

Bound Words

- Bound words (BWs): Words that only occur in a highly restricted set of environments, typically in only one multi-word expression (MWE)
- English examples: *headway* as in *make/ *show headway*; *umbrage* as in *take/ *feel umbrage*

- Usefulness of expressions with BWs for the classification and investigation of MWEs
- On-line collection of BWs in German and English (CoDII-BW)
- Connections between the study of BWs and PARSEME interests

Linguistic Aspects of Bound Words

BWs are documented for all languages for which phraseological research exists. Lists and classifications of BWs have been published for Dutch, English, German, Russian (Dobrovol'skij, 1988; Dobrovol'skij and Piirainen, 1994; Feyaerts, 1994; Fleischer, 1989). BWs occur in all types of MWEs, if we take the classification in Sag et. al. (2002) and Baldwin & Kim (2010).

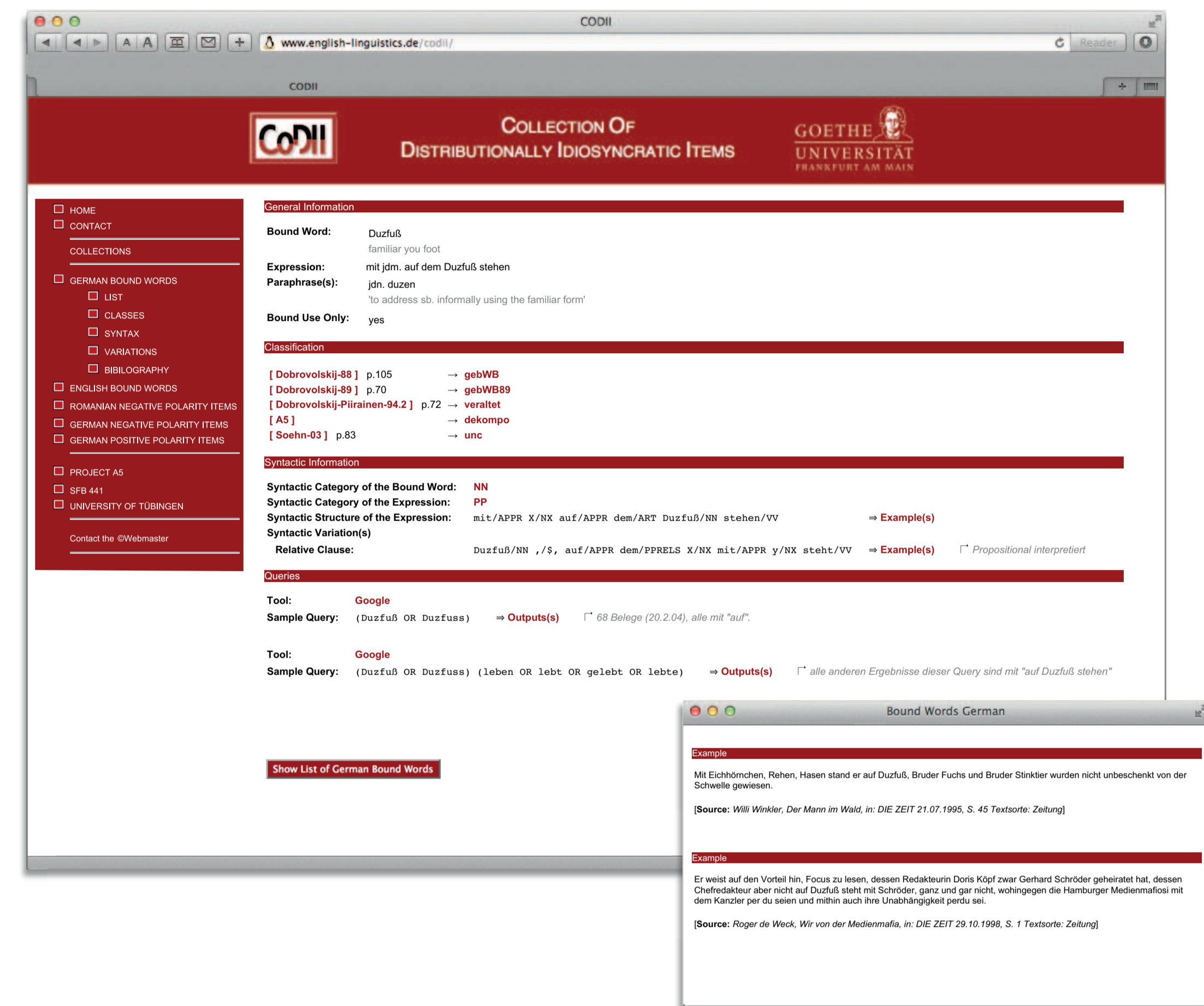
We see this in the following list, using expressions with BWs from German, the BW being underlined.

