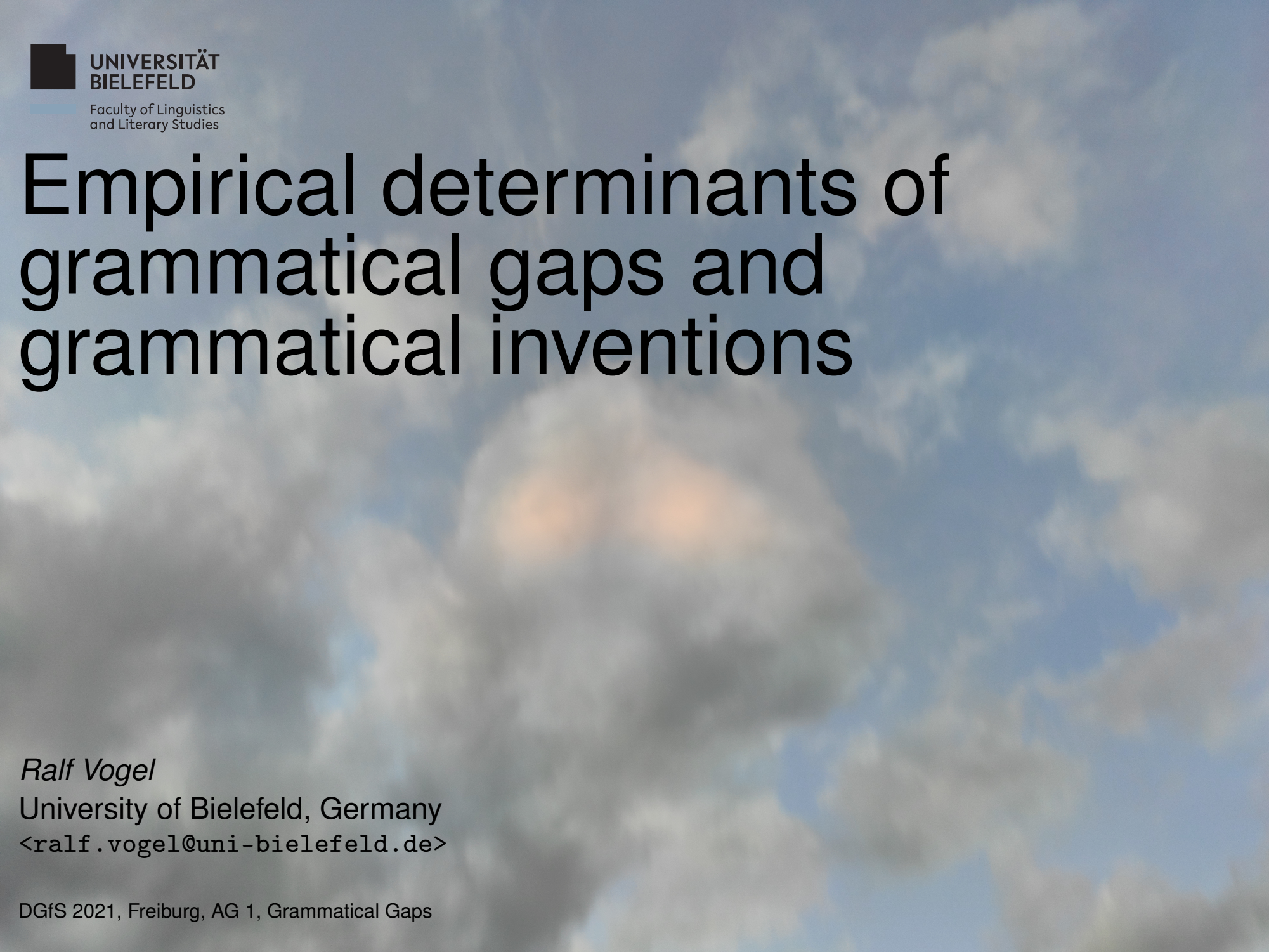


UNIVERSITÄT BIELEFELD

Faculty of Linguistics and Literary Studies

Empirical determinants of grammatical gaps and grammatical inventions

Ralf Vogel
University of Bielefeld, Germany
<ralf.vogel@uni-bielefeld.de>

DGfS 2021, Freiburg, AG 1, Grammatical Gaps

Contents

Aims

1. Which kinds of observations lead to the assumption of a gap?
2. Proposal on observational heuristics
3. Account of speakers' resolution strategies when faced with gaps

