A model for corpus-driven exploration and presentation of multi-word expressions

Annelen Brunner[1] and Kathrin Steyer[1]

Institute for the German Language, Mannheim

Abstract. In this paper we outline our corpus-driven approach on detecting, describing and presenting multi-word expressions (MWE). We make use of large corpora and statistical data to explore and visualize the rich interrelations, patterns and types of variances of MWE. In order to do this, we have developed a method of linguistically interpreting collocational data in a structured way (cf. [Steyer/Brunner 2009]). Several levels of abstraction build on each other: surface patterns, Lexical realizations (LR), MWE and MWE patterns. Generalizations are made in a controlled way and in adherence to corpus evidence. The method helps to identify and describe MWE in a way that gives credit to their flexible nature and their role in language use.

1 Methodological foundations

Our approach is corpus-driven as defined by Tognini-Bonelli who states:

> "In a corpus-driven approach the commitment of the linguist is to the integrity of the data as a whole, and descriptions aim to be comprehensive with respect to corpus evidence." [Tognini-Bonelli 2001, p. 84]

Following this basic principle, we work empirically with large quantities of corpus data and generate our linguistic hypotheses and generalizations bottom up. The following steps are crucial to our interpretative practice (cf. [Steyer/Lauer 2007, p. 494]):

- Study of all evidence of the corpus and acceptance of this evidence: We use collocation profiles as well as pattern matching to get a starting point for our analysis that is as close to real life usage of language and as objective as possible.
- Generation of hypotheses on the basis of the evidence: We take interactive steps in formulating and refining pattern matching queries to study the evidence.
- Empirical checking of those hypotheses: We check the results of our queries for plausibility and revise if necessary.
- Generalization leads to usage rules: In our model, generalization happens on several hierarchical levels and is detailed by narrative comments if necessary. Usage is always the key factor for justifying generalization.

We have a broad concept of MWE, which is heavily influenced by experience with empirical language data and centers around usage. Our German label - Usuelle Wortverbindungen (first used in [Steyer 2000]) - can be translated as 'MWE which are common in usage'. We adhere to Firth's contextual theory of meaning, here summarised by Tognini-Bonelli:

> "In the Firthian framework the typical cannot be severed from actual usage, and 'repeated events' are the central evidence of what people do, how language functions and what language is about." [Tognini-Bonelli 2001, p. 89]

Following this idea, we regard MWE as conventionalized patterns of language use that manifest themselves in recurrent syntagmatic structures and have acquired a distinct function in communication (cf. [Feilke 2004]; cf. [Brunner/Steyer 2007], [Steyer/Brunner 2009]). They must have at least two concrete lexical components, but may also contain abstract components representing a certain subset of lexical items or even a general grammatical class. Neither idiomaticity nor idiosyncrasy on a grammatical or lexical level is a necessary criterion for MWE in our definition. MWE can have a perfectly regular structure, as long as they work as functional chunks in language use.

Our approach to analysis is similar to that of Hanks detailed in the description of his Corpus Pattern Analysis (CPA):

> "Concordance lines are grouped into semantically motivated syntagmatic patterns. Associating a 'meaning' with each pattern is a secondary step, carried out in close coordination with the assignment of concordance lines to patterns. The identification of a syntagmatic pattern is not an automatic procedure: it calls for a great deal of lexicographic art. Among the most difficult of all lexicographic decisions is the selection of an appropriate level of generalization on the basis of which senses are to be distinguished." [CPA]

CPA aims at describing single words (cf. also [Hanks 2008]), while we are interested in MWE, which adds an additional level of complexity as identifying the surface form itself requires an interpretative effort. To handle the difficulties of generalization, our model has several hierarchical levels, which will be presented below.

2   Model of analysis

2.1   Prerequisites

The basis of our work is collocation profiles, computed from a very large corpus of written German, DeReKo (Deutsches Referenzkorpus), which currently consists of over three billion tokens (cf. [KLa]). The sophisticated method used for generating these profiles was developed by Cyril Belica ([Belica 1995]). It takes a target word form as input and computes the word forms that appear in the vicinity of this target word more often than statistically expected.

