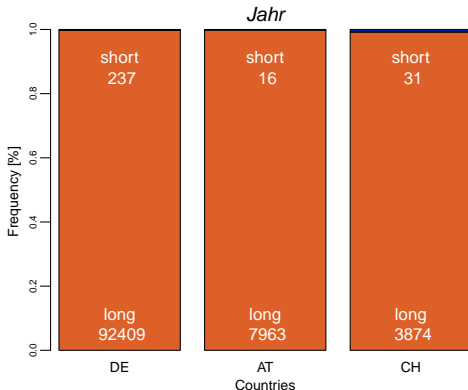
Analogy between *Depart(e)ment* and *Parlament*?





More short genitive forms in CH

Extremely strong tendency towards long genitive forms in all countries, but not quite so strong in CH, e.g. *Jahr* ('Year').

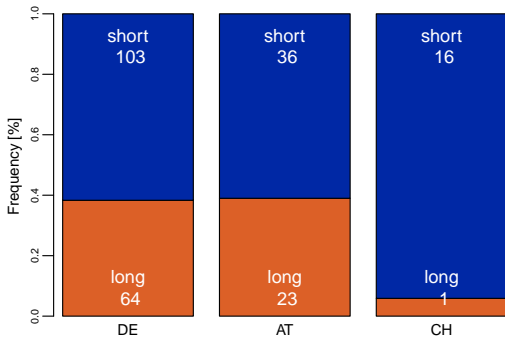




More short genitive forms in CH

Tendency towards short genitive forms in all countries, but decidedly stronger in CH, e.g. *Delikt* ('crime').

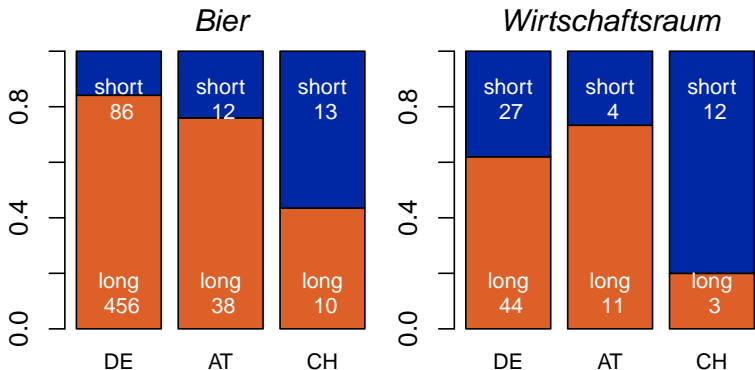
Delikt





More short genitive forms in CH

Tendency towards long genitive forms in other countries, but towards the short form in CH, e.g. *Bier* ('beer') and *Wirtschaftsraum* ('economic area').





More short genitive forms in CH

Inverted tendency for CH

Of all 781 lemmas with > 10 occurrences in DE and CH:

- ▶ 86 lemmas show a tendency towards long genitive forms in DE, but towards the short form in CH.
- ▶ 54 of them are significantly different.

Of all 596 lemmas with > 10 occurrences in AT and CH:

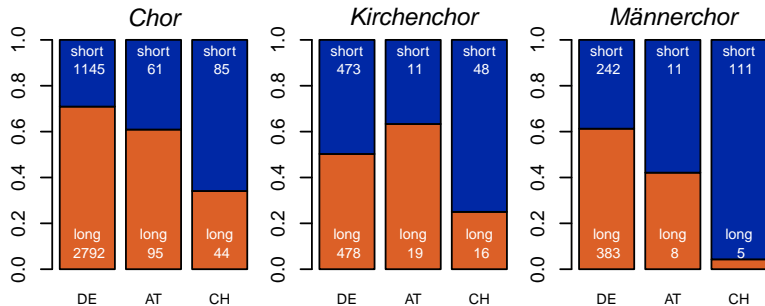
- ▶ 71 lemmas show a tendency towards long genitive forms in AT, but towards the short form in CH.
- ▶ 45 of them are significantly different.