1. **fixed expressions:** *by and large*. with BWs: *Krethi und Plethi* (people of all kind)
2. **semi-fixed expressions:** complex proper names; compound nominals; non-decomposable idioms with BWs: *der gordische Knoten* (the Gordian knot); *Schornstein* (chimney); *wider den Stachel löcken* (kick against the goads)
3. **syntactically flexible idioms:** verb particle constructions; decomposable idioms; light verb constr. with BWs: *klein bei-geben* (to give in); *die Spendierhosen anhaben* (be generous); *jmdn. zum Hahnrei machen* (cheat on s.o.)
4. **institutionalized phrases:** idioms of encoding, not decoding ("collocations"); conversational formulae with BWs: *Eier abschrecken* (dip eggs into cold water), *klar wie Kloßbrühe* (quite plain/obvious)

Advantages of using BWs in the study of MWEs:

1. Restricted in distribution → naturally occurring examples are easy to extract by searching for the BW in any corpus, annotated or not.
2. Using a BW outside its usual context leads to ungrammaticality → strong argument for a collocational component in competence-oriented language descriptions, including parsing systems.
3. Collocation research: frequent MWEs (*auf Anhieb* (at first go)); transparent but infrequent expressions with BWs (*jmdm. etwas madig machen* (spoil s.o.'s pleasure in s.th.)) → collocational strength is not a mere frequency phenomenon.

CoDII Screenshot *Duzfuß*



The Collection of Distributionally Idiosyncratic Items (CoDII)

- Basic idea: Linguistic documentation of items with unexpected distribution. BWs as prime case.
- History: Developed since 2002 with contributions from the University of Tübingen and the University of Göttingen; now permanently maintained at the University of Frankfurt. Example publications: Trawiński et. al. (2008) and Richter, Sailer & Trawiński (2010).
- Present state of CoDII-BW:
 - CoDII-BW.de contains 444 German BWs, and
 - CoDII-BW.en contains 77 English BWs.

Information blocks in CoDII entries:

1. general information (glosses, translation to English),
2. syntactic information (including possible syntactic variation),
3. classification found in the literature,
4. usage examples from corpora and the Internet (including search patterns to make it easy for users to find more examples).

All information is encoded in XML, with a uniform underlying DTD for all collections. Because of this architecture, new collections on different languages can easily be added.

On-line bibliography on BWs at www.english-linguistics.de/sfb441/a5/bwb

XML Encoding

```
<?xml-stylesheet href="http://www.english-linguistics.de/sfb441/a5/bwb/dtd/coDII.dtd" type="text/xml" />
<entry id="duzfuss">
  <word>
    <orth>Duzfuß</orth>
    <familar you foot</fam>
    <only given</only>
  </word>
  <expression>
    <orth>mit jdm. auf dem Duzfuß stehen</orth>
    <paraphrase>jdm. diesem</paraphrase>
    <ortho address sb. informally using the familiar form</ortho>
  </expression>
  <classification>
    <class category="bw" class="gebhd" original-class="yes" type="Dobro88">
      <bibliography bib-item="Dobrovol'skij-88" p.199/>
    </class>
    <class category="bw" class="gebhd89" original-class="yes" type="Dobro89-strukty">
      <bibliography bib-item="Dobrovol'skij-89" p.199/>
    </class>
    <class category="bw" class="vealtat" original-class="yes" type="Dobro-Piir">
      <bibliography bib-item="Dobrovol'skij-Piirainen-94.2" p.72/>
    </class>
    <class category="bw" class="dekompo" original-class="yes" type="AS">
      <bibliography bib-item="AS"/>
    </class>
    <class category="bw" class="umt" original-class="yes" type="AS">
      <bibliography bib-item="Soehn-03" p.83/>
    </class>
  </classification>
  <expression-syntac cat="PP" nit/APPR X/RX auf/APPR dem/ART Duzfuß/NN stehen/VV</expression-syntac>
  <notation hist="duzfuss-Beis" kind="relative-class">
    <expression-syntac cat="NP">
      Duzfuß/NN /s. auf/APPR dem/PPRNS X/RX nit/APPR y/RX steht/VV
    </expression-syntac>
    <comment status="external">Prepositional interpretation</comment>
  </notation>
  <query type="google">
    <query-text>Duzfuß OR Duzfuss</query-text>
    <comment status="external">Beleg (20.2.04), alle mit "auf".</comment>
  </query>
  <query type="google" hist="duzfuss-leben">
    <query-text>
      (Duzfuß OR Duzfuss) (leben OR lebt OR gelebt OR lebte)
    </query-text>
  </query>
  <comment status="external">
    alle anderen Ergebnisse dieser Query sind mit 'auf Duzfuß stehen'
  </comment>
</entry>
```