These partner word forms are clustered on multiple levels. The KWIC (key word in context) lines from the corpus are grouped into collocation clusters, according to the word forms they contain. The total of all clusters generated for a target word form is called its collocation profile. Collocation analysis is available for the IDS corpora via the COSMAS II corpus research tool (`http://www.ids-mannheim.de/cosmas2`) and can be customized in various ways. For details see [KLb].

LLR	kumul.	Häufig	links	rechts	Kookkurrenzen	syntagmatische Muster
69527	597	597	-1	-1	**taube** gestoßen	100% auf taube [...] Ohren [...] gestoßen
	983	386	-1	-1	taube stoßen	67% auf taube Ohren [...] stoßen
	1206	223	-1	-1	taube stießen	73% stießen ... auf taube Ohren
	4630	3424	-1	-1	taube	99% auf taube [...] Ohren
58806	4634	4	-1	-1	**offene** gestossen Anliegen	100% Anliegen ... auf offene Ohren gestossen
	4747	113	-1	-1	offene gestossen	100% auf offene Ohren [...] gestossen
	4751	4	-1	-1	offene stiess Anliegen	50% Anliegen stiess auf offene Ohren
	4856	105	-1	-1	offene stiess	75% stiess ... auf offene Ohren
	4968	112	-1	-1	offene Anliegen	64% mit ... Anliegen [... auf] offene Ohren
	9418	4450	-1	-1	offene	99% auf offene [...] Ohren
39570	9613	195	-5	5	**Augen** offen halten	74% Augen [und] Ohren [...] offen [zu] halten
	10018	405	-5	5	Augen offen	92% Augen [und] Ohren [...] offen halten
	10019	1	-5	5	Augen offenhalten halten	100% Augen ... Ohren offenhalten ... halten
	10073	54	-5	5	Augen offenhalten	96% Augen und Ohren [...] offenhalten und
	10298	225	-5	5	Augen halten	69% Augen [und] Ohren [offen zu] halten
	16266	5968	-5	5	Augen	81% Augen [und] Ohren
12095	16268	2	1	1	**gekommen** noch sind Klagen	50% noch ... Klagen ... Ohren gekommen sind
	16281	13	1	1	gekommen noch sind	76% sind [...] noch keine|nicht Beschwerden zu Ohren gekommen
	16290	9	1	1	gekommen noch Klagen	100% noch keine Klagen zu Ohren gekommen
	16434	144	1	1	gekommen noch	92% noch nicht zu Ohren [...] gekommen
	16443	9	1	1	gekommen sind Klagen	55% sind ... keine Klagen zu Ohren gekommen
	16558	115	1	1	gekommen sind	57% sind ... keine ... zu Ohren gekommen
	16602	44	1	1	gekommen Klagen	100% noch keine Klagen ... zu Ohren gekommen
	18150	1548	1	1	gekommen	99% zu Ohren [...] gekommen
11769	18257	107	1	1	**gestoßen** sind	76% sind wir auf offene|taube Ohren [...] gestoßen
	19276	1019	1	1	gestoßen	99% auf taube|offene Ohren [...] gestoßen

Fig. 1. Collocation profile of *Ohren* (computed 29 July 2009 via COSMAS web)

Starting point for our study of MWE is the collocation profile of a target word form. Figure 1 shows a snapshot from such a profile for the word form *Ohren [ears]*. Though collocation analysis offers lemmatization, we do not use this setting, neither for the target word form nor when computing its collocates. This is because empirical research shows that the behavior and contexts of different realizations of a lemma are often quite different. These distinctions would be obfuscated by lemmatization. Also, lemmatization of a word form is already an abstraction and we want to be very careful not to make assumptions. In this respect, we follow Sinclair who pointed out:

> "There is a good case for arguing that each distinct form is potentially a unique lexical unit, and that forms should only be conflated into lemmas when their environments show a certain amount and type of similarity." [Sinclair 1991, p. 8]

Consequently, we study the profiles of several word forms which belong to the same lemma separately and make generalizations only at a much later stage of analysis.

