

Exemplars and analogy: Semantic extension in constructional networks

Arne Zeschel

Abstract

How (and when) do speakers generalise from memorised exemplars of a construction to a productive schema? The present paper presents a novel take on this issue by offering a corpus-based approach to semantic extension processes. Focusing on clusters of German ADJ N expressions involving the heavily polysemous adjective *tief* 'deep', it is shown that type frequency (a commonly used measure of productivity) needs to be relativised to distinct semantic classes within the overall usage spectrum of a given construction in order to predict the occurrence of novel types within a particular region of this spectrum. Some methodological and theoretical implications for usage-based linguistic model building are considered.

Keywords: semantic extension, productivity, type frequency, Correlation Analysis, usage-based model

1. Introduction

How (and when) do speakers generalise from specific learned expressions of a particular type to a productive constructional schema?¹ Usage-based approaches to language generally emphasise that these two notions are in fact endpoints on a continuous scale. Still, positing a blurred boundary between lexical and grammatical constructions does not in itself explain how speakers derive the latter from the former. For instance, consider the status of the putative schema in (1.a) in view of the acceptability of the expressions in (1.b-d):²

- (1) a. *to go ADJ*
b. *to go mad / bankrupt / quiet ...*
c. *to go crazy / insane / mental ... broke / bust / ?skint... calm / silent / ?still...*
d. *to go ?famous / ?popular / ??renowned ...*

In a paper about a group of similar constructions in Spanish, Bybee and Eddington (2006) argue that acceptability reflects familiarity: an expression is acceptable to the extent that it is a.) directly retrievable from memory as

a stored unit (as an effect of high input frequency), or b.) semantically similar to such a stored unit.

This account offers a both simple and intuitively plausible explanation for the differences in idiomatic feel or ‘naturalness’ between the routine formulas in (1.b), their close semantic variants in (1.c), and the less obviously related verb-adjective combinations in (1.d). As it stands, however, it has little to say about the status of (1.a): on the one hand, the dubious acceptability of (1.d) suggests that the adjective position in this pattern is not a fully general ‘open slot’. On the other hand, the verb *go* nevertheless *does* occur with several dozens of adjectives in such change-of-state expressions in the British National Corpus, and many of these combinations have a distinctly unconventional ring to them (suggesting that they are in fact creative formations that were computed from scratch rather than stored and retrieved ‘fixed expressions’). So where do these come from? Since a fully general productive schema [*go* ADJ] does not seem to exist, it would appear that such combinations are formed in analogy to specific existing expressions of the relevant type, and this is indeed what Bybee and Eddington conclude from an analysis of their Spanish data. However, this answer raises an important new question: if semantic extension proceeds via analogy to entrenched exemplars, then how is it that some high-frequency expressions are easily extended in this way (cf. 2.a), whereas others permit considerably less or even little to no such variation at all (cf. 2.b, c)?

- (2) a. *to go mad / crazy / insane / wild / furious / ballistic ...*
 b. *to go bankrupt / broke / bust / ?skint / ?insolvent / ?illiquid ...*
 c. *to go cold / ?cool / ?freezing / ?icy / ?chilly / ?frosty ...*

One idea is that the crucial additional property is *semantic type frequency* (Israel 1996): on this account, the likelihood (in production) and acceptability (in comprehension) of a particular novel extension depends on the *amount* of similar expressions that are already known to the speaker. The present paper evaluates this hypothesis with a quantitative corpus study of a group of idiomatically restricted collocations. Focusing on figurative adjective-noun expressions involving the heavily polysemous adjective *tief* (‘deep’) in German, it introduces a corpus-based approach to the reconstruction of semantic extension processes in constructional networks. The prediction to be evaluated is the following:

- (3) *Semantic type frequency hypothesis*
 The likelihood/acceptability of a particular novel extension of an established collocation depends on the number of previously encountered variants of the same semantic type

The paper is structured as follows: I begin with a brief sketch of the theoretical motivations for the above assumption (section 2) and introduce methods and data of the study (section 3). Section 4 evaluates the hypothesis on three successively more detailed levels of semantic analysis. Section 5 offers a brief summary and conclusion.

