A Survey of
Idiomatic Preposition-Noun-Verb Triples
on Token Level

Fabienne Fritzinger, Marion Weller and Ulrich Heid

University of Stuttgart
Institute for Natural Language Processing
– Computational Linguistics –
Azenbergstr. 12
D 70174 Stuttgart
[fritzife, wellermn, heid]@ims.uni-stuttgart.de
Motivation

In previous work, we focused on the identification of idiomatic multiword expression (MWE) **types**, like e.g.:

| MWE                          | Idiomatic meaning          | Literal meaning                                |
|------------------------------|-----------------------------|-------------------------------------------------|
| *in Leben rufen*             | to initiate                 | “to call to life”                               |
| *unter Teppich kehren*       | to hide/conceal             | “to sweep under carpet”                         |
| *auf Kopf stellen*           | to turn sth. inside out     | “to place sth. on head”                         |
| *(sich) aus Staub machen*    | to leave                    | “to make oneself out of the dust”               |
Motivation

In previous work, we focused on the identification of idiomatic multiword expression (MWE) types, like e.g.:

| MWE                                      | Idiomatic meaning                              | Literal meaning                             |
|------------------------------------------|------------------------------------------------|---------------------------------------------|
| *in Leben rufen*                         | to initiate                                    | “to call to life”                           |
| *unter Teppich kehren*                   | to hide/conceal                                | “to sweep under carpet”                     |
| *auf Kopf stellen*                       | to turn sth. inside out                        | “to place sth. on head”                     |
| *(sich) aus Staub machen*                | to leave                                       | “to make oneself out of the dust”           |

... but if we have a closer look at them, we see that
Motivation

In previous work, we focused on the identification of idiomatic multiword expression (MWE) types, like e.g.:

| MWE                        | Idiomatic meaning       | Literal meaning                  |
|---------------------------|-------------------------|----------------------------------|
| *in Leben rufen*          | to initiate             | “to call to life”                |
| *unter Teppich kehren*    | to hide/conceal         | “to sweep under carpet”          |
| *auf Kopf stellen*        | to turn sth. inside out | “to place sth. on head”          |
| *(sich) aus Staub machen* | to leave                | “to make oneself out of the dust”|

... but if we have a closer look at them, we see that → some of them can only have an idiomatic meaning
Motivation

In previous work, we focused on the identification of idiomatic multiword expression (MWE) types, like e.g.:

| MWE                  | Idiomatic meaning                  | Literal meaning                               |
|----------------------|------------------------------------|-----------------------------------------------|
| *in Leben rufen*     | to initiate                        | “to call to life”                             |
| *unter Teppich kehren* | to hide/conceal                   | “to sweep under carpet”                       |
| *auf Kopf stellen*   | to turn sth. inside out            | “to place sth. on head”                       |
| *(sich) aus Staub machen* | to leave                           | “to make oneself out of the dust”            |

... but if we have a closer look at them, we see that
→ some of them can only have an idiomatic meaning
→ while others could also have a literal meaning (theoretically)
Motivation

In previous work, we focused on the identification of idiomatic multiword expression (MWE) types, like e.g.:

| MWE                  | Idiomatic meaning | Literal meaning                                      |
|----------------------|-------------------|------------------------------------------------------|
| *in Leben rufen*     | to initiate       | “to call to life”                                    |
| *unter Teppich kehren* | to hide/conceal  | “to sweep under carpet”                              |
| *auf Kopf stellen*   | to turn sth. inside out | “to place sth. on head”                            |
| *(sich) aus Staub machen* | to leave         | “to make oneself out of the dust”                    |

→ Tools to identify idiomatic MWEs shall be aware of the actual meaning
In previous work, we focused on the identification of idiomatic multiword expression (MWE) types, like e.g.:

| MWE                     | Idiomatic meaning | Literal meaning                  |
|-------------------------|-------------------|----------------------------------|
| in Leben rufen          | to initiate       | “to call to life”                |
| unter Teppich kehren   | to hide/conceal   | “to sweep under carpet”          |
| auf Kopf stellen       | to turn sth. inside out | “to place sth. on head”       |
| (sich) aus Staub machen | to leave          | “to make oneself out of the dust” |

→ Tools to identify idiomatic MWEs shall be aware of the actual meaning
→ We do not present a method of how to deal with this!
Motivation