We thus start with a collection of KWIC lines which contain the non-lemmatized target word form (e.g. *Ohren [ears]*) and are grouped according to the collocates that have been identified for this target word. This gives us a very good starting point, as the statistical method has already detected regularities within the data with extremely few a priori assumptions. Our goal now is to make use of the information given by corpus data and statistical analysis in a structured and controlled way. We have designed four different levels of abstraction:

– The level of surface patterns
– The level of Lexical realizations (LR)
– The level of MWE
– The level of MWE patterns

Each of these levels builds on the previous and on each level we work manually when grouping, correlating and commenting on the phenomena we observe. The model is thus based on automatically pre-structured data, but is itself a strategy of controlled human interpretation.

As a main example in this paper, we will look at the MWE *Musik in den Ohren*, literally translated as *music in the ears*, which is similar to the English MWE *music to the ears. Musik* is a significant collocation partner of *Ohren* and the KWIC lines containing these two word forms in an appropriate distance are grouped in a collocation cluster. This cluster is the starting point for our analysis. Table 1 shows an excerpt of the relevant KWIC lines.

```
M06      "klingt wie Musik in meinen        Ohren",    sagte die Sozialbürgermeisterin und
MLD      "Das ist Musik für unsere          Ohren",    sagt Ursula Schmitz. Musik, von
S94      Die Musik kann für westliche       Ohren      eine Qual sein; sie ist
O94      Chancen.. ." Musik in unseren      Ohren.     Doch auch wenn uns wohltut,
O94      Uhren,                             Ohren,     Musik In der Grazer ESC
O94      wahre Musik in Horst Fendrichs     Ohren.     "Es leben die PS.. ."
O94      das war Musik in den               Ohren      des Akustik-Sachverständigen. Auch der
O95      Motoren wieder Musik in seinen     Ohren.     Die ersten vier
O95      Sonderling, Künstler, der "die     Ohren      voll Musik hat und den
O95      sind auch Musik in den             Ohren      des Liebespartners. Was gehört noch
O95      Lust sind Musik in den             Ohren      des Liebespartners
O96      ist wahre Musik in seinen          Ohren:     Seine Liebe zur Eisenbahn ließ
O96      Gästen. Musik in Carsten Kelms     Ohren,     denn wenn's bei denen im
O96      narrative Musik halten Augen und   Ohren      immer wieder lustvoll auf Trab.
O97      - zeitgenössische Musik für junge  Ohren      (Ossiach, 13. bis 16. 7.).
O97      Während es etwa in den             Ohren      vieler wie Musik klingt, wenn
O97      Musik in des versöhnten Ombudsmans Ohren!     Die Starparade ist der Höhepunkt
O97      euch" wie Musik in den             Ohren.     Aber die Papst-Visite wurde ebenso
O97      angekündigt. Musik in meinen       Ohren.
O99      der lauten Musik kurz die          Ohren      zuhielt... Gefeiert wird in St.
```

Table 1. KWIC lines from the collocation cluster *Ohren - Musik*

2.2   The level of surface patterns

On this level the KWIC lines which have been grouped by collocation analysis are subjected to further structuring. For this we use a query syntax based on

regular expressions. The queries are used to identify and group lines with a similar syntagmatic structure which serve then as a basis for the analytic steps that follow.

This step is necessary, as our definition of MWE calls for a common syntagmatic structure of the instances of an MWE while collocation analysis looks at word form surfaces without considering the syntactic connection between them. Thus, it sometimes sorts instances of different MWE which share the same lexical material into a single cluster or assigns instances of the same MWE which have different lexical material (e.g. because of orthographical variance or different realizations of the same lemma) to different clusters. Humans, as opposed to the computer, can decide which surface similarities are important for the task of identifying and describing MWE and formulate surface patterns designed to gather the correct instances.

Pattern matching is also a valuable asset when exploring the variability of an MWE. The patterns can be formulated more or less restrictively and the researcher can observe how many KWIC lines - i. e. instances of realization of the MWE in the corpus - are captured. It is also possible to define gaps in the patterns and study the fillers for these gaps. This gives a very good indication what features are really important for the structure of a MWE, how it can be modified and what is its core meaning.

