



## Univerbation and categorical fuzziness

What spelling variants of preposition-noun  
combinations reveal about linguistic  
categorization

Stefan Hartmann

hartmast@hhu.de



Data & script:  
<https://hartmast.github.io/zuhause/>



- Preliminaries: What spelling reveals about linguistic categorization
- Case study: *zuhause* – *zu Hause* – *zuhause*
  - Data and methods
  - Results
  - Discussion
- Conclusion

- Spelling often reflects knowledge about linguistic categories:
  - capitalization at the beginning of sentences → textual units
  - capitalization of nouns in German → word class categories
  - open, solid, and hyphenated spelling → words vs. syntactic unit (see Sanchez-Stockhammer 2018)
- Spelling variation can reveal uncertainty about categories, and provide cues to categorical fuzziness
  - e.g. spelling errors in English derivational suffixes depending in morphological boundary strength (Gahl & Plag 2019)

## Spelling variation and "doubtful cases"

- Instances of spelling variation as prime example of "doubtful cases" in the sense of Klein (2003)
- Klein (2003) speaks of a doubtful case when we are dealing with
  - two or more variants that are nearly identical in form,
  - which are equally acceptable in a given context,
  - but that competent speakers are unsure about when to use which variant and in what situations (see Schröder 2019: 250)
- Doubtful cases in the graphemic domain often relate to
  - open vs. solid (vs. hyphenated) spelling,
  - capitalization

# Spelling and categorization

## Spelling variation and "doubtful"



## Setzfehler

Ronald M. Filkas: Korrektorat, Lektorat, Redaktion, Text, Satz

SETZFEHLER

KONTAKT

LEISTUNGEN

MEDIENGALERIE

ÜBER MICH



### Top 100 Rechtschreibfehler - Richtige Schrei

SPRACHWISSEN

## Zuhause, zuhause oder zu Hause?

Veröffentlicht am 31. März 2020 von Ronald M. Filkas / 8 Kommentare

### Rechtschreibfehler - zu Hause

## mandlogs

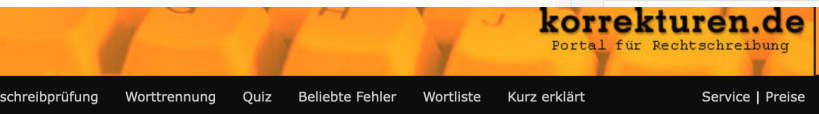
Deutschland, Europa und der Welt

ZEIT&HOBBY GESUNDHEIT HAUS&GARTEN KUNST&KULTUR

## oder „Zuhause“ – wie heißt es richtig?

, 2017 0

reibung fühlen sich nur wenige so richtig heimisch. Dabei erreichen, wenn man sich ein wenig mit ihr beschäftigt. ichtig „zu Hause“ oder „Zuhause“ fühlen, erklären wir heute deren Unterschied.



### Rechtschreibforum

ragen zur deutschen Rechtschreibung, nach Duden richtigen Schreibweise, zu Grammatik oder Kommasetzung, utung oder Synonymen sind Sie hier richtig. Bevor Sie eine Frage stellen, nutzen Sie bitte die Suchfunktionen.

Wortsuche

bo verwalten] [Suche] [Benutzerprofile] [Mein Profil] [Mobile Version]

bersicht] [Antwort schreiben] [Antworten lesen] [Vorheriger Beitrag]

ächster Beitrag]

Forumsuche

Ansicht: einzelne Beiträge

erarbeiten von Zuhause aus

## *zu*hause: Prescriptive perspective

- in standard German orthography, only two spelling variants are permitted:
  - *zu Hause*
  - *Zuhause*
- They clearly correspond to different usage contexts:
  - adverbial: *Ich habe meinen Schlüssel zu Hause vergessen* 'I've forgotten my key at home'
  - nominalized: *mein Zuhause* 'my home'

## *zu*hause: Descriptive perspective

- In actual usage, we find three different spelling variants:
  - open spelling: *zu Hause*
  - solid spelling, lowercase: *zuhause*
  - solid spelling, uppercase: *Zuhause*
- Each spelling variant is attested in virtually every possible context (apart from nominalized uses, where *Zuhause* seems clearly dominant)
- This is especially the case in non-standard written language, but also in newspaper texts:

## *zu*hause: Descriptive perspective

- Meine Mutter hat sowieso immer **Zuhause** am Computer arbeiten können.  
'My mother could always work from home (lit. at home) at the computer.' (ZEIT Magazin, 10.08.2017, DWDS)
- Der Traum vom modernen Familienleben - Mutter arbeitet, Vater bleibt **zu**hause.  
'The dream of modern family life: Mother is working, father stays at home.'  
(ZEIT, 11.11.2013, DWDS)
- Patienten müssen auf diese Weise nicht mehr ihr **zu Hause** verlassen. 'That way patients don't have to leave their home any more.' (spiegel.de, Corona-Korpus)

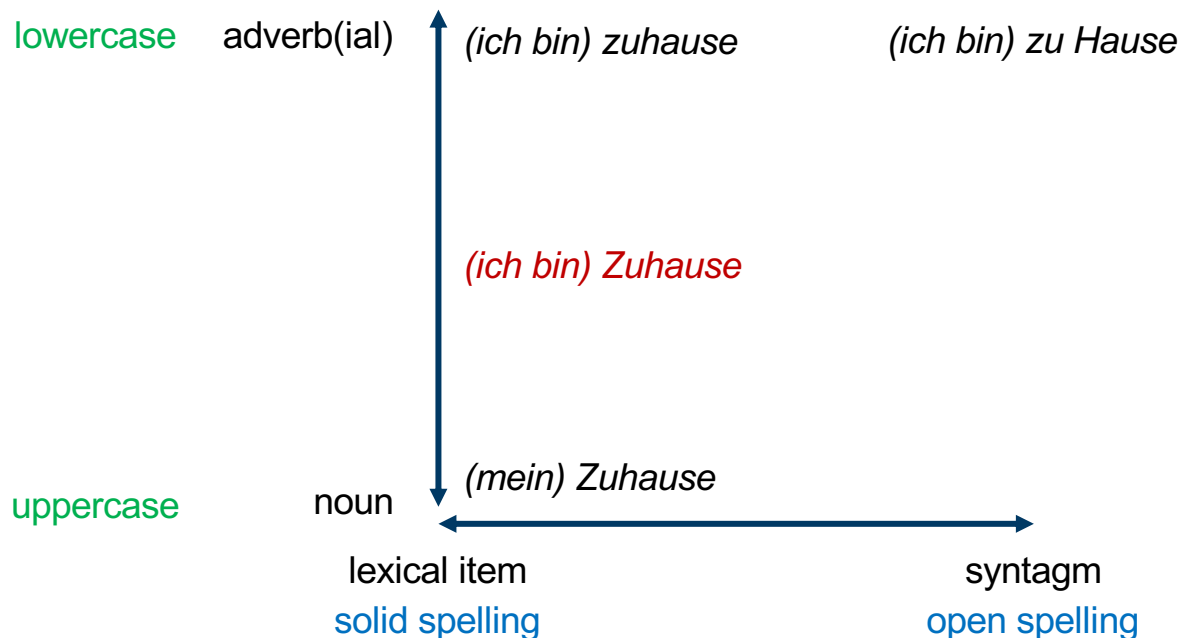


## *zuhause*: A quick diachronic outlook

- Solt yr dan eym kranckenn das heylig Sacrament oder die heiligenn ölung **zu hauß** tragenn/ (Bucer, Ayn schoener Dialogus, 1521, DTA)
- Die gepurt Christi war aber alßo gethan/ Als Maria seyne muter dem Joseph vertrawet war/ ehe sie mit eynander **zu hauß** sassen/ erfand sichs das sie schwanger war/ von dem heyligen geyst. (Luther, Das Newe Testament Deutzsch, 1522)

→ lexicalization / univerbation

## zuhaus: Spelling variants



## *zu*hause: usage contexts

- predicative:

*ich bin zu Hause, ich bleibe zu Hause*

- adverbial:

- *sie arbeitet zu Hause,*

## zuhouse: Database

- Corona corpus (Barbaresi 2020)
- c. 50 mio. tokens from c. 71.000 Covid-related texts
- advantage: variants of *zu Hause* can be expected to occur very often