In previous work, we focused on the identification of idiomatic multiword expression (\textsc{Mwe}) types, like e.g.:

| \textsc{Mwe}                      | Idiomatic meaning       | Literal meaning                                      |
|----------------------------------|-------------------------|------------------------------------------------------|
| \textit{in Leben rufen}          | to initiate             | “to call to life”                                     |
| \textit{unter Teppich kehren}    | to hide/conceal         | “to sweep under carpet”                               |
| \textit{auf Kopf stellen}        | to turn sth. inside out | “to place sth. on head”                               |
| \textsc{(sich) aus Staub machen} | to leave                | “to make oneself out of the dust”                     |

→ Tools to identify idiomatic \textsc{Mwe}s shall be aware of the actual meaning
→ We do \textbf{not} present a method of how to deal with this!
→ \textbf{Instead}, we give an impression of the quantitative dimension of the problem:
Motivation

In previous work, we focused on the identification of idiomatic multiword expression (MWE) types, like e.g.:

| MWE                  | Idiomatic meaning           | Literal meaning                          |
|----------------------|------------------------------|------------------------------------------|
| in Leben rufen       | to initiate                  | “to call to life”                        |
| unter Teppich kehren| to hide/conceal              | “to sweep under carpet”                   |
| auf Kopf stellen    | to turn sth. inside out      | “to place sth. on head”                   |
| (sich) aus Staub machen | to leave                    | “to make oneself out of the dust”         |

→ Tools to identify idiomatic MWEs shall be aware of the actual meaning
→ We do **not** present a method of how to deal with this!
→ **Instead**, we give an impression of the quantitative dimension of the problem:

  we manually annotated a huge dataset
Preprocessing
Extraction of MWEs to be annotated
Preprocessing
Extraction of MWEs to be annotated

Corpus → Extract PNVs → List of PNVs
Preprocessing

Extraction of \textit{MWEs} to be annotated
Preprocessing

Extraction of **MWEs** to be annotated

- Corpus
- Extract PNVs
- List of PNVs
- Duden Lookup
  - only literal
Preprocessing

Extraction of MWEs to be annotated

Corpus → Extract PNVs → List of PNVs → Duden Lookup

only idiomatic

only literal
Preprocessing
Extraction of MWEs to be annotated

Corpus → Extract PNVs → List of PNVs → Duden Lookup →
only literal
only idiomatic
idiomatic&literal
Preprocessing
Extraction of MWEs to be annotated

Corpus → Extract PNVs → List of PNVs → Duden Lookup

- only literal
- only idiomatic
- idiomatic & literal
Preprocessing
Data and Tools

Data:

| corpus                        | size  | years   |
|-------------------------------|-------|---------|
| *Frankfurter Allgemeine Zeitung* | 70Mio | 97/98   |
| *EUROPARL*                    | 35Mio | 96-06   |

→ Assumption: literal instances in newspaper, but more rarely in *EUROPARL*
Preprocessing
Data and Tools

Data:

| corpus                          | size  | years |
|---------------------------------|-------|-------|
| *Frankfurter Allgemeine Zeitung* | 70Mio | 97/98 |
| *EUROPARL*                      | 35Mio | 96-06 |

→ Assumption: literal instances in newspaper, but more rarely in EUROPARL

Tools:

- **Fspar**, German dependency parser
- **PERL** scripts
Thus, the rumour is still to be dealt with.
Thus, the rumour is still to be dealt with.
Thus, the rumour is still to be dealt with.
Thus, the rumour is still to be dealt with.
Duden Lookup

Corpus → Extract PNVs → List of PNVs → Duden Lookup

- only literal
- only idiomatic
- idiomatic&literal

Idiomatic PNVs amongst 1,000 most frequent:
- Corpus
- Europarl
- Total 196
Duden Lookup

Idiomatic PNVs amongst 1,000 most frequent:

| Corpus | in *Duden* |
|--------|------------|
| FAZ    | 155        |
| EUROPARL | 108       |
| Total  | 196        |
Duden Lookup

Idiomatic PNVs amongst 1,000 most frequent:

| Corpus   | in Duden | only idiom. | idiom.&lit. |
|----------|----------|-------------|-------------|
| FAZ      | 155      | 86          | 69          |
| EUROPARL | 108      | 73          | 35          |
| Total    | 196      | 119         | 77          |