The definition of surface patterns is an iterative process. Often the form of the patterns has to be adapted when observing the results of the previous try. This process reflects an ever deepening understanding of what is relevant in the MWE structure.

The cluster *Ohren - Musik* in Table 1 above is a good example for a collocation cluster which contains instances of different MWE, for example *westliche Ohren [western ears]* or *die Ohren zuhalten [to cover the ears]*. Here are some examples of search patterns which are used to filter out instances which are relevant for the description of the MWE *Musik in den Ohren*. (`#*` stands for an arbitrary number of unspecified word forms; `N(den)` stands for "not *den*"; `ist|war` stands for "*ist* or *war*".)

(1) `Musik in #* Ohren`
(2) `Musik in den Ohren`
(3) `Musik in #* N(den) Ohren`
(4) `Das ist|war Musik in #* Ohren`

You can see that the search patterns differ in their restrictiveness and that, though for sake of simplicity the 'name' of the MWE was given above as *Musik in den Ohren*, not all realizations actually take exactly this form.

Search pattern 3 is specifically designed to capture all realizations that do *not* use the definite determiner *den*. We can now examine the hits of this search pattern and learn from the filler list for the gap that possessive pronouns and genitive forms referring to persons are also common when this MWE is realized. However, frequency counts show that the realization with *den* appears almost three times as often in our corpus.

Search pattern 4 is an example of a very specific pattern. It captures a common way of using the MWE which is very stable, though it covers only a relatively small section of all instances of the MWE in our corpus.

2.3   The level of Lexical realizations

The level of Lexical realizations (LR) is a step between the surface patterns and the actual MWE. LR represent typical realizations of an MWE in the corpus. This intermediate level of analysis has been introduced to account for the fact that MWE are very flexible and subject to much variation. When generalizing immediately to a single typical form, many of these nuances would be lost. LR focus on different kinds of realization of the same MWE and offer a chance to comment on them.

From empirical experience, we have defined several types of LR. This typology is quite general and reflects very basic mechanisms of language. The types will be presented below and exemplified by the LR of the MWE *Musik in den Ohren*. Figure 2 illustrates how the KWIC lines from the collocation cluster are assigned to several LR (each subsuming the hits of one or more search patterns) and how the LR tree for this MWE is built up.

Core LR We assume that there is a core structure which is necessary for the MWE to be recognizable. This structure is captured in the Core LR. Often, the structure of the Core LR is also the most general and subsumes the largest number of KWIC lines, i.e. instances of the MWE in the corpus.

Example: The Core LR *Musik in den Ohren* subsumes more than half of all the instances of the MWE and can thus be regarded as the most common realization - especially as the surface forms of alternative realizations are not as stable.

Core Variant LR The core can be subject to variations on the surface, which are documented in the Core Variant LR. The Core Variant LR is defined relative to the Core LR and differs from its structure in at least one respect.

Example: The Core Variant LR *Musik in N(den) Ohren* subsumes all cases in which the determiner *den* is not present. The fillers for the search pattern gap `N(den)` are presented prominently in form of a filler list and it becomes clear that especially possessive pronouns or genitive forms referring to persons take the place of the determiner. It is justified to differentiate between a Core LR and a Core Variant LR instead of just defining a more general Core LR with the search pattern `Musik in #* Ohren`, because it allows us to highlight the fact that these two different types of realization exist and to show their relative frequencies and nuances in meaning.