Variant		Freq
zu Hause		11882
zuhouse		2682
Zuhause		1093

random sample of 2000 annotated for construction

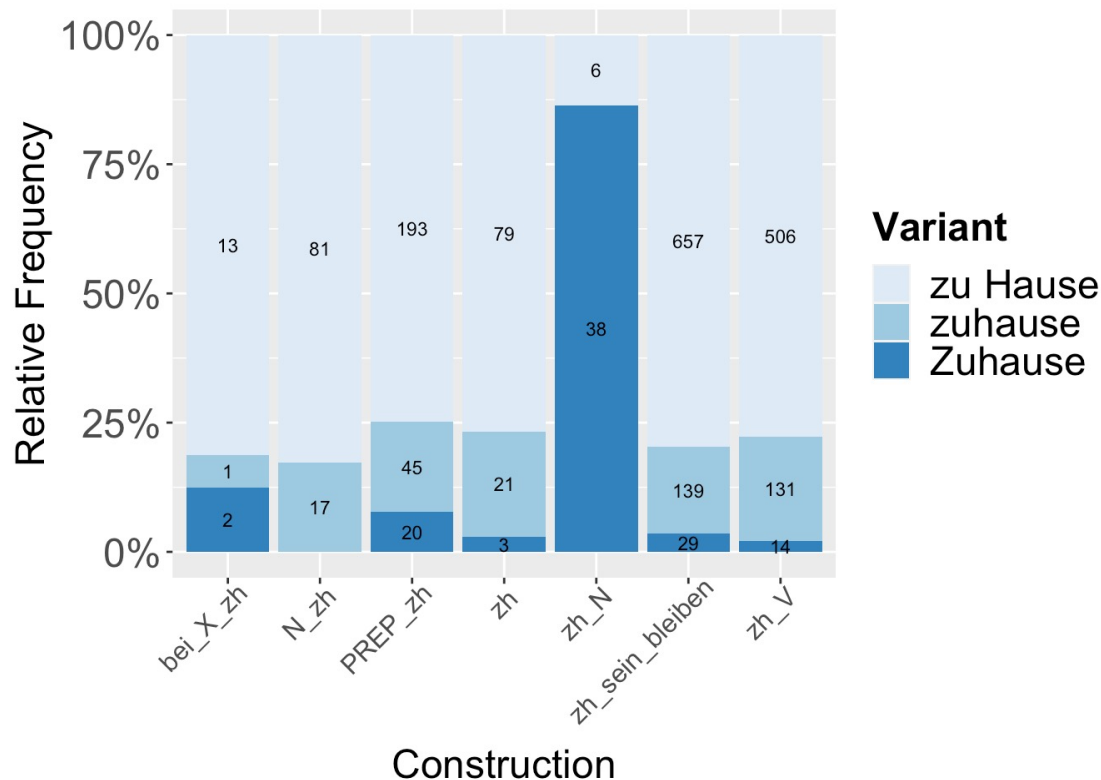
## *zu*house: annotation categories

- zh sein/bleiben: *ich bin/bleibe zu Hause* 'I am / stay at home'
- zh V: *ich arbeite zu Hause* 'I work at home'
- N zh: *die Kinderbetreuung zu Hause* 'childcare at home'
- bei X zh: *bei Hoppenstedts zu Hause* 'at the home of the H. family'
- PREP zh: *ich arbeite von zu Hause* 'I work [lit.] from at home'
- zh (N): *das Zuhause, mein Zuhause* 'the home, my home'
- zh: *Zu Hause ist es am schönsten* [lit.] 'At home it's the most joyful' / [roughly:] 'There's no place like home'

