

Collocations and the Domains of Idiosyncratic Distribution*

Manfred Sailer
Seminar für Englische Philologie
Universität Göttingen
manfred.sailer@phil.uni-goettingen.de

May 18, 2006

1 Introduction: Collocations

- (1) a. strong/*powerful tea; powerful/*strong car
b. take a shower
c. hotel at; at the hotel (Kjellmer, 1994; Hausmann, 2004)
d. in connection with
e. cow — milk

Purely Occurrence-based Definitions

- (2) (Sinclair, 1991, p.170)
“Collocation is the occurrence of two or more words within a short space of each other in a text. The usual measure of proximity is a maximum of four words intervening.”
- “Word” as an orthographic entity: needs to be generalized to capture a lexeme’s inflected forms.
- (3) a. Kompromisse eingehen/ einen Kompromiss eingehen
compromises enter/ a compromise enter
b. Kompromisse machen / *einen Kompromiss machen
compromises make/ a compromise make
- Restriction to a text window defined by word distance. Problematic for German N-V collocations (Breidt, 1993):
- (4) a. (eine) Frage stellen
a question put
b. Diese **Frage** ist Josef Ebenöther im Laufe der Jahrzehnte oft **gestellt** worden.
‘J. E. has often been asked this question in the course of the decades.’
- (5) (Hoey, 1991, p.7)
“Collocation has long been the name given to the relationship a lexical item has with items that appear with greater than random probability in its (textual) context.”
- Includes an explicit notion of relative co-occurrence frequency.
 - What is a *lexical item*? see (3)

*The material presented in this talk is based to a large extent on joint work and fruitful discussions with Timm Lichte, Frank Richter, Jan-Philipp Soehn, and Beata Trawińska. They might, however, not share all of my interpretations and conclusions — especially not the wrong ones.

- Also includes pairs of the type *cow — milk*.
- (6) (Butler, 1985, p.130)
“The defining feature of a lexical item, by which such an item is recognized, is its pattern of co-occurrence with other items, that is its *collocational* behavior. A lexical item is recognized as different from other lexical items because its total pattern of collocation is unique.”
- also includes less-frequent co-occurrence partners
 - very much like Hoey (1991)

Bringing in Competence van der Wouden (1992, 1997) generalizes from the notion of collocations from a lexeme-lexeme relation to more abstract patterns:

- (7) (van der Wouden, 1997, p.5)
“Collocation: idiosyncratic restriction on the combinability of lexical items.”
- only the “idiosyncratic” occurrence aspects should be considered collocations: differentiate between *hotel at* and *at the hotel*;
 - i.e. notions from “outside” the observable corpus are considered: grammar, semantics, pragmatics, ...
- (8) (van der Wouden, 1997, p.50)
“If all restrictions on the distribution of lexical items are taken to be collocational, uniform listing of collocational combinations in the lexicon is impossible: certain open classes of lexical items (e.g. negative polarity items) are restricted to the neighbourhoods of member of another open class (viz. monotone decreasing operators).”
- An attempt to represent collocations in a compact form by generalizing over context by means of grammatical terms.
 - Predictive theory! Links up with “speaker’s knowledge”.

1.1 What is Regular — What is Irregular?

1.1.1 Lexis and Grammar

Regularity: Grammar, Open Choice Principle

- Rules/Principles of how to combine words and phrases to larger syntactic units.
- Rules/Principles of how to interpret syntactic combinations (compositionality)
- Allows for creativity
- Predominant perspective in Formal Grammar

Irregularity: Lexis, Idiom Principle

- Idiosyncratic information about the “basic units” that Grammar can manipulate.
- Information accessed by syntactic/semantic rules: phonological shape; part of speech, selection; semantics.
- Additional idiosyncratic distributional information (collocations).
- Makes speech more predictable (and, thus, faster in production and processing)
- Predominant perspective in lexicography.

- Problems:

- Where do idiosyncratic syntactic and semantic combinations belong? (especially phraseological patterns (*Phraseolschablonen*, Fleischer (1997))?)
- How should/can tendencies/preferences be integrated?
Note: tendencies also exist at the Grammar site: why is there a preference for a particular structure in a given variety/situation/...?