Word formation

E.g. *Chor* ('choir')

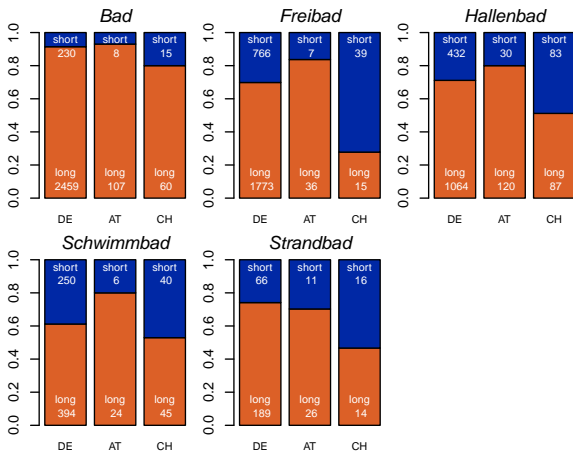
Reminder: Compounds and derivations do not behave like the nouns they end on, they rather tend towards the short genitive.





Word formation

E.g. *Bad* ('bath', 'pool')

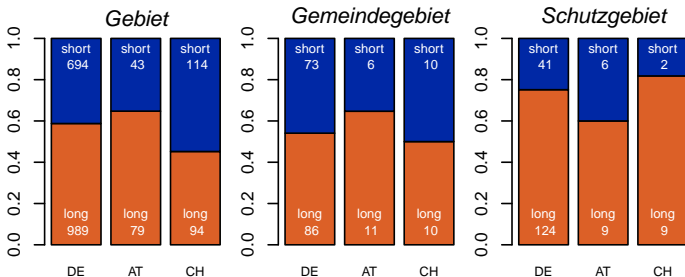




Word formation

E.g. *Gebiet* ('area')

Only few occurrences for the outlier.





Word formation

Concatenating word formations (compounds and derivations)

- ▶ Treat occurrences that end on the same lemma (not morpheme) as one group.
- ▶ If the members of a group show the same distribution, they enforce the significance.
- ▶ If they show an contrary distribution, significance can be lost.



Concatenating word formations

E.g. *Brauch* ('custom, use')

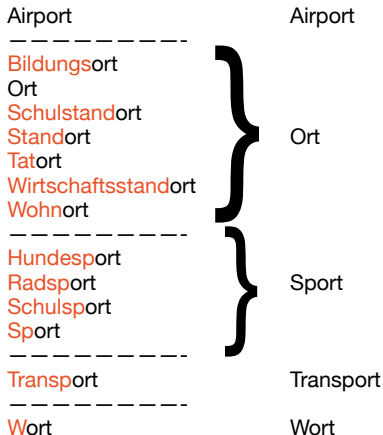
Brauch
Energieverbrauch
Missbrauch
Stromverbrauch
Verbrauch

} **Brauch**



Concatenating word formations

E.g. *Ort* ('location, place')





Concatenating word formations

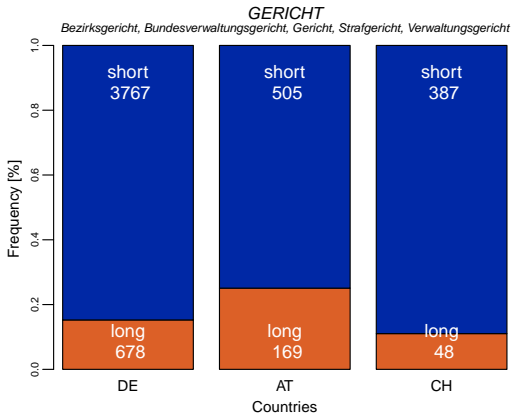
- ▶ 192 Simplexes
- ▶ 124 Concatenated forms
- ▶ 316 forms total

for 63 (22%) of these forms, CH has significantly more short genitive forms than either of the other countries.



Concatenating word formations

More short forms in CH: *Gericht* (1% 'meal', 99% 'legal court')





Switzerland

Summary

- ▶ Clear tendency towards more short genitive endings than in either DE or AT.
- ▶ Tendency is stronger, if the number of occurrences per lemma(-group) is larger.
- ▶ Other regularities, that are given by phonetics or by word formation are maintained and regional distribution is superposed.



Outline

Alternating Genitive Endings in German

The corpus

Evaluation of the Data

Switzerland

Austria

Influences

Conclusions



Individual lemmas

χ^2 -tests of independence with 0-hypothesis:

- ▶ Same ratio between long and short genitive endings in AT and in DE.
- ▶ Same ratio between long and short genitive endings in AT and in CH.