Extension LR The structure of the Core LR can be extended by additional elements, which are not mandatory for the structure of the MWE, but are still
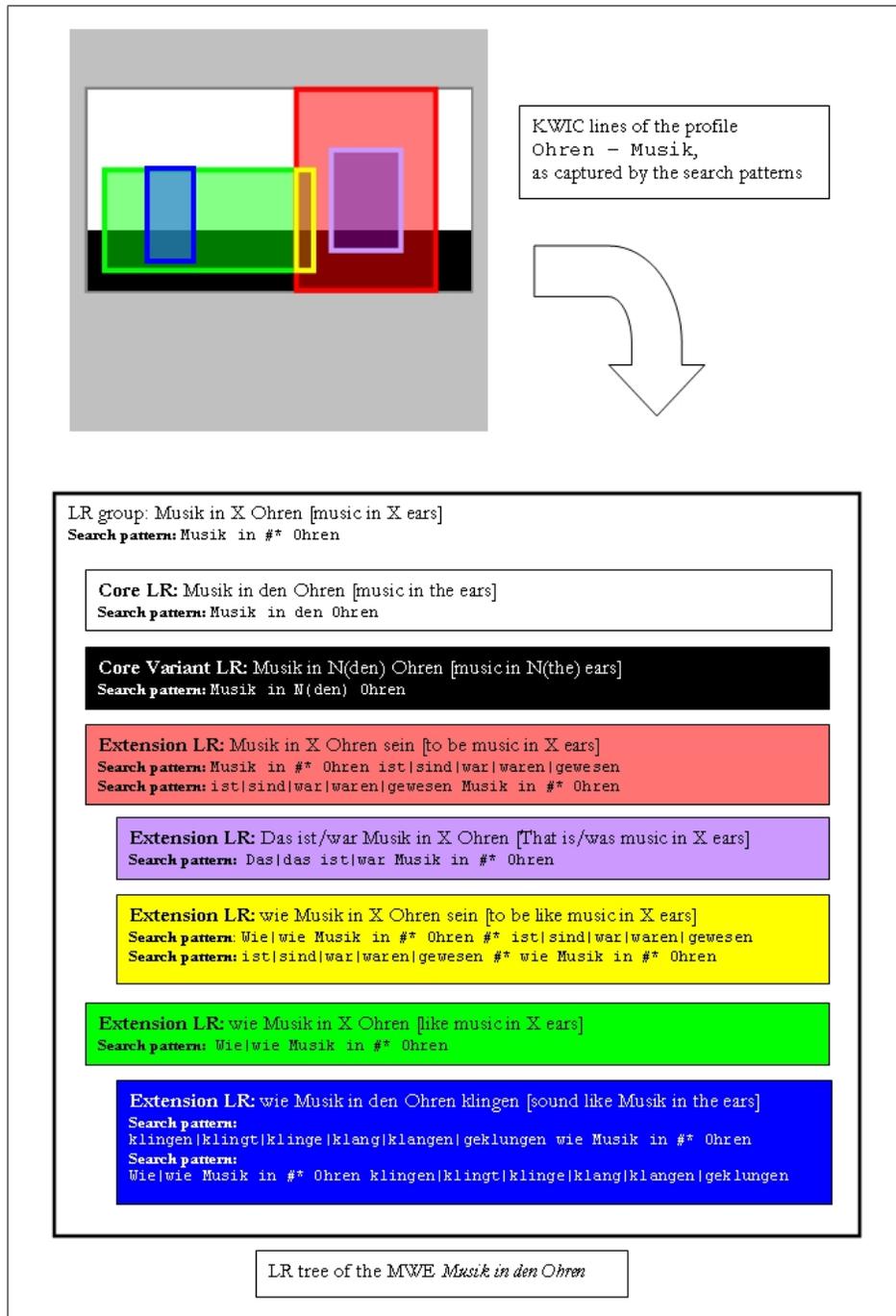
Fig. 2. LR structure of the MWE *Musik in den Ohren* in relation to the KWIC lines of the cluster *Ohren - Musik*

frequent and typical for the way the MWE is used. There can be internal extensions, which appear between the elements of the core structure, or external extensions, which are added before or after the core. Extensions have to be connected to the core syntactically. They are for example verbal constructions, modifiers, object extensions or prepositional phrases.

Example: There are several Extension LR in the LR tree for the MWE *Musik in den Ohren*. All of them extend the structure captured by the Core LR and Core Variant LR in different ways. As you can see in Figure 2, some of them demonstrate that an LR can subsume more than one search pattern. Usually search patterns that only account for word order or grammatical variances are bundled together, unless one of them is extremely prominent or idiosyncratic.