Ins Leben rufen ("to call to life", to create sth.)
auf Schlauch stehen ("to stand on a hose", to have a mental block)
Idiomatic PNVs amongst 1,000 most frequent:

| Corpus   | in *Duden* | only idiom. | idiom.&lit. |
|----------|------------|-------------|-------------|
| Faz      | 155        | 86          | 69          |
| Europarl | 108        | 73          | 35          |
| Total    | 196        | 119         | 77          |

*ins Leben rufen* ("to call to life", to create sth.)
Duden Lookup

Idiomatic PNVs amongst 1,000 most frequent:

| Corpus   | in *Duden* | only idiom. | idiom.&lit. |
|----------|------------|-------------|-------------|
| FAZ      | 155        | 86          | 69          |
| EUROPARL | 108        | 73          | 35          |
| Total    | 196        | 119         | 77          |

*ins Leben rufen* ("to call to life", to create sth.)
*auf Schlauch stehen* ("to stand on a hose", to have a mental block)
Manual Annotation

- 100 sentences for each PNV-Triple, randomly extracted
- annotated by two German native speaker (independently)
  → agreement in 6,550 of 6,690 cases of FAZ (97.9%)
  → remaining 140 discussed with a third native speaker

| corpus    | all  |
|-----------|------|
| FAZ       | 6,690|
| EUROPARL  | 3,050|
| total     | 9,740|
Manual Annotation

- 100 sentences for each $P_{NV}$-Triple, randomly extracted
- annotated by two German native speaker (independently)
  → agreement in 6,550 of 6,690 cases of FAZ (97.9%)
  → remaining 140 discussed with a third native speaker
- annotated categories: idiomatic (I), literal (L), ambiguous (A), parsing/extraction error (X).

| corpus     | all   | I   | L   | A   | X   |
|------------|-------|-----|-----|-----|-----|
| FAZ        | 6,690 | 6,176 | 345 | 75  | 94  |
| Europarl   | 3,050 | 2,937 | 31  | 14  | 68  |
| total      | 9,740 | 9,113 | 376 | 89  | 162 |
## Selected Examples

| PNV                                      | FAZ   | Europarl |
|------------------------------------------|-------|----------|
| **auf Seite stehen**                     | 16 75 3 | 12 86 1  |
| to stand on (so.’s) side                 |       |          |
| **zu Einsatz kommen**                    | 1 97 1 | 1 98 0   |
| to come into operation                   |       |          |
| **zu Fall bringen**                      | 17 82 1 | 0 47 0   |
| to bring down sth./sb.                   |       |          |
Selected Examples

| PNV                              | FAZ | Europarl |
|----------------------------------|-----|----------|
|                                   | L   | I        | A   |
| auf Seite stehen                  | 16  | 75       | 3   |
| to stand on (so.’s) side         |     |          |     |
| zu Einsatz kommen                 | 1   | 97       | 1   |
| to come into operation           |     |          |     |
| zu Fall bringen                  | 17  | 82       | 1   |
| to bring down sth./sb.           |     |          |     |

Literal instance (FAZ):

*Emmerling hatte im Strafraum Benedyk zu Fall gebracht.*

“Emmerling had brought down Benedyk in the penalty area “
## Selected Examples

| PNV                                      | Faz  | Europarl |
|------------------------------------------|------|----------|
| auf Seite stehen                         | 16   |          |
| to stand on (so.’s) side                 | 75   |          |
| zu Einsatz kommen                        | 1    |          |
| to come into operation                   | 97   |          |
| zu Fall bringen                          | 17   |          |
| to bring down sth./sb.                   | 82   |          |

Literal instance (Faz):

*Emmerling hatte im Strafraum Benedyk zu Fall gebracht.*

“Emmerling had brought down Benedyk in the penalty area “

Idiomatic instance (Ep):

*Der Versuch der Tabaklobby, dieses Gesetz zu Fall zu bringen, ist absurd.*

”The attempt of the tabacco lobby to kill the law is absurd.*
Distribution of Morpho-Syntactic Features