Both 0-hypothesis are wrong with a probability of 95% ($p < 0.05$):

- ▶ With AT having significantly more lemmas with short genitives or
- ▶ With AT having significantly more lemmas with long genitives.

	for DE & AT		DE (AT = \emptyset)		AT (DE = \emptyset)	
	(n=601)	[%]	(n=789)	[%]	(n=5)	[%]
Short	4	0.7	26	3.3	0	
Long	62	10.3	95	12	0	



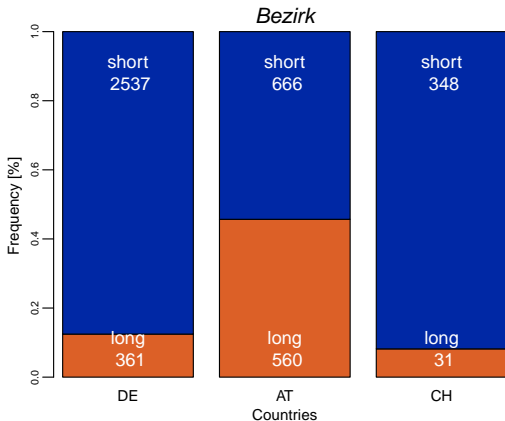
More short genitive forms in AT

- ▶ **Experiment:** Analogy to *Departement*?
- ▶ **Parlament:** Analogy to *Departement*?
- ▶ **Bundesamt:**
 - ▶ 47 of 77 short forms come from the same newspaper (Krone Zeitung).
 - ▶ All compounds of *Amt* have more short genitives than DE, but fewer than CH.
- ▶ **Weltkrieg:**
 - ▶ *Krieg* does not show a significant distribution, neither as a compound nor as a simplex.



More long genitive forms in AT

E.g. *Bezirk* ('district')





Compound groups

χ^2 -tests of independence with 0-hypothesis:

- ▶ Same ratio between long and short genitive endings in AT and in DE.
- ▶ Same ratio between long and short genitive endings in AT and in CH.

Both 0-hypothesis are wrong with a probability of 95% ($p < 0.05$):

- ▶ With AT having significantly more lemmas with short genitives or
- ▶ With AT having significantly more lemmas with long genitives.

	for DE & AT		DE (AT = \emptyset)		AT (DE = \emptyset)	
	(n=6289)	[%]	(n=155)	[%]	(n=5)	[%]
Short	7	2.4	10	6.5	0	
Long	35	12.1	16	10.3	0	



Compound groups: More short gen. forms in AT

For AT-DE-CH and AT-DE (n=17)

All 5 compound groups that end on *-ment* and show a significant difference between AT and DE:

- ▶ Experiment, Parlament, Dokument, Kontingent, Testament

Monosyllabic:

- ▶ Block, Dorf, Geld, Rat, Schritt, Frost, Fund, Kleid, Loch, Schirm, Schlaf, Sinn



Austria

Summary

- ▶ Slight tendency towards more long genitive forms
- ▶ Possible other influences:
 - ▶ Certain endings (-*ment*)
 - ▶ Style/register (Krone Zeitung)?
 - ▶ Phonological parameters?



AT and CH compared to DE

If the values for DE are taken as an in-between value, the following percentage of lemmas show a significantly different number of occurrences with long or short genitive endings:

%	AT	CH
Short	5.1	20.0
Long	13.9	2.9



Outline

Alternating Genitive Endings in German

The corpus

Evaluation of the Data

Influences

Conclusions



Phonological influences on regional distribution

Question

Do any of the following attributes trigger the data from AT/CH to use significantly more long/short genitive forms compared to DE for a specific lemma?

- ▶ The length of the lemma in syllables
- ▶ the phonology of the ending
- ▶ The stress pattern

Data

Lemmas, that show significantly more long/short genitive forms than DE for AT/CH.

Control Group

Other lemmas, that exist in the corpus from AT/CH.