On the first level, there are the LR *wie Musik in X Ohren* and the LR *Musik in X Ohren sein*. Those LR have in common that they add one element to the core structure. Note that they both capture instances of the Core LR structure as well as of the Core Variant LR structure - the difference between those structures has been highlighted already, so it is not necessary to separate them in the Extension LR. The variable component is marked in the name of the LR by the letter X.

Both Extension LR have subordinate Extension LR which further differentiate the structure. LR *wie Musik in X Ohren* can be extended to *wie Musik in X Ohren klingen*. This LR is interesting because it is in fact a combination of two MWE: *Musik in den Ohren* and *in den Ohren klingen [resound in the ears]*. *Musik in X Ohren sein* can be extended to *Das war/ist Musik in X Ohren* and *wie Musik in X Ohren sein*. In fact, the latter LR could also be defined as a child element of *wie Music in X Ohren* as it represents a combination of the two extensions.

Context LR These LR serve as a focus on typical contexts in which the MWE is used. They highlight word forms that commonly appear close to the MWE, but are not directly connected to its structure (as opposed to the elements of the Extension LR). Context LR show the associative frame of the MWE and are thus useful to understand the pragmatics of its usage.

Example: The MWE *Musik in den Ohren* does not have Context LR. However, a typical Context LR would be *Töne ... das menschliche Ohr* which belongs to the MWE *das menschliche Ohr [the human ear]* and subsumes the search patterns `Töne #* das menschliche Ohr` and `das menschliche Ohr #* Töne`. The Context LR highlights a word form, *Töne [sounds]*, which appears significantly frequently in the vicinity of the MWE's Core LR. This is an indicator that the MWE typically refers specifically to the human ability of hearing as opposed to other characteristics of the human ear.

LR tree and LR group The different types of LR can be arranged in a hierarchical structure and may be assigned a narrative comment explaining their specifics. Together, they give a differentiated picture of the MWE in its realizations according to the observed corpus evidence.

Apart from the specialized LR types listed above, there is also the LR group. It serves as a container for collecting all relevant instances of an MWE at once and is always used as the trunk of an LR tree. It captures the overall frequency of the MWE and also preserves instances of realization which are not frequent enough to be highlighted by specialized LR, but may still be of interest for a researcher working with our results.

2.4   The level of MWE

An MWE subsumes an LR tree and is assigned a paraphrase that describes a generalized meaning. The special nature of MWE in our approach becomes evident at this point: An MWE is not a static form, but a complex set of realizations from which a common communicative meaning emerges.

As our approach is based on collocation profiles of target word forms, we established the rule that an MWE must contain at least two concrete and immutable lexical elements - the target word form and one collocation partner. This may seem a somewhat artificial restriction, but it helps greatly in structuring the rich and often overlapping structures that can be detected when analysing language in this manner.

To account for the complex interrelations between MWE, it is possible to define links between them. MWE which have a similar or opposing meaning are connected and their relationships are commented on. Also, structural overlap between MWE is pointed out. Links can be defined between MWE which arise from the same collocation profile (and contain the same target word form), but also between MWE from different profiles.

The example MWE *Musik in den Ohren* is assigned the LR tree shown in Figure 2 and a paraphrase if its general meaning is added: "This MWE is used to express that something is received positively and considered pleasing or beneficial."

Also, links to related MWE are defined. Within the same profile these are for example the MWE *in den Ohren klingen [to resound in the ears]*. As mentioned above, these two MWE are commonly combined. There is also a link to the MWE *Musik für X Ohren [music for X ears]* which has a quite similar meaning, but is more often used to refer to actual musical preferences. In addition to that, there is a connection to a different profile, that of the singular form *Ohr [ear]* where the MWE *Musik in X Ohr [music in X ear]* can be found, which is nearly identical in meaning, but much less frequent.

2.5   The level of MWE patterns

MWE patterns are abstractions over several MWE. This level accounts for the fact that structures in language can be much more general than the restrictions we imposed on MWE allow for.

Much regard is given in our model to mutable elements in fixed structures, which also appear on the level of surface patterns and LR. On the level of MWE patterns, this concept is brought to a higher degree of abstraction: MWE

patterns generalize over MWE which are structurally similar, but different in some aspect of their lexical structure. Therefore these patterns always contain underspecified components.