Contents

Six methodological remarks

1. "Recognition" of a gap is based on our knowledge, analyses and expectations about a given language and language(s) more generally
   gaps ≈ failed expectations (of linguists)

   Where do these expectations come from?
   ➔ "typicality" inferences from established system, typological considerations

   Typical cases: a) exceptions to productivity
   b) incomplete grammaticalisation
   c) complicated/rare contexts without standard solution

2. reliance on function/meaning:
   English: **gooder* ➔ no function gap, only form **blocked** by *better*
   everything can be expressed in every language: **no function/meaning gaps**

   gaps ≈ non-use of an established construction for a particular function/meaning in a particular situation

Six methodological remarks

3 reliance on **conventionalisation**: distinguish the *collective* and the *individual* dimension
a standardised resolution for a gap may be lacking, but still speakers may systematically employ particular resolution strategies, based on linguistic criteria
→ a window into general linguistic competence
gaps ≈ poverty of the stimulus scenarios

4 relevance of **frequency**:
"Words and concepts are (in a certain sense) tools for communication and thought. Tools are a means of providing standard solutions for recurrent problems. Logically and technically, it is entirely possible that there be a tool for getting tennis balls out of milk bottles. It is solely because a solution to this problem is too infrequently required that no such instrument exists." (Keller, 1998, 65)

gaps ≈ rare communication problems

Six methodological remarks

5 Analogy as expected resolution strategy:

(1) a. What do you think [that Paul thinks [that Anne thinks [that John thinks [that Mary suggested __]]]] ?
b. #**Quid** do you think [that Paul thinks [that Anne thinks [that John thinks [that Mary suggested **it**]]]] ?

(1-a,b): same very rare context;
(1-a): recursive rule or **construction by analogy to simpler cases**;
(1-b): hypothetical ad hoc solution;

because of (1-a), we do not diagnose a gap;
if (1-b) was the chosen construction, we would do so
but (1-a) might be as ad hoc as (1-b) would be, and the situation is identical

gaps $\approx$ failure of construction by analogy / contexts where strange things happen

Six methodological remarks

6 Reliance on established language system:

German:

A) no infinitive form of past tense stems
zu lachten*, 'to laugh-INF.PAST': not attested (as yet?) ➔ **gap

B) no *conventionalised* infinitive of future tense
lachen zu werden ('laugh to AUX.FUT.INF'), occasionally attested (Reiner, 2015)

ad A) blocking by established infinitive of present perfect construction
(*gelacht zu haben*, 'laugh-PTCP to AUX.INF')
plus decline of past tense ('Präteritumschwund')

ad B) present tense dominantly used for future meaning, but no blocking of analytic future (present tense as default tense cannot block other tenses)

Ad hoc constructions

Claim

Grammatical gaps are situations which lack a grammatical solution within the conventionalised language system. In order to solve that problem, speakers need to *invent* new constructions. I call these *ad hoc constructions*. They fulfill the following criteria:

1. Their frequency is below the threshold for grammaticalisation.
2. They have unexpected or irregular properties.
3. They are nevertheless accepted (to a sufficient degree) by speakers.
4. They may alternate with variants that are equally rare.
5. They are preferred over other alternatives which are expected by analogical construction. (failure of analogy)

Ad hoc constructions are newly invented *each time* they are being used. Because of the rarity of their triggering contexts, they do not grammaticalise.
They are examples of the *grammatical creativity* of speakers.

The concepts of grammatical gaps, linguistic invention and ad hoc constructions helps to find more adequate soulutions to some interesting problems for traditional analyses.
➔ two case studies.