2. Semantic extension: A usage-based account

Especially in morphology, type frequency has long been recognised as a useful measure of productivity (Bybee 1985, 1995): on this account, a good indication that e.g. affixation with *-s* is a more productive strategy for expressing the meaning PLURAL in English than affixation with *-en* is that there are substantially *more* plural formations (types) that involve an allomorph of *-s* (e.g. *box* > *boxes*) than in the case of *-en* (e.g. *ox* > *oxen*). Since usage-based approaches see no categorical difference between morphology and syntax in principle (in the sense that they involve qualitatively different representations or processing operations), an extension of these findings to word combinations in syntax would appear natural. On the other hand, even in a model that sees lexicon and syntax as a continuum, what is a useful measure for a highly generalised schema such as [N-*s*] may well be less appropriate for a much less inclusive (and in fact maybe just hypothetical) generalisation such as aspectual [*go* ADJ]. Specifically, it has been argued that speakers normally take both formal and semantic properties into account when they assess the combinability of two linguistic elements, and it is only in the limiting case of maximally generalised semantic restrictions that a consideration of formal properties alone will do (Zeschel 2007). It follows that for less generalised constructions such as the collocations investigated in this paper, type frequencies should be counted per semantically distinct usage class rather than for a given form in aggregate: for instance, for categories that could be approximated as [*go* ADJ_{-NORMAL MENTAL PROCESSING}] (cf. 2.a) and [*go* ADJ_{-ABLE TO PAY}] (cf. 2.b) in separate measures rather than for [*go* ADJ_{PROPERTY}] at large. Insofar as such classes have in fact been abstracted and one is not dealing with a genuinely isolated expression, the current approach is therefore sympathetic to Erman and Warren's (2000: 41) assumption that "what we store in some cases [of slots in 'idioms' and 'fixed expressions'] is a meaning rather than a specific word."

Apart from that, the present study embraces the usage-based assumption that speakers retain memory traces of (various aspects of) fully specific

utterance tokens that they experience (i.e. exemplar representations of particular expressions), but also extract generalisations from these memorised units (i.e. schemas) which may become entrenched routines themselves (cf. Abbot-Smith and Tomasello 2006; Goldberg 2006, Bybee 2006). Following Langacker (2000), such schemas can be seen as reinforced commonalities between co-activated exemplars: every time a set of different units is activated in language processing because they are all relevantly similar to a given target, those (and only those) properties of the activated exemplars that are shared will be jointly reinforced. Over time, the repetition of this process may produce a stable pattern that is not identical to, but nevertheless completely immanent in, its individual instantiations. Put differently, schemas are emergent generalisations over a set of relevantly similar instances (including lower-level schemas) that are repeatedly co-activated in processing. Second, clustering exemplars for similarity (i.e. the crucial prerequisite for later schema extraction) is assumed to be an automatic by-product of language processing: metaphorically speaking, hearers/language learners are assumed to either map an incoming target onto an already existing identical representation in their repertoire, strengthening it, or if none such unit is available, to store it 'next to' the most similar existing unit in their system, creating a new representation (Bybee 2006). Over time, growing exposure to expressions that are in some respect similar to already encountered ones therefore creates progressively dense exemplar clusters, with various local schemas/similarity classes emerging from the numerous individual categorisations that were involved.

Applied to the question of semantic extension that is investigated in this paper, I assume that different variants of a given construction may start out (again metaphorically speaking) from different regions of semantic space (due to e.g. distinct conceptual motivations). Driven by speakers' changing communicative needs and the demand of modelling novel expressions on closely related established ones, such variants may then begin to expand in different semantic directions analogically. At some point, such a cluster of related expressions may come into contact with a hitherto distinct cluster of other such expressions, which then permits the formation of a more inclusive schema. The result is an interplay of analogical extension and subsequent schema extraction such that "[l]ong strings of analogical extensions lead to discrete clusters of usage, which then license the extraction of more abstract schemas for the construction" (Israel 1996: 220).