1.1.2 Linguistic Theories

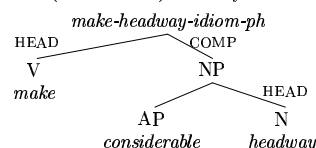
- Construction Grammar:

- There is no Grammar/Lexis distinction, but a continuum of regularity/abstractness in one dimension and a continuum of complexity of the basic units in a second dimension (Fillmore et al., 1988; Goldberg, 1995).
- Must treat collocations as constructions.
- Usage-based branches (Croft, 2001)

- Head-driven Phrase Structure Grammar (HPSG)

- Pollard and Sag (1987, 1994): Lexis is confined to phonological, syntactic and pragmatic information about a lexical item itself. No treatment of idioms or collocations.
- Constructional HPSG: like Construction Grammar, but: since every word in a construction must be licensed by a lexical entry, a fully constructional approach (à la Abeillé (1995)) is not possible. (a version of Gert's "handshake" problem, not entirely solved in Riehemann (2001))

(9) make (considerable) headway



2 Grammaticalized Collocations

2.1 Bound Words

(10) (Aronoff, 1976, p.15)

"There are words which, like cranberry morphs, concatenate only with specific words and not with syntactic classes. For example the noun *headway* occurs only as the direct object of the verb *make*, just as *cran* occurs only in *cranberry*."

- collected and described in the phraseological literature: Dobrovol'skij (1988, 1989); Dobrovol'skij and Piirainen (1994a,b); Fleischer (1989); Feyaerts (1990, 1992)
- Described in HPSG: Richter and Sailer (2003); Soehn (2003); Soehn and Sailer (2003)
- Online collections: *Sammlung unikaler Wörter des Deutschen* (450 bound words) and *Collection of English Bound Words* (75 bound words) (Sailer and Trawinski, ta; Sailer and Trawiński, 2006) available at:
www.sfb441.uni-tuebingen.de/a5/codii

(11) Descriptive findings:

A **BW** must co-occur with another lexeme (**LEX**). There may only be one of the following dependency relations between BW and LEX:

- Angst einjagen** (LEX is a complement of BW)
 fear in chase 'frighten'
 English: wend one's way; cudgel one's brain
- Tacheles reden** (BW is a complement of LEX)
 goal talk 'talk straight'
 English: **make** headway; **on** tenterhooks; **put** the kibosh on someone
- der **große Zampano** (LEX is an adjunct to BW)
 the big Zampano 'the big doer'
 English: the **whole** caboodle
- der **gordische Knoten** (BW is an adjunct to LEX)
 the gordic knot
 English: the curule chair 'highest rank'; eliminary matches; lending library
- sich **freuen** wie ein **Schneekönig** (BW is compl. of *wie*; *wie*-PP is adjunct to LEX)
 oneself be happy like a snow king 'be in a state of extraordinary joy'
 English: **sell** like hotcakes; **happy** as a sandboy
- klipp und klar** (BW is a sister-conjunct to LEX)
 ?? and clear 'very clearly'
 English: **to** and fro;

- Of course, LEX can also be a BW! (*flotsam and jetsam*, *tit for tat*)

- These dependency relations are relatively local, but go beyond the reach of selection!

- (12) A Bound Word is a lexical item which requires the occurrence of
 Licensor: an instance of a particular other lexical item
 Domain: with at most one intervening dependent.

2.2 Negative Polarity Items

(13) Negative polarity items:

- Kim didn't lift a finger to help Chris.
- *Kim lifted a finger to help Chris.

- Kim hasn't seen any student here for a long time.
- *Kim has seen any student here for a long time.

(14) "Non-negative" licensing:

- Have you made any changes to the slides yesterday? (neutral)
- Have you ever lifted a finger to help me? (biased towards a negative answer)
- Chris denies to have anything to do with this problem.

- Widely studied within formal syntax, semantics and pragmatics (Klima, 1964; Krifka, 1995; Ladusaw, 1980; Progovac, 1988; Zwarts, 1997)
- Compilations of NPIs: Kürschner (1983); Hoeksema (2002); (NPI versions of the Collection of Distributionally Idiosyncratic Items are planned in Tübingen and Göttingen)

(15) Descriptive findings:

- An NPI can be in the immediate scope of negation:
 - Chris wasn't one bit happy about this result.
 - Chris didn't lift a finger to help.
 - Chris didn't give any answer.

- b. An NPI can be in the scope of $\neg\exists x \dots$:
- *No one was one bit happy about this result.
 - No one** lifted a finger to help.
 - No one** ever called.
- c. An NPI can be in the scope of $\neg Qx \dots$:
- *Not many students were one bit happy about this result.
 - *Not many students lifted a finger to help.
 - Not many** students have ever used this library.

There cannot be a second quantified expression intervening:

- (16) *No student gave every teacher any apple
 $\neg\exists x(\text{student}'(x) \wedge \forall y(\text{teacher}'(y) \rightarrow \exists z(\text{apple}'(z) \wedge \exists e\text{give}'(e, x, y, z))))$
- (17) A Negative Polarity Item is a lexical item which requires the occurrence of
 Licenser: an instance of negation
 Domain: with at most one intervening quantifier.