Case Studies

- The following two syntactic phenomena from German will be discussed:
  - A infinitival variants of modal verb constructions in present perfect tense
  - B cases of reflexivisation, where the reflexive element is not bound by the subject of the clause
- The steps in the analysis are as follows:
  1. determine their deviant property
  2. determine their rarity
  3. determine their acceptability
  4. grammatical analysis / determine crucial factors for preferring them over more regular alternatives

Contents

Contents

Reflexivisation in German

- German has a reflexive pronoun, *sich*, that is used in place of a dative or accusative noun phrase.
- *Sich* must be coreferent with a clause-mate constituent, usually the subject:

(2) *Maria sah sich im Spiegel*
M. see.PST REFL.ACC in.the.DAT mirror
"Maria sah herself in the mirror"

- Standard assumption: reflexive binding follows a hierarchy of grammatical functions (Pollard & Sag, 1992; Grewendorf, 1988)
- The antecedent must be to the left of the reflexive in the hierarchy (3).

(3) Grammatical function hierarchy for German:
subject $\succ$ accusative object $\succ$ dative object $\succ$ prepositional object

Reflexivisation in German

- Prediction of this account: a dative *sich* can be bound by an accusative antecedent, as in (4):

(4) *Ich zeigte den Patienten$_i$ sich$_i$ im Spiegel*
I show.PST the.ACC patient REFL.DAT in.the.DAT mirror
≈ "I showed the patient to himself in the mirror".
(cf. Grewendorf 1988)

- This empirical claim has been under dispute from the beginning.
- Most authors find (4) odd, if not ungrammatical.
- Reis (1976) already claimed that *sich* has to be accompanied by the intensifier *selbst* 'self' in such cases.

→ binding by object might be a grammatical gap due to rarity of the context

→ Whatever speakers do in that situation, they *invent* it

Acceptability

- Featherston & Sternefeld (2003) report an acceptability study on such object-to-object binding relations, using 16 different possible variants.
- Two clear results:
  1. The antecedent should have dative case and the reflexive accusative case.
  2. The reflexive may be *sich* or a personal pronoun, but has to be accompanied by the intensifier *selbst*.
- ad (1): Vogel (2014), referring to Jackendoff (1992): there is a further semantic factor, namely that the antecedent must not be the copy, while the reflexive is the original.
- ad (2): simple *sich* cannot be used, because binding by object is much too rare, binding by subject being the typical case.

➔ The obligatoriness of *(sich/ihn) selbst* here is the surprising property.

➔ The failure of analogy lies in the low acceptability of plain *sich*

Frequency of the construction

First study: *sich*

- Random sample of 2000 occurrences of *sich* in the corpus research system COSMAS II, IDS Mannheim, archive W-ALL:
- not a single item was found where *sich* was bound by a non-subject complement
- the antecedent always was a subject or an implicit subject (as in infinitival clauses)
- about 80% were cases of reflexive verbs, only 20% were anaphoric uses
- in 5% of the anaphoric cases (1% overall), *sich* was accompanied by the intensifier *selbst*

Frequency of the construction

Second study: *sich selbst*

- Random sample of 2000 occurrences of *sich selbst* in the corpus research system COSMAS II, IDS Mannheim, archive W-ALL:
- 47 errors, 1953 hits.
- 1913 cases (= 98%) of binding by subject (including 454 with implicit subjects). 3 (0.15%) instances of logophoric binding.
- 37 cases of binding by object.
  - 21 cases of ACC→DAT binding, all with the same verb:
    "X_{ACC} $sich_{DAT}$ selbst überlassen" ('leave X to X-self')
    clearly idiomatic, no indicator of grammaticalisation.
  - none of the remaining cases is ACC→DAT or DAT→ACC binding.
- based on the remaining 16 items (= 0.8%), we can now estimate the frequency of object-bound *sich selbst* in speakers' experience.

Frequency of the construction

- total of *sich selbst* in the corpus: 610 809
  size of the corpus: 9.7 billion words
  average number of words speakers experience by day:
  100 500 (Bohn and Short 2009, for English)
  (certainly on the upper end: with 2-words/sec rate $\sim$ 14 hrs. nonstop talking)
- number of days needed to read the corpus:
  9 700 000 000/100 500 = 96 517.41 days (= 264.43 years)
- 0.82% of non-idiomatic object-bound *sich selbst*.
  confidence interval: 0.47 - 1.33% (with *R*'s binomial test).
- in totals (from 610 809): 5 008.6 (2 870.8 - 8 123.8)
- How many days does it take to experience one instance?
  96 517.41/4 688.9 = 19.3 days (11.8 - 33.6 days)
- Is that sufficiently often to call *sich selbst* a grammaticalised object-bound reflexive expression?