3. Case study: Data and method

The present section describes how the data used in this study were extracted, postprocessed, coded, and analysed. Recall from section 1 that the overall aim of the study is to predict the distribution of *novel expressions* (of a specific semantic type) from the distribution of *established expressions* (of the same semantic type) across the relevant regions of semantic space (i.e. those that are covered by attested instances), and to do so on three different levels of semantic analysis: first, on the level of all figurative uses of the German adjective *tief* in attributive adjective-noun contexts at large (subdivided into individual figurative senses; cf. 4.1); second, on the level of the most frequent of these figurative senses (subdivided into different conceptual mappings that give rise to this sense; cf. 4.2); and finally, on the level of the most frequent conceptual mapping within this sense (subdivided into individual semantic fields that are covered by the attested instances; cf. 4.3). On all three levels, it is predicted that the distribution of novel adjective-noun combinations across semantic space is not random, but that it will correlate positively with the number of established adjective-noun combinations in different regions of the overall semantic map. Diagrammatically, therefore, the study progressively ‘zooms in’ on a subset of the data, with each new analysis concentrating on just the most frequent category (solid line circle in figure 1 overleaf) emerging from the previous classification step.

The following pages describe the methodological procedure of the study step by step. To begin with, 10,000 tokens of the target adjective *tief* in all relevant inflectional forms were extracted from PUBLIC/COSMAS-II, a giga corpus of written German. All concordance lines were removed in which the retrieved form of *tief* did not function as an attributive adjective which modified a noun that followed it within a span of up to three words. Due to the high frequency of nominal *Tiefe* in my data, this reduced the data set by more than 50%.

These data were then classified semantically using the lexicographic descriptions of the *Duden Deutsches Universalwörterbuch*, a large corpus-based monolingual dictionary of German.² This dictionary lists seven distinct senses for *tief*, each of them possessing further subsenses. Simplifying things a little, the main distinctions can be glossed and exemplified as in (4):

