

# Argument selection and alternations in German VP idioms

Christiane Fellbaum  
Princeton University  
and  
Berlin-Brandenburg Academy of  
Sciences

## Acknowledgments

### Work supported by

Alexander von Humboldt-Foundation (ZIP)  
ARDA/AQUAINT  
National Science Foundation

### Collaborators:

A. Firenze, E. Gehweiler, A. Geyken, A. Herold,  
I. Hoerer, C. Huemmer, F. Koerner, U. Kramer,  
K. Krell, G. Neumann, A. Sokirko,  
D. Stantcheva, E. Stathi, A. Urbschat

## Background: The Berlin project

- Corpus-based research on the grammar of German idioms
- Create a database of hundreds of German idioms as a resource for future research
- Lexicographic experiment:  
combine corpus data with linguistic annotation

## Why idioms?

- Idioms are frequent; not a marginal phenomenon
- Insufficient lexical resources for idioms
- Need for large-scale, empirical work
- Data so far show that most idioms are neither “long words” nor merely preferred co-occurrences
- How do VP idioms fit into the grammar?

## Motivation

No lexical resource at present lists syntactic, morphological, lexicographic properties of (German) collocations and idioms in a way that fully reflects their use in connection with a linguistic analysis

## Goals

Linguistically deep, (relatively) theory-neutral analysis  
Empirically based work: each linguistic phenomenon is linked to corpus examples  
Allows retrieval of examples for a given syntactic, lexical, semantic phenomenon

## Empirical basis for work:

- DWDS Corpus (Berlin-Brandenburg Academy of Sciences)
- 1 billion words of running text
- Subcorpus of 100 mio words is balanced across genres, covers entire 20th cty.  
<http://www.dwds.de>

## Main components

- digital knowledge base created via structured annotations
- corresponding sub-corpus of examples for each idiom

## Working Procedure

- Establish list with target idioms
- Formulate search queries
- Retrieve, sort hits
- Construct example corpora
- Linguistic analysis of idioms
- Link linguistic analyses to corresponding examples

## The Example Corpora

- For each target idiom, create a sub-corpus of examples
- Write search queries to generate candidate set of idiom tokens (using in-house search engine)
- Search queries include lexical and morphological elements (incl. semantically related words), Boolean operators

## Linguistic Search Engine

- Apply to 1 bio. word annotated corpus
- Avoid over- and undergeneration
- DDC: a custom-developed search engine
- Can deal with linguistically sophisticated queries

## Linguistic Search Engine

- Linguistic and Boolean query:

Example: `NEAR ((Tier) [SUB sin] 2)*  
animal NP sg`

Find all sentences with all morphological variants of „animal“ and a subcategory (cat, snail,...) and a singular noun within a distance of no more than 2 words

- Establish query bank; continually update this bank
- classification of queries

## Linguistic Annotation

- Manually classify retrieved examples: idiomatic vs. non-idiomatic (create potential training corpus for contexts for non-idiomatic strings)
- Idiomatic tokens are annotated via “template,” a kind of linguistic questionnaire
- Template info is stored in MySQL database
- Many types of linguistically motivated queries possible
- For each query, get examples and analyses (e.g., find all idioms with ditransitive verbs and a noun referring to an animal)

## Strawman/Classical view of idioms

- Idioms are long words
- Standard example: *kick the bucket*
- Not available for syntactic operations
- Non-compositional
- Semantically opaque constituents/non-referential

## Idioms are not just (long) words!

- different semantic/ontological properties
- varying degrees of compositionality, semantic opacity
- varying degrees of syntactic freedom
- much lexical variation

## Semantic/ontological status of VP idioms

- In some cases, the VP expresses a simple concept (disregarding style/connotations):

*kick the bucket/ins Gras beissen* = “die”  
*take a powder/sich aus dem Staub machen* = “disappear, vanish”

These are the exceptions, not the rule!

- not only structurally complex: V + complement, but also
- semantically complex: non-atomic concepts
- “pre-packaged” concepts that are would otherwise have to freely composed in the non-idiomatic language

“Ready-made, pre-fabricated” messages

- save effort necessary to compose messages
- often carefully nuanced (many idioms have synonyms differing in their shade of judgment, emotional strength, etc.)
- express emotions, interpersonal relations, social interaction, judgments of others,..)