Grammatical Analysis

- An easy account of this can be given in terms of Levinson's (2000) theory of generalised conversational implicature, in particular, his M-implicatures:

M-implicature (modality) The use of unusual expressions signals a deviation from the stereotype. (my wording)

- The stereotypical use of *sich* is binding by subject. The addition of the intensifier signals a deviation from this stereotype, here: binding by non-subject.
- Similarly, it can be argued for "PRON *selbst*" that the additional intensifier signals exceptional binding of the pronoun from within the clause.
- The two variants are preferred over alternatives in those rare cases of binding by object, presumably without being grammaticalised (therefore *two* variants).
  → **ad hoc constructions**

Contents

German Perfect

- German perfect constructions come in several variants
- The auxiliary may be a form of *haben* ‘have’ (5-a) or *sein* ‘be’ (5-b)

(5) *Ich habe gelacht*
I have.3SG.PRS laugh.PTCP

(6) *Ich bin angekommen*
I be.1SG.PRS arrive.PTCP

- Choice of perfect auxiliary is dependent on lexical, constructional and/or semantic preferences, and also variable among different regions in the German speaking countries.

Constructions

- I use a version of construction grammar:

  "Each construction will be a form-meaning pair (F,M), where F is a set of conditions on syntactic and phonological form and M is a set of conditions on meaning and use." (Lakoff 1987, 467; emphasis mine, RV)

- On the form side, constructions are sets of construction *components* (= elements of Lakoff's F).
- Syntactic constructions are categorised (as sentence, verb phrase, noun phrase etc.).

German perfect

- The two constructions:

(7) *haben* perfect
category: VP
comp 1: HABEN
comp 2: [$_{VP}$ ... V_{PTCP} ...]

(8) *sein* perfect
category: VP
comp 1: SEIN
comp 2: [$_{VP}$ ... V_{PTCP} ...]

(capital letters: some word form of *sein* or *haben*;
VP = verb phrase; comp = component)

German perfect

- *Infinitival* variants of the perfect construction are possible in case of embedding (9) or infinitival clauses (10):

(9) *Grindel soll Zuwendungen verheimlicht haben.*
G. is said to gifts.ACC conceal.PTCP have.INF

(10) *Grindel wird verdächtigt, Zuwendungen verheimlicht zu haben.*
G. is being suspected, gifts.ACC conceal.PTCP to have.INF

"Grindel is suspected of having concealed gifts."

Infinitivus pro participio (IPP)

- Some verbs which embed another VP diverge from this pattern.
- These are first of all modal verbs: *wollen* ‘want’, *müssen* ‘must’, *können* ‘can’, *sollen* ‘shall’ etc.
- These verbs select *haben* ‘have’ as auxiliary, but they have an infinitive instead of the participle:

(11) *Er hat es nicht verheimlichen können.*
He has it not conceal-INF can-INF
“He wasn’t able to conceal it.”

- Other such verbs are *lassen* ‘let’ and the perception verbs *hören* ‘hear’ and *sehen* ‘see’.

Infinitivus pro participio (IPP)

- Most modal verbs can also be used with an accusative object instead of an infinitival complement.
- In this case, the participle must be used in the perfect construction:

(12) *Er hat es nicht gekonnt.*
He has it.ACC not can.PTCP
"He wasn't able to do it."

- But this use is exceptional. In 90+% of all cases, modal verbs are used with an infinitival complement. (estimation based on random corpus samples)
- This strong preference seems to justify the IPP in the first place.

Infinitivus pro participio (IPP)

- We have two infinitives, the modal verb and a lexical verb, but because modals usually embed an infinitive, the lack of morphological contrast is doing no harm.

(13) *Er hat es nicht verheimlichen können.*
He has it not conceal.INF can.INF
"He wasn't able to conceal it."

- Given the German language experience of an average speaker, it is highly likely that *können* embeds *verheimlichen* in (13), irrespective of morphosyntactic properties.