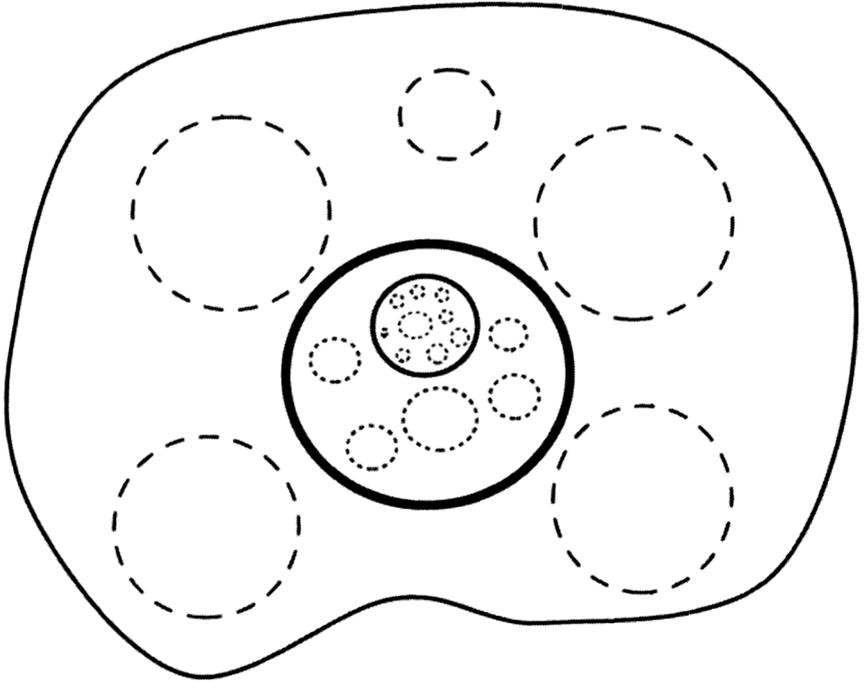


Figure 1. Outline: Classification steps

- (4) a. 'of considerable extension (vertically) downwards, reaching far below'
ein tiefer Abgrund 'a deep abyss'
- b. 'of considerable extension backwards, reaching far behind'
ein tiefer Wald 'a deep forest'
- c. 'of considerable extension inwards, reaching far inside'
eine tiefe Wunde 'a deep wound'
- d. 'temporally advanced, late', 'in the middle of a timespan'
bis tief in die Nacht 'until deep in the night'
- e. 'intense, strong, given to a large extent'
tiefe Trauer 'deep sadness'
- f. 'profound; not superficial, but reaching to the core'
eine tiefe Einsicht 'a deep insight'
- g. 'intense, full, dark' (of colours); 'low, dark' (of voices or sounds)
ein tiefes Rot 'a deep red'

It is not difficult to see how these senses are connected and how some of them might have evolved from others on the list. Consequently, semantic coding decisions were not always entirely straightforward. For instance, is the expression *tiefe Überzeugung* ('deep conviction') an instance of sense 5 (i.e., a *strong* conviction) or an instance of sense 6 (i.e., a *profound* conviction), and does it in fact make sense to draw a line at this point? Because the classification task is really the crucial part of the study, the data were analysed by two separate coders and interrater agreement measures were calculated to assess the reliability of the obtained judgments. Consistency was assessed in two steps: First, the second coder (a linguistically naïve native speaker of German) coded 10% of the data (424 observations), and a first measure of interrater agreement was calculated. Second, all cases in which there was disagreement between coders were inspected again in order to remove differences due to oversight or carelessness, and to resolve cases that had been explicitly coded as 'unclear' by one of the raters. When these were settled, the remaining differences were used to calculate the final agreement score. Interrater agreement was quantified using Brennan and Prediger's free-marginal kappa (κ , cf. Brennan and Prediger 1981), a measure that is less likely to exaggerate the amount of actual agreement between coders than a simple percentage because it also takes the amount of expected chance agreement into account. As a rule of thumb, a kappa of .7 or higher is commonly accepted as an indication of promising interrater agreement.

In addition to sense group membership, each concordance line was coded for whether *tief* was used figuratively (again by both coders, following the procedure laid out above). Instances of senses 4-7 (cf. 4.d-g) were coded as figurative across the board. By contrast, instances of the three spatial senses 1-3 (cf. 4.a-c) were treated as non figurative, with two important provisos: first, the dictionary listed the meaning 'low on a scale' as a subsense of spatial sense 1, 'reaching far downwards, low'. Instances of this subsense (e.g., *tiefe Temperaturen*, 'low temperatures') were classified as figurative throughout. Second, judgments of figurativity were based on assessing the meaning of the composite adjective-noun expression in its full context (i.e. by inspecting each individual concordance line) rather than by considering just the decontextualised noun alone. This was necessary because one and the same adjective-noun combination can be unambiguously concrete in meaning in one context and unambiguously abstract / metaphorical in others:

- (5) a. *Zwei tiefe Wunden mussten genäht werden*
 two deep wounds must sutured become
 'Two deep wounds had to be sutured.'
- b. *In die Seele der Kinder schlägt*
 into the soul of.the children strikes
die Obdachlosigkeit tiefe Wunden.
 the homelessness deep wounds
 'Homelessness causes severe damage to children's souls.'

To elaborate, the point is that if *Wunden* in (5.b) conveys the schematic meaning 'damage' rather than the more specific meaning 'laceration of the skin', it also coerces a fitting more abstract reading on its modifier *tief* (i.e., intensifying 'severe' rather than spatial 'reaching far inwards').

Next, the modified nouns were lemmatised: case, number and gender markers were stripped off, spelling variants were homogenised and diminutive affixes were removed. In contrast to diminutive formations, compounds were left intact, to the effect that words like *Krise*, 'crisis', *Formkrise*, 'form crisis' and *Ehekrise*, 'marriage crisis' etc. were all counted as separate types. The resulting list was then partitioned into 'established' types on the one hand and 'novel' types on the other by using the simplest possible operationalisation: any combination that was *recurrent* was counted as 'established', and any combination that occurred only once was counted as 'novel'.³