Are there items which require to have scope over a negation?

- (18) Jan wollte überhaupt *(**nicht**) anrufen.
 Jan wanted at all not call
 'Jan didn't want to call at all.'

2.3 Other Types of Strict Collocations

Sailer and Richter (2002) and Sailer (2004) studied the words *fackeln* (*dither*) and *wahrhaben* (*accept*), comparing corpus data from the IDS-corpora and introspective data.

- (19) *fackeln* (Canonical example: negative context + *lange*; 94,9%):
- Der neue Geschäftsführer fackelte nicht lange.
 the new manager dithered not long 'The new manager didn't dither very long.'
 - collocational requirements:
 - Dependency domain: *fackeln* must occur with an indication of duration.
 - logical form domain: *fackeln* must occur in the scope of a negation.
- (20) *wahrhaben* (Canonical example: negative context + *want*; 96%):
- Man wolle die mögliche Diagnose Herzinfarkt nicht wahrhaben, sagte S.
 one wants the possible diagnosis heart attack not true.have said S.
 'People didn't want to face the reality of the possible diagnosis heart attack, said S.'
 - collocational requirements:
 - Dependency domain: *wahrhaben* must be a dependent of an expression of volition
 - logical form domain: *wahrhaben* must be in the scope of a negation

3 Collocations with Polysemous Words

3.1 Lexeme-Lexeme Collocations

- (21) a. ein Ei abschrecken (head-complement relation)
 an egg scare away 'chill an egg'
 English: take a shower
- b. schwer enttäuscht (head-adjunct relation)
 heavily disappointed 'badly disappointed'
 English: strong tea; powerful car
- c. schöne Augen machen (one intervener)
 beautiful eyes make 'give someone the glad eye'

3.2 Decomposable Idioms with Free Readings

- (22) a. den Mund vollnehmen (head-complement relation)
 the mouth full take 'talk big'
 English: spill the beans
- b. fest im Sattel (sitzen) (head-adjunct relation)
 firmly in the saddle sit 'firmly in the saddle'
- c. schöne Augen machen (one intervener)
 beautiful eyes make 'give someone the glad eye'

More complex idioms have non-decomposable parts:

- (23) wissen [wo Barthels den Most holt]
 know where ?? the ?? gets
 'know every trick in the book' (Sailer, 2003)

Other dependency oriented theories of idioms:

- O'Grady (1998) assumes that there is a head of an idiom and there is an uninterrupted dependency connection of arbitrary depth between the head and the other components of the idiom.
 - O'Grady (1998) does not distinguish between decomposable and non-decomposable idioms.
 - non-headed idioms cannot be described: *[the cat] [out of the bag]* see Riehemann (2001).
 For us: even though there is an intervening verb, *the cat* is a direct dependent of *out of the bag* in: *let the cat out of the bag, the cat is out of the bag*. Similar: *Schmalhans (ist) Küchenmeister (thin-Hans is chief)* (someone who cannot spend a lot of money is in charge of the expenses) (Soehn, 2006).
- Everaert and Kuiper (1996) and Kuiper (2004) propose an X-bar based theory which assumes that idioms have a head which can select lexemes in its dependency positions. Idiom parts are, thus, combined in syntax, but the idiom head has access to all information of the selected element that is normally available only in morphology (i.e. phonological shape, particular lexeme etc.) This theory is similar in its predictions and problems to O'Grady (1998), but does not use dependency structures explicitly and works with syntax-morphology analogy.

4 Collocations in HPSG

The following types of attempts have been made to capture collocations in HPSG:

4.1 Allowing for lexeme selection

- Krem and Erbach (1994); Sag et al. (2003); Soehn (2006)
- Problem: the collocational data go beyond the selection domain
- Even within the dependency domain, lexeme information is necessary, but not enough: *fackeln/wahrhaben* allow for a range of lexemes.

4.2 An explicit collocation module

- shared assumption: there is an attribute COLL such that:
 - a "trivial" COLL value is assumed for all regular phrases
 - a "non-trivial" COLL value is assumed for all lexical items (simple words or internally irregular phrases [such as non-decomposable idioms and phraseological patterns]).