Clause-final complexes

- Clause-final verbal complexes display “132” order and “312” order:

(14) a. . . . *hat verheimlichen können* 132-IPP
b. . . . *verheimlichen hat können* 312-IPP

- 132-IPP is unmarked, 312-IPP is marked and especially common in Austria, Switzerland and the Southern areas of Germany
- by analogy, infinitival variants like (15) are expected to be used:

(15) a. . . . *haben verheimlichen (zu) können* 132-IPP
b. . . . *verheimlichen haben (zu) können* 312-IPP

- But this is extremely rarely the case.
  ➔ **failure of analogy**
- The default 321 pattern for perfect constructions in (16) is also rarely attested:

(16) . . . *verhindern gekonnt zu haben* 321

Infinitival perfect of modal verbs

- In the mostly used variant of an **infinitive of a perfect of an embedding modal verb**, the participle surprisingly reappears, but in the wrong place (participium pro infinitivo, PPI).

(17) *Das Glück der FN ist, niemals regiert haben zu müssen.*
the luck of the FN is never govern.PTCP have.INF to must.INF
"The FN is lucky that it never had to govern." (Die Presse, 28.08.1997)

- The lexical verb *regieren*, being embedded by the modal verb, should be an infinitive.
- This construction seems to be an isolated instance of PPI in German.
  ➔ the exceptional property of the construction.
- Haider (2011); Reis (2017): "grammatical illusion"; Vogel (2009); Wurmbrand (2012): follows from properties of German grammar
- Claim: somehow, both camps are partially right and wrong . . .

Frequency of use

- Vogel (2009): corpus study (Cosmas II, IDS Mannheim), full counts of the four attested variants of such *zu*-infinitival 3-verb complexes:

X = ...	312-IPP/PPI V_{PTCP}-Aux_{INF}-X_{INF}	312-IPP V_{INF}-Aux_{INF}-X_{INF}	132-IPP Aux_{INF}-V_{INF}-X_{INF}	321 V_{INF}-X_{PTCP}-Aux_{INF}
causative *lassen* 'let'	**12**	4	0	2
modal verbs	**62**	0	1	6
non-causative *lassen*	3	2	0	**19**
perception verbs	0	0	0	**32**
sum	77	6	1	59
total		143		

Table 1: Distribution of *zu*-infinitival 3-verb complexes in the archive *W-public* of the "Deutsches Referenzkorpus", IDS Mannheim, (Vogel, 2009, 311)

- Preferred variants are those with participles.
- PPI for modal verbs and causative *lassen*, and a construction with a participial variant of X for perception verbs and non-causative *lassen*.

Frequency of use

- IDS archive W-public (Cosmas II): by the time of the corpus search, it contained 1.15 billion word forms (Vogel, 2009, 309).
- Average linguistic input of speakers: 100 500 words per day (see above)
- 1.15 bln / 100500 / 365 = 31.35 years to read the corpus.
- This would result in an average of 4.56 (= 143/31.35) zu-infinitival 3-verb complexes per year, only about half of which (77/31.35 = 2.46) would show PPI.

➔ This is presumably too rare for the construction to become grammaticalised.

- Still, within these small numbers, the preferences are clear.

Acceptability Study

- 48 students in Bielefeld, 7-point rating scale, written questionnaire, for finite and infinite versions of the four variants found in the corpus search. Results:

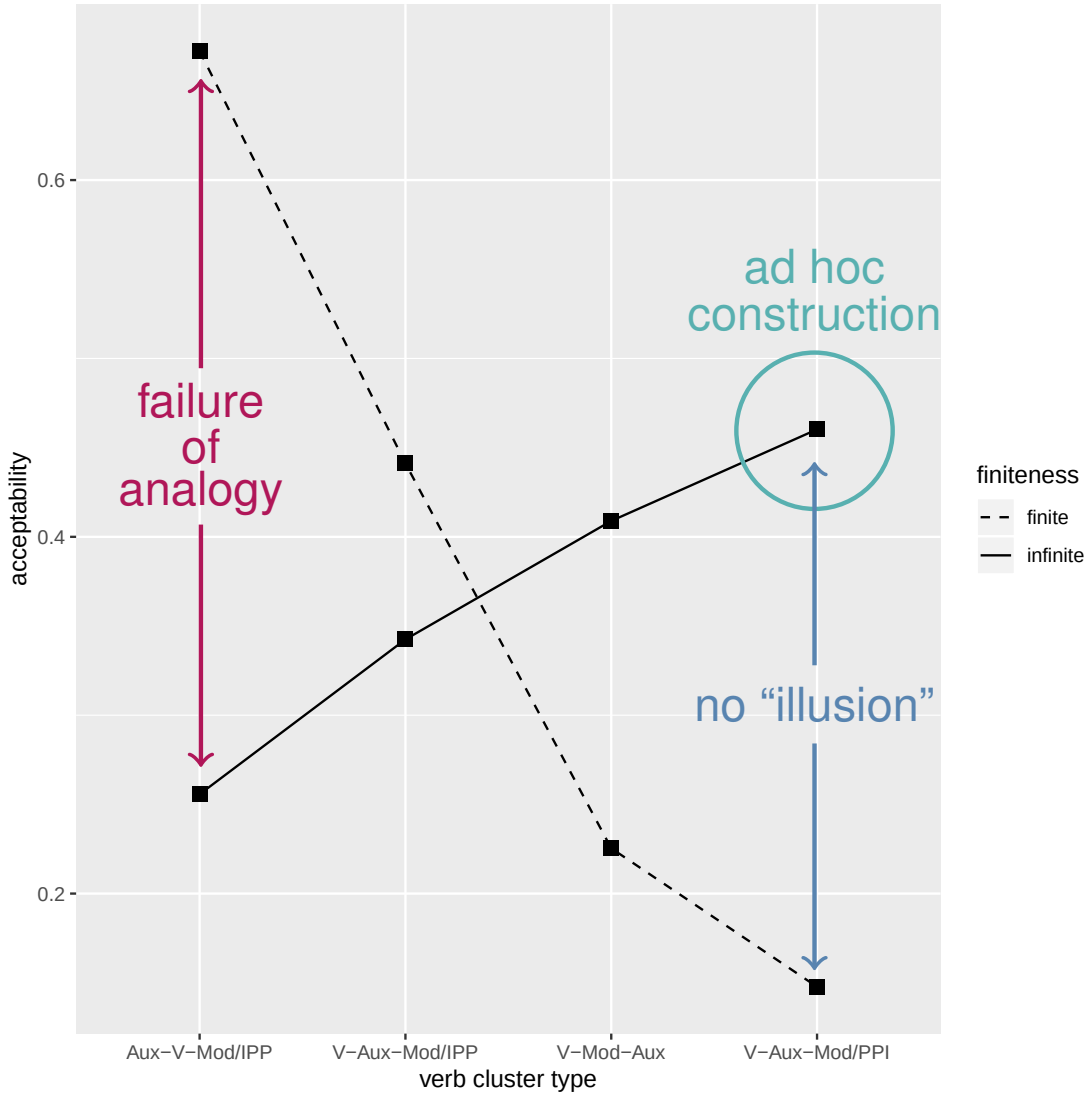
	Aux-V_{INF}-Mod_{INF}	V_{INF}-Aux-Mod_{INF}	V_{INF}-Mod_{PTCP}-Aux	V_{PTCP}-Aux-Mod_{INF}
Aux = finite	0.672	0.441	0.225	0.148
Aux = infinite	0.256	0.342	0.409	0.460
difference	–0.416	–0.099	+0.184	+0.312

Table 2: Mean acceptabilities for four verb cluster types with finite and infinite auxiliary (present perfect of modal verb constructions); 0 = out, 1 = perfect

% acceptable	**Category**	
0.90 – 1	✓	unmarked
0.60 – 0.80	?	slightly marked
0.20 – 0.50	??	marked
0 – 0.10	*	ungrammatical

Table 3: Correlation of acceptability and gradient grammaticality in Vogel (2019)