Method

Variable Rule Analysis with Rbrul (cf. Johnson, 2009)



Annotation of the lemmas

Length

- ▶ In phonological syllables
- ▶ Problem: Differences in pronunciation (cf. *Kiosk*)

Endings

- ▶ Phonological values (e.g. *-ng* is one phonological consonant)
- ▶ Values: One consonant, more than one consonant, sibilant, diphthong
- ▶ Problem: Depends on the region (cf. *Département*)

Stress pattern

- ▶ Values: last syllable, not last syllable
- ▶ Problem: Depends on the region (cf. *Labor*)



Outline

Alternating Genitive Endings in German

The corpus

Evaluation of the Data

Influences

Switzerland

Austria

Summary

Conclusions



Endings

%	others	more short
Diphthong	2	0
1 Consonant	58	61
> 1 Consonant	37	37
Sibilant	3	2

$p = 0.0807$



Stress pattern

%	others	more short
Last syllable	30	42
other syllable	70	58

$p = 0.00644$



Syllables

%	others	more short
1 syllable	16	27
2 syllables	37	35
3 syllables	26	25
4 syllables	13	9
5 syllables	6	3
6 syllables	1	1
7 syllables	1	0

$p = 0.000643$



Outline

Alternating Genitive Endings in German

The corpus

Evaluation of the Data

Influences

Switzerland

Austria

Summary

Conclusions



Endings

%	others	more long
Diphthong	1	4
1 Consonant	57	75
> 1 Consonant	41	21
Sibilant	2	1

$p = 2.19e-08$ (without diphthongs: $p = 2.03e-07$)



Stress pattern

%	others	more long
Last syllable	28	22
other syllable	72	78

$p = 0.108$



Syllables

%	others	more long
1 syllable	14	15
2 syllables	34	30
3 syllables	29	24
4 syllables	13	20
5 syllables	6	8
6 syllables	1	0
7 syllables	1	0
8 syllables	0	0

$p = 0.317$



Outline

Alternating Genitive Endings in German

The corpus

Evaluation of the Data

Influences

Switzerland

Austria

Summary

Conclusions



Influences Summary

CH:

- ▶ Lemmas that have the stress on the last syllable have a slightly stronger tendency towards the short genitive form than others.
- ▶ The most important influencing factor is the length of the lemma. Monosyllabic lemmas have a significantly stronger tendency to take the short genitive ending than others.
- ▶ Combining factors does not strengthen the influence.

AT:

- ▶ Lemmas ending on one consonant are the ones that show the strongest tendency towards the long genitive ending.
- ▶ Combining factors does not strengthen the influence.



Outline

Alternating Genitive Endings in German

The corpus

Evaluation of the Data

Influences

Conclusions



Conclusions

- ▶ In the corpus investigated, data from CH show a clear tendency towards shorter genitive endings than in the other countries.
- ▶ In the corpus investigated, data from AT show a remarkable tendency towards longer genitive endings than in the other countries.
- ▶ In the data from CH, the syllabic length is the strongest influencing factor, in AT the mono-consonant ending.
- ▶ In the regional distribution, compounds and derivations often show the same regional distribution as their respective basis.
- ▶ Other extralinguistic influencing factors, e.g. style, cannot be investigated, but seem to be hinted at by individual lemmas.
- ▶ Phonological and regional influencing factors interact.



Thank you

Handouts and presentation on www.ueberwasser.eu



University of
Zurich ^{UZH}

Deutsches Seminar

Appendix





Literatur I

- ▶ **Ammon, Ulrich et al. (2004):** *Variantenwörterbuch des Deutschen. Die Standardsprache in Österreich, der Schweiz und Deutschland sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol.* Berlin, New York: de Gruyter.
- ▶ **Duden (2009):** *Die Grammatik. Unentbehrlich für richtiges Deutsch.* 8., überarbeitete Auflage. (= Der Duden in zwölf Bänden 4). Mannheim: Bibliographisches Institut.
- ▶ **Johnson, Daniel Ezra (2009):** "Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis". In: *Language and Linguistics Compass*, 359–383. url: http://www.danielezrajohnson.com/johnson_compass_final.pdf.



Literatur II

- ▶ **Szczepaniak, Renata (2010):** “Während des Flug(e)s/des Ausflug(e)s? German Short and Long Genitive Endings between Norm and Variation”. In: *Grammar between Norm and Variation*. Ed. by Alexandra N. Lenz and Albrecht Plewnia. (= VarioLingua 40). Frankfurt a. M. et al.: Peter Lang, 103–126.