- only lexical items can have collocational restrictions; for collocationally restricted items, the restriction is expressed in the COLL value.
- a general principle will insure that the lexically specified collocational requirements are met in the structure in which a given lexical item occurs.
- Syntactic barriers and limited licenser information (Richter and Sailer, 1999; Soehn, 2006):
 - The collocational restrictions (be it “lexeme” restrictions or “logical form” restrictions) must hold within a certain syntactic domain. For example: *tenterhooks* must be the argument of *on* within the smallest PP which contains *tenterhooks*.
 - In practice, only “smallest complete clause” is used as a domain in Soehn (2006). Smaller domains (such as “pp”) can be accommodated through the fact that the licenser’s full LOC value is given.
 - Depending on the kind of collocational restriction, only part of the information on the licenser is present:
 - * its LOC-value for dependency/lexeme collocations
 - * its logical form for NPIs etc.
- Full utterances in COLL (Sailer, 2003):
 - Every lexical item’s COLL value is identical to the entire utterance in which the lexical item occurs. Any property of this utterance can, then, be said to be relevant.
 - Too powerful! too unconstrained!!

4.3 A Domain-based Collocation Module

4.3.1 The New COLL

1. There is a feature COLL defined on all signs.
2. Regular phrases have a trivial (i.e. atomic) COLL value *empty-coll*.
3. Lexical items have a non-trivial COLL value of the following form:

$$(24) \quad \text{COLL value of lexical items:}$$

$$\left[\begin{array}{l} \textit{nonempty-coll} \\ \text{DEP-DOMAIN } \textit{list}(\textit{synsem}) \\ \text{LF-DOMAIN } \textit{list}(\textit{meaningful-expression}) \end{array} \right]$$

4. Two COLLOCATION DOMAIN PRINCIPLES:

- (25) a. DEPENDENCY DOMAIN PRINCIPLE:
- For every *sign* object *s*, and for every *sign* object *o*,
- if *o* has a COLL DEP-DOMAIN value *d*, then every *synsem* object *o*
- that occurs in *s*
- and that is at dependency-distance ≤ 2 to *o* in *s*,
- is on *d*.
- And if *s* is unembedded or the lexical head of *s* is at dependency-distance > 2 to *o* in *s*, then nothing else is on *d*.

b. LOGICAL FORM DOMAIN PRINCIPLE:

For every *sign* object *s*, and for every *sign* object *o*,

if *o* has a COLL LF-DOMAIN value *d*, then every *meaningful-expression* object that occurs in *s*’s logical form and that is at lf-distance ≤ 2 to *o* on *s*’s logical form, is on *d*.

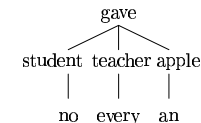
And if *s* is unembedded or the highest expression in the logical form of *s* is at lf-distance > 2 to *o*, then nothing else is on *d*.

- (26) a. dependency-distance:
- For each *sign* objects *s* and *o*, such that *o* is a lexeme dominated by *s*,
- (i) A *synsem* object *y* is on dependency-distance 0 to a *sign* *o* on *s* iff *y* is *o*’s SYNSEM value.
- (ii) A *synsem* object *y* is on dependency-distance $x + 1$ to a *sign* *o* in *s*, iff there is a *sign* *o*’ such that
- y* is on dependency-distance *x* to *o*’ on *s*
- and *y* is the SYNSEM value of (the lexical head) of a *sign* *o*’ such that *o*’ and *o*’ are in a head-complement relation or in a head-adjunct relation.
- b. lf-distance:
- (i) A *meaningful-expression* ϕ is on lf-distance 0 to a *sign* *o* iff ϕ is the content of *o*;
- (ii) A *meaningful-expression* ϕ is on lf-distance *x* to a *sign* *o* iff there is a *meaningful-expression* ϕ ’ such
- ϕ ’ is on lf-distance *x* to *o*,
- and either ϕ is a proper subexpression of ϕ ’ or the other way around,
- and ϕ and ϕ ’ do not differ with respect to occurrences of negation operators, and (non-existential) quantifiers
- (iii) A *meaningful-expression* ϕ is on lf-distance $x + 1$ to a *sign* *o* iff there is a *meaningful-expression* ϕ ’ such
- ϕ ’ is on lf-distance *x* to *o*,
- and either ϕ is a proper subexpression of ϕ ’ or the other way around,
- and ϕ and ϕ ’ differ with respect to the occurrence of exactly one of a negation operator or a (non-existential) quantifier.

(27) Example:

- a. No student gave every teacher an apple.