## Many concepts expressed by VP idioms don't fit into the lexicon

- Many idioms cannot be integrated into a semantic network (like WordNet)
- cannot be connected to other word-concept pairings with existing relations
- (e.g., what are synonyms/superordinates of *den Sack schlagen und den Esel meinen? sich einen Korb holen?*)
- differentiae not the same (greater semantic distance between super- and subordinate--generic and more specific--concepts)
- e.g., *aus der Not eine Tugend machen -> improvisieren?*
- kein Blatt vor den Mund nehmen -> sprechen?*

## Modals, NPIs, passives, states,...

- die Puppen tanzen lassen
- sich sein Lehrgeld zurueckgeben lassen koennen
- sich einen Korb holen (be rebuffed/rejected)
- etw. auf dem Kerbholz haben (be guilty)
- in den Tuschkasten gefallen sein (be haevily made up)
- fuer X keinen Pfennig/Pfifferling geben (not give a dime for)
- Y nicht ueber die Lippen bringen (be unable to utter Y)

## How much variability?

## Grammatical Operations

- Corpus shows: syntactic operations are more widespread than the classical view of idioms suggests
- No clear relation to semantic opacity

Even semantically opaque idioms are subject to operations inside the VP:

*Er biss ins texanische Gras*

He bit into the Texan grass

He bit the Texan dust: “he died in Texas”

*Gras* does not refer; adj has scope over entire VP

*das Gras, in das er gebissen hat*  
the grass into which he bit (relativization)

*er beisst in das Gras der endlosen Ebenen*  
he bites into the grass of the endless steppes  
*Russlands*  
of Russia (genitive appositive)

Such data contradict Nunberg et al.'s  
(1994) claim that semantic opacity of the  
constituents precludes syntactic flexibility  
(conversely, semantic transparency is not  
always reflected in syntactic freedom)

Cf. also Dobrovolskij (1999)

### Regular grammatical processes Aspect:

den Stein ins Rollen bringen (causative)  
get the ball (lit. stone) rolling

der Stein kommt ins Rollen (inchoative)  
the ball starts to roll

der Stein ist im Rollen (resultant state)  
the ball is rolling

(cf. also *in Harnisch, in Rage,...*)

### As with lexical variation, limitations seem arbitrary:

die Kuh vom Eis bringen

\*die Kuh kommt vom Eis

\*die Kuh ist vom Eis

### Lexical variation

- Idiomatic readings are stable across a wide range of lexical variations

### Corpus shows significant lexical variation

--highly productive

*jmd. nimmt jmdm. die Butter vom*

*s.b. takes from s.b. the butter from the*

*Brot*

bread

“s.b. gets the better of s.b.”

## Lexical variations

*jmd. nimmt jmdm. die Margarine/Wurst vom Brot*

s.b. takes the margarine/sausage from s.b.'s sandwich

## Productive, ad-hoc variations:

*jmd. nimmt jmdm. die Argumentbutter vom Wahlkampfbrot*

“s.b. takes s.b.’s argument butter from his election campaign sandwich”

## Lexical productivity

- frequent substitution of synonyms, antonyms
- adjectival modification, compounding

sich in die Wolle kriegen/bekommen  
each other into the wool get/become  
“get into a conflict/adversarial situation”

sich in der Wolle liegen/haben  
each other in the wool lie/have  
“be in a conflict/adversarial situation”

- But lexical variation, even synonym substitution, is limited (not clear whether systematically/how):

*jn. uebers Ohr schlagen/hauen*  
s.b. over the ear hit/beat (cheat)

*jmd. aus den Stiefeln/vom Hocker hauen/\*schlagen*  
s.b. out of his boots/off the stool hit/\*knock  
(knock s.b. off his socks)

Not related to frequency!

## So what makes an idiom?

- With lexical variation, how much does the syntax contribute to the idiomatic meaning?
- Vice versa, with syntactic behavior approximating that of the free language, how much lexical fixedness is required?

Some kind of trade-off between lexical and syntactic fixedness

Big question:

What are (lexical, semantic) necessary and sufficient conditions for idiomatic readings?

(Varies across idioms/idiom classes)

## Chinese cheng-yu

- Idioms expressed by four characters
- Allow substitutions (but limited to one pair)
- Some syntactic freedom

## VP idioms V + complement (NP,PP,...)

## Idiomatic reading requires complement

jmdm. schlaegt die Stunde

for s.b. tolls the hour

the bell tolls for s.b.

“s.b.’s time has come/s.b. has to die”

cf. die Stunde schlaegt: literal reading only

vom Regen in die Traufe kommen  
from rain in the gutter come  
(come) from the frying pan into the fire  
“move from a bad situation into one that’s at least equally bad”

cf. literal reading only:  
(?) vom Regen kommen  
in die Traufe kommen

## Many idioms have rich argument structure

## Argument augmentation

- Benefactives
- Instruments
- Resultatives
- Obligatory PP adjuncts

## Benefactives (and malefactives)

Alternation: DO-PP (headed by *fuer*)

*fuer* jn. eine Bresche schlagen  
ein gutes Wort *fuer* jn. einlagen

jmdm. den Stuhl vor die Tuer stellen  
jmdm. den Steigbuegel halten  
jmdm. die Hoelle heiss machen

## Benefactives

- optional argument in non-idiomatic language
- PP-alternant has few distributional restrictions (just about any event can be construed as a benefit for s.b.)
- DO-alternant is restricted (semantics)
- Restrictions are not well understood (but there are fewer in German, which has “ethical dative”)

## Confusion Benefactive-Dative Alternation (Beneficiary/Recipient argument)

--some overlap in the verb classes  
*read a poem to/for s.b.*  
*write a letter to/for somebody*

Distinction is neutralized in DO alternant:  
*read Y a poem*  
*write y a letter*

## Same ambiguity in German

--jm. einen Brief schreiben  
--einen Brief *fuer/an* jn. schreiben

## Benefactives in the literal language

DO alternant is confined to a few semantic classes:  
created, prepared entities; performance verbs  
In these cases, Benefactive alternation is fairly regular because it's class-based.