Example taken from FAZ

|                  | #P_NVs | DET  | FUS | NUM |
|------------------|--------|------|-----|-----|
|                  |        | no   |     |     |
| literal          | 33     | 2    | 19  | 28  |
| idiomatic        | 53     | 24   | 53  | 53  |
| ambiguous        | 2      | 1    | 53  | 2   |
|                  |        | def  |     | pl  |
| literal          |        | 2    | 14  | 5  |
| idiomatic        |        | 53   | 0   | 0   |
| ambiguous        |        | 2    | 0   | 0   |
|                  |        | quant |     |     |
| literal          |        | 1    | 28  |     |
| idiomatic        |        | 0    | 53  |     |
| ambiguous        |        | 0    | 2   |     |

*in Raum stehen*
Distribution of Morpho-Syntactic Features

Example taken from FAZ

|                         | #PNVs | DET | FUS | NUM |
|-------------------------|-------|-----|-----|-----|
|                         |       | no  | +   | sg  |
| literal                 | 33    | 2   | 19  | 28  |
| idiomatic               | 53    | 0   | 53  | 53  |
| ambiguous               | 2     | 0   | 2   | 2   |
|                         |       | def | -   | pl  |
| literal                 | 24    | 14  | 53  | 53  |
| idiomatic               | 53    | 0   | 53  | 0   |
| ambiguous               | 2     | 0   | 2   | 0   |

**Literal use**

*In den besseren Räumen standen Öfen, an denen er sich wärmte.*

“In the superior rooms stood stoves, at which he himself warmed.”

*In the superior rooms were stoves at which he warmed himself.*
Distribution of Morpho-Syntactic Features
Example taken from FAZ

|                   | #PNVs | DET | FUS | NUM |
|-------------------|-------|-----|-----|-----|
|                   |       | no  | def | quant | + | - | sg | pl |
| literal           | 33    | 2   | 24  | 1    | 19 | 14 | 28 | 5  |
| idiomatic         | 53    | 0   | 53  | 0    | 53 | 0  | 53 | 0  |
| ambiguous         | 2     | 0   | 2   | 0    | 2  | 0  | 2  | 0  |

*Idiomatic use*

_Widersprüche stehen ungelöst im Raum._
“Contradictions stand unresolved in the room”
There are unresolved contradictions to be dealt with.
Distribution of Morpho-Syntactic Features

Example taken from FAZ

in Raum stehen

| #P\text{N}Vs | DET | FUS | NUM |
|-------------|-----|-----|-----|
|             | no  | def | +   | sg  | pl  |
| literal     | 33  | 24  | 19  | 28  | 5   |
| idiomatic   | 53  | 53  | 53  | 53  | 0   |
| ambiguous   | 2   | 2   | 2   | 2   | 0   |

Ambiguous use

Sie steht als Manifest einer neuen Wohnkultur im Raum und macht heute jeden Besucher zum potentiellen Bewohner.

“It stands as manifest of a new living home decor in the room and makes today every visitor to a potential resident.”
Conclusion

▶ We presented a dataset of 9,700 manually analysed instances of idiomatic preposition-noun-verb triples, provided along with their morpho-syntactic properties.
Conclusion

- We presented a dataset of 9,700 manually analysed instances of idiomatic preposition-noun-verb triples, provided along with their morpho-syntactic properties.
- We intend to make this dataset available to the research community.
Conclusion

- We presented a dataset of 9,700 manually analysed instances of idiomatic preposition-noun-verb triples, provided along with their morpho-syntactic properties.
- We intend to make this dataset available to the research community.
- It can be used to train supervised classification approaches (cf. Diab and Bhutada, 2009)
Conclusion

- We presented a dataset of 9,700 manually analysed instances of idiomatic preposition-noun-verb triples, provided along with their morpho-syntactic properties.
- We intend to make this dataset available to the research community.
- It can be used to train supervised classification approaches (cf. Diab and Bhutada, 2009)
- Similar resources available for English (cf. Cook, 2008; Sporleder, 2010)
We presented a dataset of 9,700 manually analysed instances of idiomatic preposition-noun-verb triples, provided along with their morpho-syntactic properties.

We intend to make this dataset available to the research community.

It can be used to train supervised classification approaches (cf. Diab and Bhutada, 2009)

Similar resources available for English (cf. Cook, 2008; Sporleder, 2010)

but no similar resource for German!