- (i) Dependency structure:



- (ii) Logical form:

$$\neg \exists x(\text{student}'(x) \wedge \forall y(\text{teacher}'(y) \rightarrow \exists z(\text{apple}'(z) \wedge \exists e(\text{give}'(e, x, y, z))))))$$

- b. collocation domains of *apple*:

- (i) dependency domain:
- | | |
|-------------|------------------|
| distance 0: | apple |
| distance 1: | an, gave |
| distance 2: | teacher, student |
| distance 3: | no, every |

(ii) logical form domain:

$$\begin{array}{l} 0: \neg\exists x(\text{student}'(x) \wedge \forall y(\text{teacher}'(y) \rightarrow \exists z(\text{apple}'(z) \wedge \exists e(\text{give}'(e, x, y, z)))))) \\ 1: \neg\exists x(\text{student}'(x) \wedge \forall y(\text{teacher}'(y) \rightarrow \exists z(\text{apple}'(z) \wedge \exists e(\text{give}'(e, x, y, z)))))) \\ 2: \neg\exists x(\text{student}'(x) \wedge \forall y(\text{teacher}'(y) \rightarrow \exists z(\text{apple}'(z) \wedge \exists e(\text{give}'(e, x, y, z)))))) \end{array}$$

$$3: \neg\exists x(\text{student}'(x) \wedge \forall y(\text{teacher}'(y) \rightarrow \exists z(\text{apple}'(z) \wedge \exists e(\text{give}'(e, x, y, z))))))$$

c. collocation domains for *an*:

(i) dependency domain: distance 0: an
distance 1: apple
distance 2: gave
distance 3: student, teacher
distance 4: no, every

(ii) If domain:

$$\begin{array}{l} 0: \neg\exists x(\text{student}'(x) \wedge \forall y(\text{teacher}'(y) \rightarrow \exists z(\text{apple}'(z) \wedge \exists e(\text{give}'(e, x, y, z)))))) \\ 1: \neg\exists x(\text{student}'(x) \wedge \forall y(\text{teacher}'(y) \rightarrow \exists z(\text{apple}'(z) \wedge \exists e(\text{give}'(e, x, y, z)))))) \\ 2: \neg\exists x(\text{student}'(x) \wedge \forall y(\text{teacher}'(y) \rightarrow \exists z(\text{apple}'(z) \wedge \exists e(\text{give}'(e, x, y, z)))))) \end{array}$$

$$3: \neg\exists x(\text{student}'(x) \wedge \forall y(\text{teacher}'(y) \rightarrow \exists z(\text{apple}'(z) \wedge \exists e(\text{give}'(e, x, y, z))))))$$

Remarks:

- If you replace *an apple* with *any apple* the sentence is ungrammatical (see (16)) because the negation is at If-distance 3 to *any!*
- If-distance is defined in such a way that only scopal elements can be at an If-distance to the content of a given sign, and there must be scopal relation between the content of the given sign and the If-distant operator. Consequently, an item in the scope of a quantifier cannot be If-collocationally dependent on an expression in the restrictor of a quantifier.

4.4 Examples

In the lexical entry of a collocationally restricted item, we, now, specify a member of the DEP-DOM- or the LF-DOM-list.

(28) The collocational specification for *headway*:

$$\left[\begin{array}{l} \text{PHON } (\textit{headway}) \\ \text{SYNSEM } \left[\text{LOC CONT INDEX } \boxed{\square} \right] \\ \text{COLL } \left[\begin{array}{l} \text{nonempty-coll} \\ \text{DEP-DOM } \left\langle \dots, \left[\text{LISTEME } \textit{make} \right. \right. \right. \\ \left. \left. \left. \text{ARG-ST } \left\langle \dots, \text{NP}[\text{INDEX } \boxed{\square}], \dots \right\rangle, \dots \right\rangle \right. \\ \left. \text{LF-DOM } \textit{list}(\textit{meaningful-expression}) \right. \end{array} \right] \end{array} \right]$$

Remarks: I follow Soehn (2006) in assuming a (head) attribute LISTEME whose value indicates the lexeme (=listeme). For simplicity I assume that ARG-ST is defined on *synsem*. Alternatively, a disjunction over the valence (and slash) attributes could be formulated.

- (29) a. John **made** considerable headway last year.
make is at dep-distance 1 to *headway*, and the coindexation holds.
- b. *John achieved considerable headway last year.
The DEP-DOM-list of the sentence does not contain *make*.
- c. *John **made** a mistake and, therefore, lacks headway now.
make is at dep-distance 3, and the coindexation is missing.

d. *John **made** an attempt to headway.

If argument-marking prepositions are not counted for dependency, *make* is at dep-distance 2 from *headway*, but the coindexation is missing.

(30) The collocational specification for *fackeln* (*dither*):

$$\left[\begin{array}{l} \text{PHON } (\textit{fackeln}) \\ \text{SYNSEM } \left[\text{LOC CONT } \left[\begin{array}{l} \text{INDEX } \boxed{\square} \\ \text{MAIN } \textit{hesitate}'(e, x) \end{array} \right] \right] \\ \text{COLL } \left[\begin{array}{l} \text{DEP-DOM } \left\langle \dots, \left[\text{LOC } \left[\begin{array}{l} \text{HEAD MOD } \left[\text{LOC CONT INDEX } \boxed{\square} \right] \right] \right] \right. \right. \\ \left. \left. \text{CONT } \boxed{\square} \right] \right\rangle, \dots \right\rangle \\ \text{LF-DOM } \left\langle \dots, \neg(\dots \textit{hesitate}'(e, x) \dots), \dots \right\rangle \end{array} \right] \end{array} \right]$$

where $\boxed{\square}$ expresses a duration.