Not understood: Benefactive with “symbolic action”  
(Green 1973):  
*Kill me a chicken*



## Benefactive Idioms

- Idiomatic reading is lost without Beneficiary argument:

sich (k)einen Reim auf etwas machen koennen  
for o.s (no) rhyme for s.th. make can  
be unable to make sense of  
die Stunde schlaegt jmdm.  
the hour tolls for s.b.  
the bell tolls for s.b.

## Idioms with a benefactive

Verbs in the idioms are those that do not typically select for a Benefactive  
“symbolic actions” (Green 1973)

- Beneficiaries in idioms are not present/future Recipient
- no created, obtained, prepared entities;
- no performances

- *machen, holen* (make, get)  
don't receive literal interpretation as creation, obtaining verbs

The benefit is the entire event

## Benefactive alternation in idioms

## Benefactives: alternating

jm./fuer jn. das Leben zur Hoelle machen  
“make life hell for s.b.”

jm./fuer jn. die Daumen druecken  
“keep o's fingers crossed for s.b./s.th.”

jm./fuer jn. die Hoelle heiss machen  
lit.: make hell hot for s.b.

“give s.b. a hard time”

## Non-alternating: DO alternant only

jm. einen Strich durch die Rechnung machen  
\*fuer jn. einen Strich...

jm. eine lange Nase machen  
\*fuer jn. eine lange Nase machen

jm. schoene Augen machen  
\*fuer jn. schoene Augen machen

jm. den Prozess machen  
\*fuer jn. den Prozess machen

jm./\*fuer jn. einen Streich spielen  
jm./\*fuer jn. ein Bein/eine Falle stellen

### Non-alternating: PP alternant only

fuer jn./\*jm. ein Herz haben  
fuer jn./\*jm. eine Bresche schlagen  
fuer jn./\*jm. sein Leben in die Schanze schlagen

(note: these are not “instead-of” readings  
cf. fuer jn. die Arbeit machen)

### Corpus shows:

- Most idioms require DO alternant
- A few allow alternation DO-PP
- Few require PP alternant

### How to account for alternation patterns?

Not based on verb class  
but on the meaning of the entire idiom  
Explanation analogous to one given for  
Dative Alternation:  
DO alternant: benefit is bestowed  
PP alternant: potential benefit  
(cf. Dative Shift alternation: cause-to-have  
vs. cause-to-go-to)

PP alternant does not necessarily entail benefit:  
ein gutes Wort fuer jn. einlagen  
“put in a good word for s.b.”  
eine Bresche fuer jn. schlagen  
“go to bat for s.b.”  
sein Leben fuer jn. in die Schanze schlagen  
“risk one’s life for s.b.”  
sich fuer jn. ins Zeug legen  
“work hard for s.b.’s sake/defend s.b.”

DO alternant: immediate (negative) benefit:

jm. eine lange Nase machen  
jm. schoene Augen machen  
jm. einen Streich spielen

## (Cautious) conclusion

Idioms seem to have appropriated the Benefactive, a relatively restricted construction

Restrictions on the distribution of Benefactive for “symbolic actions” (Green) are relaxed for idioms

Frame carries meaning

Idioms are often said to have “emotional” flavor  
express interpersonal relations, feelings, judgments

## Argument reduction in VP idioms

- transfer verbs without Recipient/Goal
- transitive verbs used intransitively
- verbs of movement: fahren, gehen, laufen (no LOC)
  
- Complements are mostly bare NP or definite NP (metaphor)

## Argument reduction idiom groups with idiosyncratic syntax

- ditransitives reduced to two arguments
- No third argument possible, even as a free slot:
  - Vollgas/Gummi geben
  - das Handtuch werfen (cf. Eng. throw **in** the towel)
  - Obacht geben
  - Ruhe geben
  - sein Letztes geben
  - Flagge/Rueckgrat/seine Borsten zeigen

## Motion verbs without Path, Goal

- Spiessruten laufen
- Amok laufen
- Gefahr laufen

- stehen + NP
- no LOC or Manner as in literal language
  - Pate/Modell stehen (serve as)
  - Spalier/Schlange stehen (form)
  - seinen Mann stehen (behave)

## More cautious conclusions

Idioms make use of specific valence patterns

Idioms make use of alternations to encode “pre-processed” messages

Semantic transparency is not relevant



Thank you!