Once all combinations were coded for a.) figurativity (yes/no), b.) the particular sense of *tief* that they instantiated (1-7), and c.) their conventionality (established/novel), a possible connection between the semantic distribution of established and novel types could be investigated by submitting type counts for both classes in all five figurative sense categories to a Correlation Analysis. Correlation analysis is a statistical method for assessing dependencies between different measurements of two variables X and Y (here: number of established vs. number of novel types in a given semantic category). Correlations may be negative or positive in polarity, and small ($\pm 0.1-0.3$), medium ($\pm 0.3-0.5$) or large ($\pm 0.5-1.0$) in size. Standard measures such as Pearson's product moment correlation coefficient (the one used here) can be calculated by freely available statistics packages like R or Spreadsheet software such as OpenOffice Calc. In interpreting the results of such an analysis, it is necessary to bear in mind that the existence of, say, a large positive correlation between two variables X and Y does not in itself provide evidence that values for Y in fact depend on those for X: it could just as well be the other way round (Y is the independent variable, and X's

value depends on Y), or both variables may be affected in similar ways by an unknown third variable Z without there being a direct connection between X and Y. However, given the well-established connection between type frequency and productivity, I would argue that the possibility of mixing up cause and effect (or even failing to recognise an underlying ‘true’ predictor) can be neglected here. In other words, if a significant positive correlation between the distribution of established and novel types can be found, this will be interpreted as evidence that creative extensions of different constructional variants are in fact driven by variants’ semantic type frequency. The results of the first classification step and the correlation results on this level of analysis are reported in section 4.1.

Next, the top frequent figurative sense emerging from the classification in step 1 was subject to a more fine-grained semantic analysis. Since it turned out that instances of the same figurative sense could be derived from different non-figurative source senses (and connected to these through different conceptual mappings, cf. 4.2), it seemed useful to refine the established classification first by systematising the data in terms of the different conceptual mappings that gave rise to their figurative interpretation. In this classification step, the second rater coded 50% of data (150 adjective-noun combinations). When the classification was complete, a second Correlation Analysis was conducted that is reported in section 4.2.

Third, one of these mappings emerged as the dominant mechanism that accounted for the lion’s share of expressions within the most frequent figurative sense. In the final classification step, all instances of this mapping were systematised yet further by grouping them into a number of distinct semantic fields. Again, the second rater coded 50% of the data (85 pairs) in order to assess whether the proposed categorisation also made sense to other native speakers, and a final Correlation Analysis was carried out on the level of the identified semantic fields (cf. 4.3).

4. Case study: Results

The following results are based on 4245 tokens (1110 types) of attributive *tief*-N expressions. Exactly two thirds of the data (66.6% of the tokens and 66.8% of the types) were identified as figurative. Interrater agreement on this classification was only moderate in the first classification round ($\kappa_{\text{initial classification}} = .58$) due to an unexpected misunderstanding and following systematic mistake on part of the second coder,⁴ but it rose to a substantial κ_{final}

classification = .82 after the second round. Figurative uses were found to be distributed over 1564 established (55%) and 1266 novel types (45%).

4.1. Classification I: Figurative *tief*

In spite of the difficulties anticipated in section 3, interrater agreement on sense classifications for figurative *tief* was already reasonably high in the first round ($\kappa_{\text{initial classification}} = .77$), and it rose to a kappa of .82 after a reconsideration of the controversial cases. The one principled disagreement that remained (besides a number heterogeneous others) concerned the boundary between senses 5 ('intense, strong') and 6 ('profound') that was already mentioned in section 3. In particular, there was disagreement about the classification of expressions from one specific semantic field, with controversial examples being expressions like e.g. *tiefer Glaube* 'deep faith', *tiefe Frömmigkeit* 'deep piety' and *tiefes Bekenntnis* 'deep confession'.