# Case study

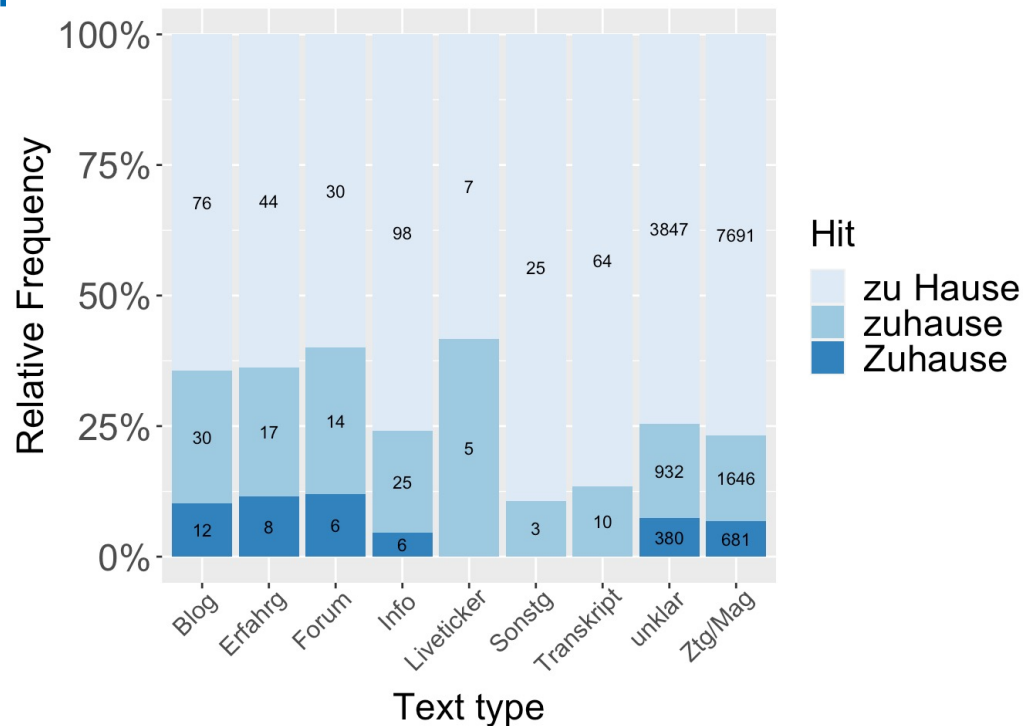
## zu Hause: Results

- Open spelling as default
- Solid spelling – lowercase across all contexts
- uppercase particularly prominent in PPs



## zu Hause: Results by text type

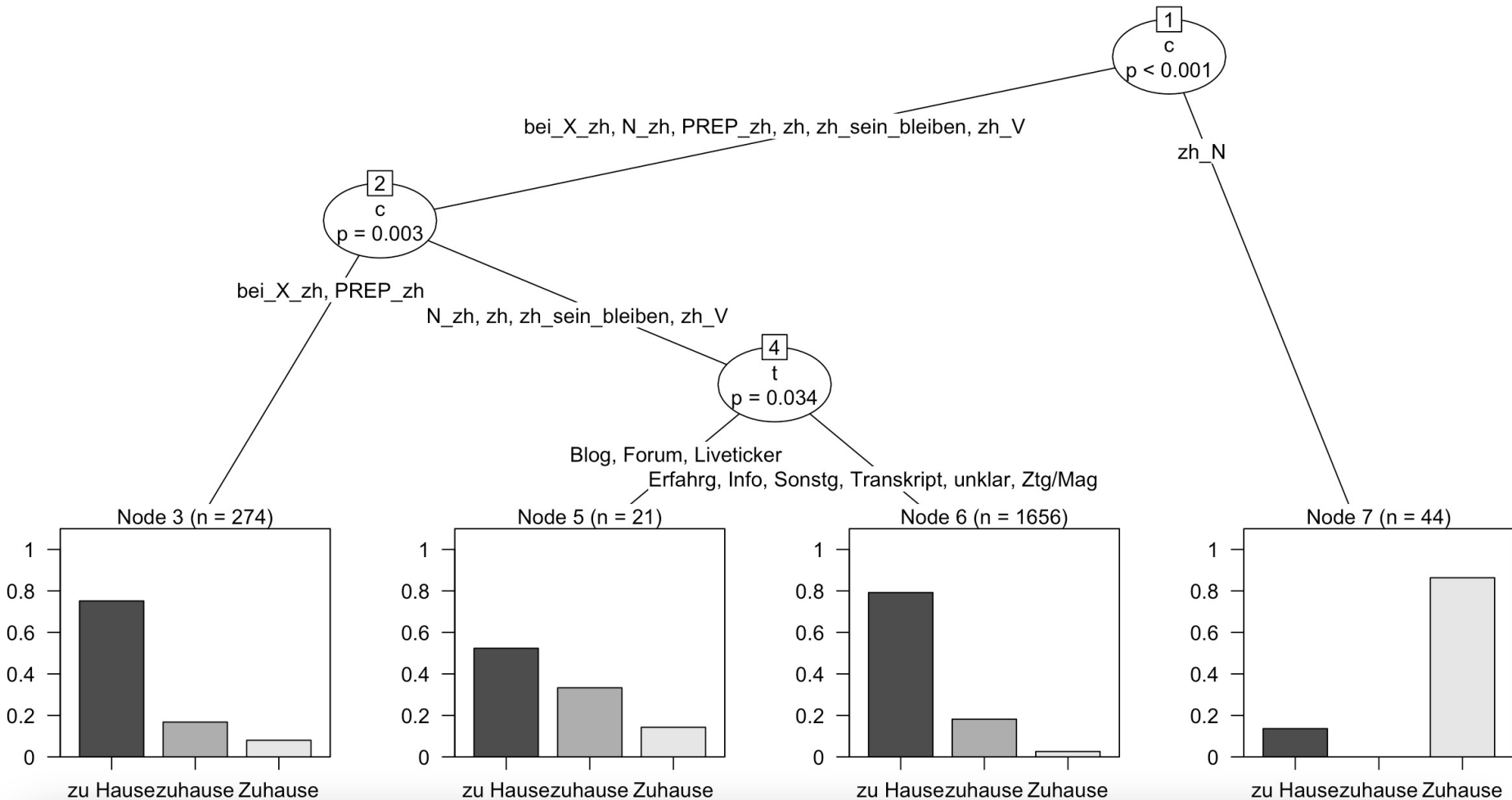
- non-standard spelling variants slightly more frequent in non-standard texts (duh!)
- but solid spelling occurs across all text types.