Figure 1: Mean acceptabilities of four (in)finite verb cluster types



Acceptability

- The preference for PPI in the infinitival versions is confirmed, as well as a strong rejection of PPI for finite verbal complexes.
  Contra Haider (2011): if high acceptability of PPI is due to local well-formedness of the sequence “participle auxiliary”, the (in)finiteness of the auxiliary shouldn’t matter
- The IPP structure that is preferred for finite complexes has the worst rating in infinitival complexes:
  ➔ failure of analogy
- contrasts between best and worst structures for finite complexes larger than for infinite complexes: indicates grammaticalisation only for finite complexes.
  All four infinitival variants in the range of markedness

Grammatical analysis

Modal perfect construction, two options:

Option A

category: VP
comp. 1: HABEN
comp. 2: [$_{VP}$ *modal*$_{INF}$]
comp. 3: [$_{VP}$ . . . V$_{INF}$]

Option B

category: VP
comp. 1: *haben*$_{FINITE}$
comp. 2: [$_{VP}$ *modal*$_{INF}$]
comp. 3: [$_{VP}$ . . . V$_{INF}$]

- Option A: three infinitives possible.
- Only 7 of the 143 items in the corpus were of this kind.
- Option B: 3-verb complexes with IPP only conventionalised for clusters with finite auxiliary.
  ➔ empirically more adequate.

➔ No construction rule for infinitival variants of the perfect of embedding modal verbs.

Grammatical analysis

(17) *. . . , niemals regiert haben zu müssen.*
never govern.PTCP have.INF to must-INF

- Speakers are looking for the optimal solution – given that there is no perfect solution.
- Use constructions that come close to the task, and
- apply some general heuristics:
  1. Auxiliary and participle are there in the PPI construction.
  2. IPP is also there.
  3. The embedded infinitive is – wrongly – realised as participle.

Grammatical analysis

(17) *. . . , niemals regiert haben zu müssen.*
never govern.PTCP have.INF to must-INF

- PPI construction: best option in comparison with all conceivable alternatives, . . .
- . . . if the constructions used in sentences like (17) are prioritised in the following way:
- infinitive construction $\succ$
  perfect construction (auxiliary + participle), modal perfect (IPP) $\succ$
  modal construction (infinitive of lexical verb)
- This hierarchy follows the direction of syntactic embedding in such sentences.

Invention as optimisation

component 1 highest verb = infinitive (infinitive construction)

component 2 auxiliary (perfect construction)

component 3 participle (perfect construction)

component 4 finite auxiliary (modal perfect construction)

component 5 infinitival modal verb (modal perfect construction)

component 6 infinitival complement verb (modal verb construction)

	cmp1	cmp2	cmp3	cmp4	cmp5	cmp6
Aux_{INF}-V_{INF}-Mod_{INF}/IPP	✓	✓	*!	*	✓	✓
V_{INF}-Aux_{INF}-Mod_{INF}/IPP	✓	✓	*!	*	✓	✓
V_{INF}-Mod_{PTCP}-Aux_{INF}	✓	✓	✓	*	*!	✓
☞ V_{PTCP}-Aux_{INF}-Mod_{INF}/IPP-PPI	✓	✓	✓	*	✓	*

Table 4: Optimality theoretic evaluation of the four candidate structures relative to the ad hoc construction based on 4 source constructions

Infinitival clusters without "*zu*"

- Vogel (2009): only *zu*-infinitival clusters inspected
- clusters without *zu* also occur, though even rarer (DeReKo):

(18) **312-IPP/PPI:**
Wie die ARD-Dopingredaktion am Sonntag berichtete, soll ein deutscher Eisschnellläufer und Olympia-Teilnehmer wiederholt sein Blut vom Erfurter Netzwerk um den seit Ende Februar inhaftierten Mediziner manipuliert haben lassen.

"As the ARD doping division reported on sunday, a German ice skater who participated in the olympic games is said to have his blood manipulated repeatedly by the network in Erfurt around the physician who has been arrested since the end of February." (Stuttgarter Nachrichten, March 24, 2019)

(19) **312-IPP:**
Gerüchten zufolge soll er sich allerdings in einer Klinik in Aspen behandeln haben lassen . . .

"According to rumours, though, he is said to have himself treated in a hospital in Aspen . . . "

(NEW08/APR.00350 NEWS, 24.04.2008, S. 22; Das Dossier Natascha)

Infinitival clusters without "*zu*"

- Patterns are very rare, in relevant numbers only with causative *lassen*.