- (31) a. **Niemand** hat **lange** gefackelt.
no one has long dithered
- dep-dom: *lange* is at dep-distance 1 and the coindexation holds.
 - If-dom: *niemand* is at If-distance 1 and the embedding holds.
- b. ***Niemand** hat immer **lange** gefackelt.
no one has always long dithered
- dep-dom: *lange* is at dep-distance 1 and the coindexation holds.
 - If-dom: *niemand* has scope over *fackeln*, but it is at If-distance 3.

5 Which Item Imposes the Restrictions?

- (32) ein Ei/ Metall/ *Gemüse abschrecken
an egg/ metal/ vegetables scare away
‘chill an egg/ metal’

The dilemma:

- We need a semantically bleached form of the verb *abschrecken* which is collocationally restricted to some nouns.
- Nonetheless, it is part of our lexical knowledge of the word *Ei* that it combines with *abschrecken* (Hausmann, 2004).

The solution:

- Basic idea: Every noun comes with its own *lexical functions* (Mel'čuk, 1999).
- For *abschrecken*: there is a lexical entry with the semantics of *chill* which is collocationally restricted as indicated.

$$(33) \left[\begin{array}{l} \text{PHON } (\textit{abschrecken}) \\ \text{SYNS } \left[\text{LOC CONT MAIN } \textit{chill}'(e, x, y) \right] \\ \text{ARG-ST } (\textit{synsem}, \left[\text{LOC CONT INDEX } \boxed{\square} \right]) \\ \text{COLL } \left[\text{DEP-DOM } \left\langle \dots, \boxed{\square} \left[\text{LOC CONT } \left[\text{INDEX } \boxed{\square} \right] \right], \dots \right\rangle \right] \end{array} \right]$$

where $\boxed{\square}$ is the listeme *Ei* or expresses a content which is ontologically below *metall*.

- For *Ei*: the lexical entry contains the following implicational constraint: if the DEP-DOM list contains an item which expresses the lexical function of “chilling”, then this item must be the listeme *abschrecken*.

$$(34) \quad \left[\begin{array}{l} \text{PHON } \langle \text{E} \rangle \\ \text{SYNS } \left[\begin{array}{l} \text{LOC CONT INDEX } \boxed{1} \end{array} \right] \\ \text{COLL } \left[\begin{array}{l} \text{DEP-DOM } \boxed{2} \end{array} \right] \end{array} \right] \text{ and}$$

if $\boxed{2}$ contains a *synsem* object $\boxed{3}$ of the form $\left[\begin{array}{l} \text{SYNS LOC CONT MAIN chill'}(e, x, y) \\ \text{ARG-ST } \langle \text{synsem}, \left[\text{LOC CONT INDEX } \boxed{1}, \dots \right] \rangle \end{array} \right]$,
then $\boxed{3}$ is the listeme *abschrecken*.

(35) Back to the data in (3)

- a. The lexical entry of *Kompromiss* (*compromise*): if the dep-domain contains an item which expresses the lexical function of “agree on (a compromise)”, then this item must be *machen* (selecting for a plural), *eingehen* (*enter*), ...

6 Conclusion

- New version of the COLL analysis which combines the ontological simplicity of Sailer (2003) and the restrictiveness of Soehn (2006).
- Empirically based on collections of bound words and on the literature about collocations, idioms and negative polarity items.
- Takes the maximal domains within which idiosyncratic distributions can be observed as point of departure.
- Third potential domain for collocational restrictions: phonology (*a/an* alternation (Asudeh and Klein, 2002)).
- Relation between qualia structure and lexical functions?
- Application: domains can be used for a more fine-tuned automatic extraction of collocations (Vilada Moiroón, 2005; Lichte, 2005)
- Very first sketch of this approach!