From a Cognitive Semantic point of view, such difficulties are of course anticipated since the sense distinctions provided by the dictionary do not define clearly delimited categories with a binary membership status. Rather, one would expect members of a category to be connected by a network of family resemblances, with clear instances of category X gradually shading off into clear instances of category Y, and certain instances falling right in between these central types. However, since only instances of the top frequent sense 5 ('intense') were included in the follow-up classifications reported in sections 4.2 and 4.3, it was necessary to make a principled decision here about what should be done with any residual disagreements. Seeing that the second coder had only coded 10% of the data, it was decided that the (final) judgments of the first coder formed the basis for all subsequent analyses, and it is these classifications that are reported in all following results. Table 1 on the following page shows the distribution of recurrent and novel types across the five recognised figurative senses. The results show that intensifying *tief* (e.g. *tiefe Dankbarkeit*, 'deep gratefulness') is both most common (i.e. token frequent) and most widely dispersed semantically (type frequent), followed by uses conveying the meaning 'low on a scale' (e.g. *tiefer Zinssatz*, 'low interest rate'), followed by uses of the type 'profound' (e.g. *tiefes Geheimnis*, 'deep mystery'), followed by metaphors for particular visual and auditory qualities (e.g. *tiefe Stimme*, 'deep voice'), and, finally, a fairly rare temporal sense (e.g. *tiefer Winter*, 'deep Winter'):

Table 1. Figurative *tief*: Distribution of established and novel types across senses

Rank	Sense	Tokens	Types	Est.	Novel
1	5 ('intense')	1456	300	143	157
2	1 ('low on a scale')	636	198	55	143
3	6 ('profound')	442	137	41	96
4	7 ('dark/low, of colour or sound')	261	97	29	68
5	4 ('temporal peak')	35	9	5	4
Total		2830	741	273	468

Moving on to the productivity issue, a Correlation Analysis for the values in the two rightmost columns in table 1 reveals the predicted positive correlation between the number of established and novel types in a given sense class ($r_{\text{Pearson}} = .82$). However, this result is only marginally significant at $p = .086$. In other words, presumably not least because there were only five categories/measurement pairs that were being related, this finding does not yet provide particularly compelling evidence for the hypothesised connection between productivity and semantic type frequency. Moreover, in order to get to the analogy issue, it will be necessary to delve deeper into the data and examine the internal structure of the clusters that make up the individual senses, which is here demonstrated on the example of the most type- and token-frequent class, i.e. *tief* as an intensifier.

4.2. Classification II: Intensifying *tief*

The 300 combination types that were judged as intensity expressions were still rather heterogeneous semantically – both in terms of the meaning of the intensified nouns and how this meaning was connected to a given non-intensifying source meaning of the adjective. Consider the examples in (5):

- (6) a. *tiefe Enttäuschung*, *tiefes Bedauern*, *tiefe Genugtuung*
 deep disappointment deep regret deep satisfaction
- b. *tiefe Krise*, *tiefe Rezession*, *tiefe Not*
 deep crisis deep recession deep affliction
- c. *tiefer Graben*, *tiefe Klüft*, *tiefer Riss*
 deep rift deep chasm deep crevice

These figurative expressions are spin-offs from different non-metaphorical source meanings, and connected to these meanings through different con-

ceptual metaphors. In particular, the collocations with emotion and mental state terms in (6.a) are motivated by an interplay between the metaphors PHYSICAL AND PSYCHOLOGICAL PROPERTIES ARE ENTITIES WITHIN A PERSON on the one hand and CENTRALITY IS IMPORTANCE on the other: the first metaphor allows the construal of abstract properties and sensations as reified objects that are located within a person, and the second metaphor contributes the intensity implication (i.e., it is ‘proximity’ to the metaphorical ‘centre’ of a person that allows the construal of strong and intensely experienced emotions as being *deep-seated*). Expressions of this type are thus metaphorical spin-offs from spatial sense 3, ‘reaching far inward’ (though their ordinary use in contemporary German is arguably better characterised as abstract and scalar rather than as concrete and spatial). The expressions in (6.b) – a class of collocations denoting unfavourable situations – are rather different: motivated by a different metaphor (GOOD IS UP), they also go back to a different spatial sense (i.e., sense 1, ‘reaching far below’). And the expressions in (6.c) are different again: though also spin-offs from sense 1 (‘low, below’), these collocations draw on a yet different metaphor. Here, the mapping is provided by SIMILARITY IS CLOSENESS: the noun denotes something that prevents two dissenting parties from moving towards each other and eventually ending up at the same position, with the addition of *tief* reinforcing the difficulty that any attempt to reconcile the different standpoints would bring.