Lexical Representation and Parsing

- Lexicography: Idiomatic key for MWEs
 - Lexicographic practice: No uniform treatment, sometimes the BW is used as a keyword, sometimes the syntactic head of the expression.
 - Related psycholinguistic question: What is/are the elements in a MWE that are most important in the recognition of the MWE? (*idiomatic key*, Titone & Connine 1999)
 - Identifying this key may be easier for expressions with a BW than for MWEs in general. Results on BWs can also be used for MWEs that are composed of free words.
- Parsing
 - Permit parsing BWs in inappropriate contexts (such as **show headway*)?
 - Over-parsing desirable, as some speakers are more flexible in the use of some BWs than others.
 - Requires 'decomposed' representation, i.e. parts of a MWE must be treated as independent lexical items, at least for some MWEs.
 - Parallel treatment of other (semi-)flexible MWEs

References

Baldwin, Timothy and Kim, Su Nam (2010). Multiword Expressions. In N. Indurkha and F. J. Damerou (Eds.), *Handbook of Natural Language Processing* (2 ed.), pp. 267–292. Boca Raton: CRC Press.

Dobrovol'skij, Dmitrij (1988). *Phraseologie als Objekt der Universalienlinguistik*. Leipzig: Verlag Enzyklopädie.

Dobrovol'skij, Dmitrij and Piirainen, Elisabeth (1994). Sprachliche Unikalia im Deutschen: Zum Phänomen phraseologisch gebundener Formative. *Folia Linguistica* 27(3–4), 449–473.

Feyaerts, Kurt (1994). Zur lexikalisch-semantischen Komplexität der Phraseologismen mit phraseologisch gebundenen Formativen. In *Sprachbilder zwischen Theorie und Praxis*, pp. 133–162. Bochum.

Fleischer, Wolfgang (1989). Deutsche Phraseologismen mit unikalener Komponente — Struktur und Funktion. In G. Gréciano (Ed.), *Europhras* 88, pp. 117–126.

Richter, Frank, Sailer, Manfred, and Trawiński, Beata (2010). The Collection of Distributionally Idiosyncratic Items. An Interface between Data and Theory. In S. Ptashnyk, E. Hallsteinsdóttir, and N. Bubenhofer (Eds.), *Korpora, Web und Datenbanken. Computergestützte und korpusbasierte Methoden in der Phraseologie, Phraseografie und der Lexikografie*, pp. 245–261. Hohengehren: Schneider Verlag.

Sag, Ivan A., Baldwin, Timothy, Copestake, Ann, and Flickinger, Dan (2002). Multiword Expressions: A Pain in the Neck for NLP. In A. F. Gelbukh (Ed.), *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2002)*, London, pp. 1–15. Springer.

Titone, Debra A. and Connine, Cynthia M. (1999). On the Compositional and Noncompositional Nature of Idiomatic Expressions. *Journal of Pragmatics* 31, 1655–1674.

Trawiński, Beata and Soehn, Jan-Philipp (2008). A Multilingual Database of Polarity Items. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.

Trawiński, Beata, Soehn, Jan-Philipp, Sailer, Manfred, and Richter, Frank (2008). A Multilingual Electronic Database of Distributionally Idiosyncratic Lexical Items. In *Proceedings of Euralex 2008*, Barcelona.