References

- Abeillé, Anne (1995). The Flexibility of French Idioms: A Representation with Lexical Tree Adjoining Grammar. In M. Everaert, E.-J. v. d. Linden, A. Schenk, and R. Schreuder (Eds.), *Idioms. Structural and Psychological Perspectives*, pp. 15–42. Lawrence Erlbaum Associates, Hillsdale.
- Aronoff, Mark (1976). *Word Formation in Generative Grammar*. MIT Press, Cambridge, Massachusetts and London, England.
- Asudeh, Ash and Klein, Ewan (2002). Shape Conditions and Phonological Context. In F. van Eynde, L. Hellan, and D. Beermann (Eds.), *Proceedings of the 8th International HPSG Conference*, pp. 20–30. CSLI Publications.
- Breidt, Elisabeth (1993). Extraction of N-V-Collocations from Text Corpora: A Feasibility Study for German. In *Proceedings of the 1st ACL Workshop on Very Large Corpora*, Columbus, Ohio. Revised version available from <http://arxiv.org/abs/cmp-1g/9603006>.
- Butler, Christopher S. (1985). *Systemic Linguistics: Theory and Application*. Batsford, London.
- Croft, William (2001). *Radical Construction Grammar. Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Dobrovolskij, Dmitrij (1988). *Phraseologie als Objekt der Universalienlinguistik*. Verlag Enzyklopädie, Leipzig.

Dobrovolskij, Dmitrij (1989). Formal gebundene phraseologische Konstituenten: Klassifikationsgrundlagen und theoretische Analyse. In W. Fleischer, R. Große, and G. Lerchner (Eds.), *Beiträge zur Erforschung der deutschen Sprache*, Volume 9, pp. 57–78. Leipzig, Bibliographisches Institut.

Dobrovolskij, Dmitrij and Piirainen, Elisabeth (1994a). PGF: Auf dem Präsentierteller oder auf dem Abstellgleis? *Zeitschrift für Germanistik* (NF 4), 65–77.

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

Everaert, Martin and Kuiper, Koenraad (1996). Theory and Data in Idiom Research. In L. McNair et al. (Ed.), *CLS32: the Parasession on Theory and Data in Linguistics*, Chicago, pp. 43–57. Chicago Linguistic Society.

Feyaerts, Kurt (1990). ‘Haben Sie auch etwas auf dem Kerbholz?’ Eine lexikalisch-semantische Beschreibung phraseologisch gebundener Wörter im Deutschen und Niederländischen aus synchronischer Sicht. Master’s thesis, Katholieke Universiteit te Leuven, Departement Germaanse Filologie.

Feyaerts, Kurt (1992). ‘Hebt U ook iets op uw kerfstok?’ Fraseologisch gebonden woorden in het Duits en het Nederlands.

Fillmore, Charles, Kay, Paul, and O’Connor, M. (1988). Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*. *Language* 64, 501–538.

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

Fleischer, Wolfgang (1997). *Phraseologie der deutschen Gegenwartssprache* (2nd, revised edition ed.). Niemeyer, Tübingen.

Goldberg, Adele (1995). *Constructions. A Construction Grammar Approach to Argument Structure*. University of Chicago Press, Chicago, London.

Hausmann, Franz-Joseph (2004). Was sind eigentlich Kollokationen? In K. Steyer (Ed.), *Wortverbindungen — mehr oder weniger fest*, Institut für Deutsche Sprache, Jahrbuch 2003, pp. 309–334. Berlin, New York: de Gruyter.

Hoeksema, Jacob (2002, November). De negatief-polaire uitdrukkingen van het Nederlands. Inleiding en lexicon. Manuskript, Rijksuniversiteit Groningen; constantly updated.

Hoey, Michael (1991). *Patterns of Lexis in Text*. Oxford: Oxford University Press.

Kjellmer, Johannes (1994). *A Dictionary of English Collocations*. Oxford: Clarendon. 3 volumes.

Klima, Edward (1964). Negation in English. In J. A. Fodor and J. Katz (Eds.), *The Structure of Language*, pp. 246–323. Prentice Hall, Englewood Cliffs, New Jersey.

Krenn, Brigitte and Erbach, Gregor (1994). Idioms and Support Verb Constructions. In J. Nerbonne, K. Netter, and C. Pollard (Eds.), *German in Head-Driven Phrase Structure Grammar*, pp. 365–396. CSLI Publications. Lecture Notes 46.

Krifka, Manfred (1995). The Semantics and Pragmatics of Weak and Strong Polarity Items. *Linguistic Analysis* 25, 209–257.

Kuiper, Koenraad (2004). Phraseologie aus der Sicht der generativen Grammatik. In K. Steyer (Ed.), *Wortverbindungen — mehr oder weniger fest*, Institut für Deutsche Sprache, Jahrbuch 2003, pp. 221–237. Berlin, New York: de Gruyter.

Kürschner, Wilfried (1983). *Studien zur Negation im Deutschen*. Gunter Narr, Tübingen.

Ladusaw, William (1980). *Polarity Sensitivity as Inherent Scope relations*. Garland Press, New York.