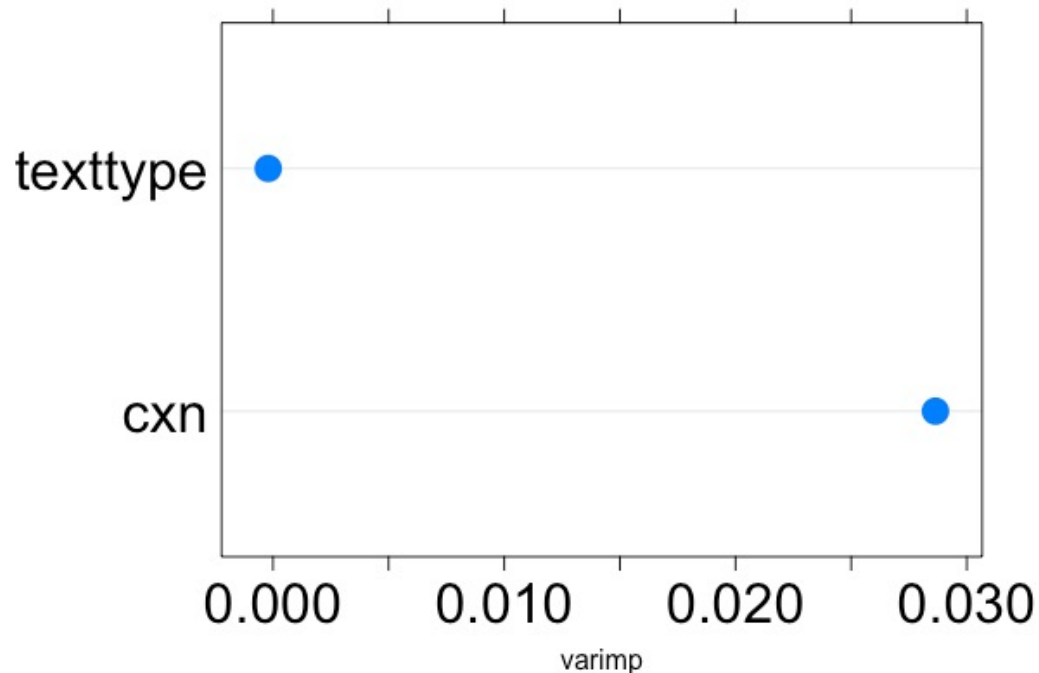
## *zuhause*: CART tree & random forest

- machine learning technique for identifying influential variables
- random forest: consists of many CART trees and averages over its predictions
- useful approach for "small  $n$  large  $p$ " problems (Tagliamonte & Baayen 2012, Levshina 2020), but some caveats have to be kept in mind (Gries 2019)