. . . sources	312-IPP/PPI X = participle	312-IPP X = infinitive	X = indifferent
German . . .	26	7	–
Austrian . . .	6	13	2
Swiss . . .	3	–	–

Table 5: Distribution of *X haben lassen* items in the W-public corpus of the IDS Mannheim, June 2019

- **Germany vs. Austria** – odds ratio = 8.05: chance of PPI in such contexts in German newspapers 8 times higher than in Austrian newspapers (large effect)
- idea: 312-IPP in finite complexes more salient/less marked in Austria ➔ potential blocking of 312-IPP/PPI by 312-IPP for infinitives in Austria

Contents

Quick summary

- Empirical heuristics for grammatical gaps and ad hoc constructions:
  - *rarity* of configuration,
  - failure of *analogy*,
  - resolution by
    - *several* variants
    - with *unexpected* properties
    - but medium to high *acceptability*

Quick summary

- Grammatical inventions offer a window into mechanisms of general linguistic competence and their interaction:
  - blending of new constructions out of old ones (optimisation)
  - generalised conversational implicature;
  - analogy;
  - blocking

  Open task:
- Crucial to this kind of research is an estimation of the frequency of occurrence of a particular invention in an average speaker's input;
- as well as an estimation of the threshold frequency for grammaticalisation.

Bohn, Roger E. & James E. Short (2009). 'How Much Information? 2009 Report on American Consumers.' Global Information Industry Center. University of California, San Diego. `http://group47.com/HMI_2009_ConsumerReport_Dec9_2009.pdf`.

Featherston, Sam & Wolfgang Sternefeld (2003). 'The Interaction of Factors in Judgments of Reflexive Structures: Data from Object Coreference in German.' *In* Gunkel, Lutz, Gereon Müller, & Gisela Zifonun, eds, *Arbeiten zur Reflexivierung*, pp. 25–50. Tübingen: Niemeyer.

Grewendorf, Günther (1988). *Aspekte der deutschen Syntax: Eine Rektions-Bindungs-Analyse*. Tübingen: Niemeyer.

Haider, Hubert (2011). 'Grammatische Illusionen – Lokal wohlgeformt – global deviant.' *Zeitschrift für Sprachwissenschaft* **30**:223–257.

Jackendoff, Ray (1992). 'Mme Tussaud meets the Binding Theory.' *Natural Language and Linguistic Theory* **10**:1–31.

Keller, Rudi (1998). *A Theory of Linguistic Signs*. Oxford: Oxford University Press. Translated by Kimberley Duenwald.

Levinson, Stephen C. (2000). *Presumptive meanings*. Cambridge, Mass. [u.a.]: MIT Press.

Pollard, Carl & Ivan A. Sag (1992). 'Anaphors in English and the Scope of Binding Theory.' *Linguistic Inquiry* **23**:261–303.

Reiner, Tabea Elina (2015). 'An emerging future infinitive in present day German?' *Open Linguistics* **1**:503–518.

Reis, Marga (1976). 'Reflexivierung in deutschen AcI-Konstrukionen: Ein transformationsgrammatisches Dilemma.' *Papiere zur Linguistik* **9**:5–82.

— (2017). 'Grammatische Variation und realistische Grammatik.' *In* Konopka, Marek & Angelika Wöllstein, eds, *Grammatische Variation. Empirische Zugänge und theoretische Modellierung*, pp. 255–282. Berlin and Boston: de Gruyter.

Vogel, Ralf (2009). 'Skandal im Verbkomplex.' *Zeitschrift für Sprachwissenschaft* **28**(2):307–346.

— (2014). 'On doctors, patients and mirrors. Binding by object and complex reflexives in German.' *Linguistische Berichte* **239**:127–140.

— (2019). 'Grammatical taboos. An investigation on the impact of prescription in acceptability judgement experiments.' *Zeitschrift für Sprachwissenschaft* **38**(1):37–79.

Wurmbrand, Susi (2012). 'Parasitic participles in Germanic: Evidence for the theory of verb clusters.' *Taal & Tongval* **24**(1):129–156.


Thank You!