- Lichte, Timm (2005). Corpus-based Acquisition of Complex Negative Polarity Items. In J. Gervain (Ed.), *Proceedings of the Tenth ESSLLI Student Session*, Edinburgh. Heriot-Watt University.
- Meřčuk, Igor (1999). Collocations and Lexical Functions. In A. P. Cowie (Ed.), *Phraseology. Theory, Analysis, and Applications*, pp. 23–53. Oxford: Clarendon.
- O’Grady, William (1998). The Syntax of Idioms. *Natural Language and Linguistic Theory* 16, 279–312.
- Pollard, Carl and Sag, Ivan (1987). *Information Based Syntax and Semantics. Vol.1: Fundamentals*. Stanford: CSLI Lecture Notes 13.
- Pollard, Carl and Sag, Ivan A. (1994). *Head-Driven Phrase Structure Grammar*. Chicago and London: University of Chicago Press.
- Progovac, Ljiljana (1988). *A Binding Approach to Polarity Sensitivity*. Ph. D. thesis, University of Southern California.
- Richter, Frank and Sailer, Manfred (1999). LF Conditions on Expressions of Ty2: An HPSG Analysis of Negative Concord in Polish. In R. D. Borsley and A. Przepiórkowski (Eds.), *Slavic in HPSG*, pp. 247–282. Stanford: CSLI Publications.
- Richter, Frank and Sailer, Manfred (2003). Cranberry Words in Formal Grammar. In C. Beyssade, O. Bonami, P. C. Hofherr, and F. Corblin (Eds.), *Empirical Issues in Formal Syntax and Semantics*, Volume 4, pp. 155–171. Paris: Presses Universitaires de Paris-Sorbonne.
- Riehemann, Susanne Z. (2001). *A Constructional Approach to Idioms and Word Formation*. Ph. D. thesis, Stanford University.
- Sag, Ivan A., Wasow, Thomas, and Bender, Emily M. (2003). *Syntactic Theory: A Formal Introduction* (2nd revised ed.). Stanford: CSLI.
- Sailer, Manfred (2003). Combinatorial Semantics and Idiomatic Expressions in Head-Driven Phrase Structure Grammar, Phil. Dissertation (2000). Arbeitspapiere des SFB 340. 161, Universität Tübingen.
- Sailer, Manfred (2004). Distributionsidiosynkrasien: Korpuslinguistische Erfassung und grammatiktheoretische Deutung. In K. Steyer (Ed.), *Wortverbindungen — mehr oder weniger fest*, Institut für Deutsche Sprache, Jahrbuch 2003, Berlin, New York, pp. 194–221. de Gruyter.
- Sailer, Manfred and Richter, Frank (2002). Not for Love or Money: Collocations! In G. Jäger, P. Monachesi, G. Penn, and S. Wintner (Eds.), *Proceedings of Formal Grammar 2002*, pp. 149–160.
- Sailer, Manfred and Trawiński, Beata (2006). The Collection of Distributionally Idiosyncratic Items: A Multilingual Resource for Linguistic Research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC 2006*.
- Sailer, Manfred and Trawinski, Beata (t.a.). Die Sammlung unikaler Wörter des Deutschen. Aufbauprinzipien und erste Auswertung. In A. Häcki-Buhofer (Ed.), *EUROPHRAS 2003*.
- Sinclair, John (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Soehn, Jan-Philipp (2003). Von Geisterhand zu Potte gekommen. Eine HPSG-Analyse von PPs mit unikaler Komponente. Magisterarbeit, Universität Tübingen, Seminar für Sprachwissenschaft.
- Soehn, Jan-Philipp (2006). *Über Bärendienste und erstaunte Banklötze. Idiome ohne freie Lesart in der HPSG*. Frankfurt am Main; Peter Lang. Ph.D. thesis, Friedrich-Schiller-Universität Jena.
- Soehn, Jan-Philipp and Sailer, Manfred (2003). At First Blush on Tenterhooks. About Selectional Restrictions Imposed by Nonheads. In G. Jäger, P. Monachesi, G. Penn, and S. Wintner (Eds.), *Proceedings of Formal Grammar 2003*, pp. 149–161.
- Wouden, Ton van der (1992). Bepkeringen op het optreden van lexicale elementen. *De Nieuwe Taalgids* 85(6), 513–538.
- Wouden, Ton van der (1997). *Negative Contexts. Collocation, Polarity and Multiple Negation*. London: Routledge.
- Villada Moiroón, Begoña (2005). *Computational Aspects of Fixed Expressions: Acquisition and Modification Potential*. Ph. D. thesis, Rijksuniversiteit Groningen.
- Zwarts, Frans (1997). Three Types of Polarity. In F. Hamm and E. W. Hinrichs (Eds.), *Plurality and Quantification*, pp. 177–237. Kluwer Academic Publishers, Dordrecht.