# Conditional variable importance



## Main results

- Solid spelling and capitalization occur predominantly in PPs
- PPs can be considered contexts in which the expression can be expected to be perceived as more "nominal":
  - (In German,) Prepositions overwhelmingly have NPs as complements;
  - in contexts like *von zu Hause, für zu Hause* etc., *zu Hause* can arguably be considered relatively "nouny"

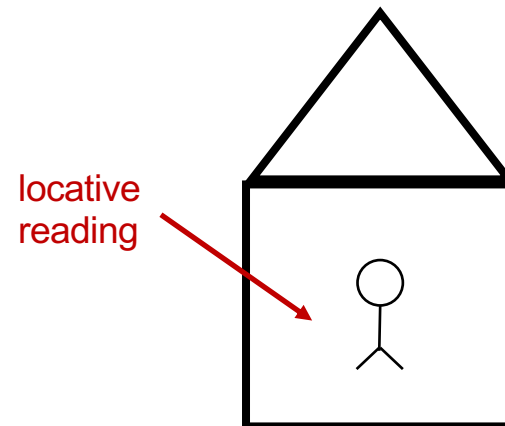
2207630	ART
992435	NN
677185	NE
562787	ADJA
399684	PP0SAT
231272	CARD
223308	PPER
210177	PDAT
204175	PIAT
155381	PRELS
129964	ADV
---	---

[https://kaskade.dwds.de/dstar/kern/dstar.perl?q=count%28%24p%3DAPPR%29+%23by%5B%24p%2B1%5D+%23desc\\_count&fmt=text&start=1&limit=100&ctx=8&debug=&export=](https://kaskade.dwds.de/dstar/kern/dstar.perl?q=count%28%24p%3DAPPR%29+%23by%5B%24p%2B1%5D+%23desc_count&fmt=text&start=1&limit=100&ctx=8&debug=&export=)

## "Nouniness"

- Cognitive Grammar (Langacker 1987, 1991, 2008):
  - A noun profiles a "thing"
  - A "thing" is any product of grouping and reification

*ich bin zu Hause* 'I am at home'



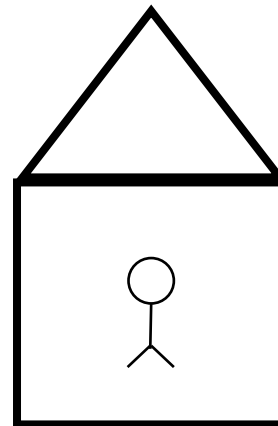
## "Nouniness"

- Cognitive Grammar (Langacker 1987, 1991, 2008):
  - A noun profiles a "thing"
  - A "thing" is any product of grouping and reification

*ich bin zu Hause* 'I am at home'

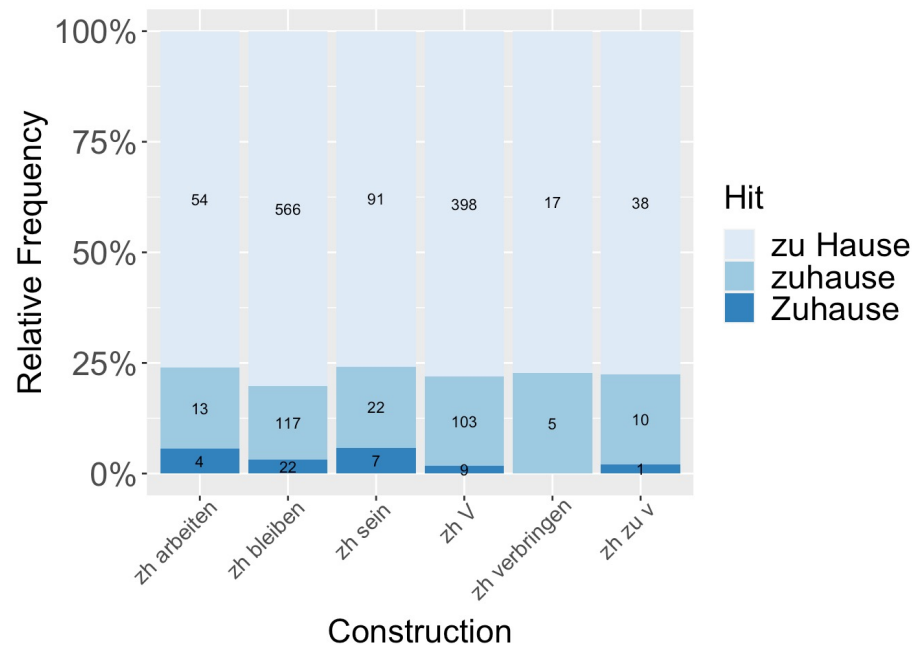
*mein Zuhause* 'my home'