Two types of MWE patterns can be distinguished. In the first case, the realizations of the underspecified components can be regarded as synonyms. The MWE pattern structurally generalizes over MWE which carry essentially the same meaning.

In the second case, the realizations of the underspecified components are a set of dissimilar lexical items. Each MWE subsumed by the MWE pattern has a different communicative function in its own right, but they share a common core meaning which can be assigned to the more generalized structure.

MWE patterns are especially interesting from the point of view of construction research as they illustrate the transition from concrete lexical items to abstract structures.

<div style="border:1px solid black; padding:10px;">

MWE pattern: aus ADJEKTIVE Gründen [for ADJEKTIVE reasons]

MWE pattern: aus ADJEKTIVE **DOMAIN** Gründen
[for ADJEKTIVE **DOMAIN** reasons]

MWE: aus politischen Gründen [for political reasons]

MWE: aus finanziellen Gründen [for financial reasons]

MWE: aus technischen Gründen [for technical reasons]

</div>
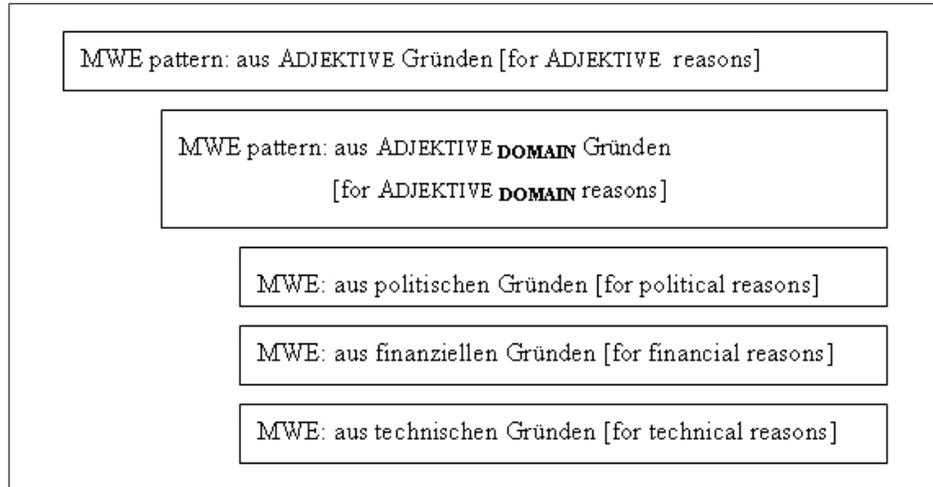
Fig. 3. Excerpt from the hierarchy of the MWE pattern *aus ADJECTIVE Gründen*

An example for an MWE pattern of the second type is the structure *aus ADJECTIVE$_{DOMAIN}$ Gründen [for ADJECTIVE$_{DOMAIN}$ reasons]*. This MWE pattern subsumes several MWE where the underspecified component is realized by a specific lexical item and which each have a different meaning. However, a communicative function which is shared by all its child MWE can be attributed to the MWE pattern: "Using this pattern makes the actions that are explained seem official and at the same time allows the speaker to be vague about the reasons for these actions by using the less specific plural form (*Gründen [reasons]*) which is mandatory for its structure."

Of course, *aus ADJECTIVE$_{DOMAIN}$ Gründen* is itself a specialization of the even more general MWE Pattern *aus ADJECTIVE Gründen.* The meaning

assigned to this most abstract MWE pattern is justification in a general sense. Figure 3 shows part of the hierarchical structure of MWE patterns and MWE.

3   Presentation and prospects

In our 'Wortverbindungsfeldern' (MWE fields) [Steyer/Brunner 2008] we have created a hypertext view which reflects the steps of interpretation detailed in this paper and gives access to the corpus data the analysis is based on. In this way, it is possible to completely retrace our interpretative steps and decisions in generalization. The presentation is enriched by comments on the meaning and specifics of MWE patterns, MWE and LR.

At the moment, the focus of our research is on the development of the model of analysis. We are planning to extend the network of interrelations especially between MWE that originate from different collocation profiles and expand the level of MWE patterns. Also, annotations for MWE should be introduced, which would allow dynamic grouping and different views.