*von zu Hause arbeiten* 'work from home'



## Frequency effects?

- it could be expected that solid spelling is more frequent in highly entrenched patterns like *zu bleiben/arbeiten*, but this does not seem to be the case:
- different spelling variants distributed quite consistently across different subcategories of *zh V*



## The categorical fuzziness of *zu Hause*

- *zu Hause* emerged as a PP but has become lexicalized to a considerable extent
- In current usage, it behaves like a noun/NP in some contexts and like an adverbial in others
- This is reflected in the different usage variants:
  - *zu Hause*, the variant sanctioned by the official spelling norm in all contexts except nominalizations, dominates;
  - solid spelling and uppercase spelling occur in specific contexts
  - uppercase spelling can potentially be explained by "nouniness"
  - no clear picture regarding the distribution of solid spelling

- Considerable amount of variation in the spelling of *zu Hause* / *zuhause* / *Zuhause*
- paradigm example of a "doubtful case" (Klein 2003)
- various hypotheses regarding the distribution of solid vs. open spelling and uppercase vs. lowercase spelling
- But: results are (at least partly) inconclusive
- Limitations:
  - only a small sample was annotated manually;
  - use of ad-hoc construction categories for annotation;
  - only one pattern taken into account – other, similar patterns (e.g. *infolge* / *in Folge*) could prove interesting as well.



- Gahl, Susanne & Ingo Plag. 2019. Spelling errors in English derivational suffixes reflect morphological boundary strength: A case study. *The Mental Lexicon* 14(1). 1–36.
- Gries, Stefan Th. 2019. On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvement. *Corpus Linguistics and Linguistic Theory*. doi:[10.1515/clit-2018-0078](https://doi.org/10.1515/clit-2018-0078).
- Hennig, Mathilde, Jan Georg Schneider, Ralf Osterwinter & Anja Steinhauer. 2021. *Duden - Sprachliche Zweifelsfälle: das Wörterbuch für richtiges und gutes Deutsch*. 9., überarbeitete und erweiterte Auflage. (Der Duden: in zwölf Bänden : das Standardwerk zur deutschen Sprache Band 9). Berlin: Dudenverlag.
- Klein, Wolf Peter. 2003. Sprachliche Zweifelsfälle als linguistischer Gegenstand: Zur Einführung in ein vergessenes Thema der Sprachwissenschaft. *Linguistik Online* 16. [http://www.linguistik-online.de/16\\_03/klein.html](http://www.linguistik-online.de/16_03/klein.html).
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Vol. 1: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar. Vol. 2: Descriptive Application*. Stanford: Stanford University Press.
- Langacker, Ronald W. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.
- Levshina, Natalia. 2020. Conditional Inference Trees and Random Forests. In Magali Paquot & Stefan Th. Gries (eds.), *A Practical Handbook of Corpus Linguistics*, 611–643. Cham: Springer International Publishing.
- Sanchez-Stockhammer, Christina. 2018. *English compounds and their spelling*. Cambridge: Cambridge University Press.
- Schröder, Daniela. 2019. Who and whom in contemporary American English: a doubtful case. In Eleonore Schmitt, Renata Szczepaniak & Annika Vieregge (eds.), *Sprachliche Zweifelsfälle: Definition, Erforschung, Implementierung*, 249–269. (Germanistische Linguistik 244-245/2019). Hildesheim: Georg Olms Verlag.
- Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24(02). 135–178.