Another research question is how to adapt the in-depth method for a larger scale description of MWE and to develop modes of presentation suitable for different user groups - for example learners of German versus researchers interested in constructions.

Bibliography

[Belica 1995] Belica, Cyril (1995): Statistische Kollokationsanalyse und Clustering. Korpuslinguistische Analysemethode. Mannheim. Internet: `http://www.ids-mannheim.de/kl/projekte/methoden/ur.html`[1]

[Brunner/Steyer 2007] Brunner, Annelen / Steyer, Kathrin (2007): Corpus-driven study of multi-word expressions based on collocations from a very large corpus. In: Proceedings of the 4th Corpus Linguistics Conference, Birmingham. Internet: `http://corpus.bham.ac.uk/corplingproceedings07/paper/182_Paper.pdf`[1]

[CPA] Corpus Pattern analysis. Internet: `http://nlp.fi.muni.cz/projekty/cpa/`[1]

[Hanks 2008] Hanks, Patrick (2008): Lexical Patterns. From Hornby to Hunston and Beyond. In: Bernal, Elisenda / DeCesaris, Janet (eds.): Proceedings of the XIII Euralex International Congress, Barcelona, 15-19 July 2008. Barcelona, p. 89-129.

[Feilke 2004] Feilke , Helmuth (2004): Kontext - Zeichen - Kompetenz. Wortverbindungen unter sprachtheoretischem Aspekt. In: Steyer, Kathrin (ed.): Wortverbindungen - mehr oder weniger fest. Jahrbuch des Instituts für Deutsche Sprache 2003. Berlin/New York.

[KLa] Projektseite Ausbau und Pflege der Korpora geschriebener Gegenwartssprache: Das Deutsche Referenzkorpus - DeReKo. Internet: `http://www.ids-mannheim.de/kl/projekte/korpora/`[1]

[KLb] Projektseite Methoden der Korpusanalyse- und erschließung. Internet: `http://www.ids-mannheim.de/kl/projekte/methoden/`[1]

[Sinclair 1991] Sinclair, John (1991): Corpus, Concordance, Collocation. Oxford.

[Steyer 2000] Steyer, Kathrin (2000): Usuelle Wortverbindungen des Deutschen. Linguistisches Konzept und lexikografische Möglichkeiten. In: Deutsche Sprache 28, 2. Berlin, p. 101-125.

[Steyer 2004] Steyer, Kathrin (2004): Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikografische Perspektiven. In: Steyer, Kathrin (ed.): Wortverbindungen - mehr oder weniger fest. Jahrbuch des Instituts für Deutsche Sprache 2003. Berlin/New York, p. 87-116.

[Steyer/Brunner 2008] Steyer, Kathrin / Brunner, Annelen (2008): Wortverbindungsfelder. Abrufbar unter "Wortverbindungen online". Internet: `http://www.ids-mannheim.de/ll/uwv/wvonline/wvfelder`[1]

[Steyer/Brunner 2009] Steyer, Kathrin / Brunner, Annelen (2009): Das UWV-Analysemodell. Eine korpusgesteuerte Methode zur linguistischen Systematisierung von Wortverbindungen. (= OPAL - Online publizierte Arbeiten zur Linguistik 1/2009). Mannheim. Internet: `http://www.ids-mannheim.de/pub/laufend/opal/privat/opal09-1.html`[1]

---

[1] Web sites were checked on 29 July 2009.

[Steyer/Lauer 2007] Steyer, Kathrin / Lauer, Meike (2007): "Corpus-Driven":
Linguistische Interpretation von Kookkurrenzbeziehungen. In: Kämper, Hei-
drun / Eichinger, Ludwig M. (eds.): Sprach-Perspektiven. Germanistische
Linguistik und das Institut für Deutsche Sprache. (= Studien zur deutschen
Sprache 40). Tübingen, p. 493-509.

[Tognini-Bonelli 2001] Tognini-Bonelli, Elena (2001): Corpus Linguistics at
Work. (= Studies in Corpus Linguistics 6). Amsterdam/Philadelphia.