All in all, nine such classes could be distinguished in the data. The remaining six (exemplified in (7) below) can be characterised as follows: ABSTRACT RELATIONS ARE PHYSICAL CONNECTIONS (7.a), CONTROL IS UP (7.b) and SOUNDNESS IS PHYSICAL INTEGRITY (7.c) in terms of metaphors, the two metonymies BOWING FOR RESPECT (7.d) and RESPIRATION FOR RELIEF (7.e, only two types), and finally a small ‘generalised intensity’ class whose members could not be traced back to any of these original motivations but seemed to have been formed in analogy to existing *tief*-+N intensity expressions (7.f):

(7) a. <i>tiefe Bindung</i> , deep bond	<i>tiefer Halt</i> , deep footing	<i>tiefe Verankerung</i> deep anchoring
b. <i>tiefe Bewusstlosigkeit</i> , deep unconsciousness	<i>tiefe Narkose</i> , deep narcosis	<i>tiefe Trance</i> deep trance
c. <i>tiefe Wunde</i> , deep wound	<i>tiefe Schramme</i> , deep scratch	<i>tiefe Verletzung</i> deep injury

d. <i>tiefe Verbeugung</i> , deep bow	<i>tiefer Kniefall</i> , deep prostration,	<i>tiefer Kotau</i> deep kowtow
e. <i>tiefes Aufatmen</i> , deep respiration	<i>tiefes Durchatmen</i> deep respiration	
f. <i>tiefes Bouquet</i> , deep bouquet	<i>tiefe Atmosphäre</i> , deep atmosphere	<i>tiefe Ungerechtigkeit</i> deepinjustice

Interrater agreement for these classifications was lower than for sense group membership in the first round ($\kappa_{\text{initial classification}} = .69$). In part, this was due to the fact that the second coder was linguistically naïve and there was some confusion about the concept of conceptual metaphors and metonymies. In particular, semantic similarity in the target domain was repeatedly favoured over isomorphism of mapping as the crucial classification criterion – for instance, expressions like *tiefe Ehrerbietung* ('deep reverence'), *tiefe Bewunderung* ('deep admiration') and *tiefe Hochachtung* ('deep respect') were grouped together with *tiefe Verbeugung* or *tiefe Verneigung* ('deep bow') even though only the latter involve the postulated metonymy (BOWING FOR RESPECT). When this was clarified, interrater agreement rose to an extraordinarily high $\kappa_{\text{final classification}} = .98$. Table 2 overleaf shows the distribution of established and novel types across the nine attested mappings. A Correlation Analysis on this level of classification revealed a strikingly large positive correlation between the counts for novel and established types ($r_{\text{Pearson}} = .97$; $p < .001^{***}$).

Table 2. Intensifying *tief*: Distribution of established and novel types across mappings

Rank	Mapping	Types	Est.	Novel
1	CENTRALITY IS IMPORTANCE	170	86	84
2	GOOD IS UP	45	17	28
3	SIMILARITY IS CLOSENESS	28	14	14
4	CONTROL IS UP	20	14	6
5	GENERALISED INTENSITY	16	15	1
6	ABSTRACT RELATIONS ARE PHYS. CONNECTIONS	10	5	5
7	METONYMY: BOWING FOR RESPECT	5	3	2
8	SOUNDNESS IS PHYSICAL INTEGRITY	4	2	2
9	METONYMY: RESPIRATION FOR RELIEF	2	1	1
Total		300	157	143

4.3. Classification III: CENTRALITY IS IMPORTANCE

Again, one class made up for the lion's share of the types in the preceding classification: these were expressions denoting emotions and mental states which involved a combination of the metaphors PHYSICAL AND PSYCHOLOGICAL PROPERTIES ARE ENTITIES WITHIN A PERSON and CENTRALITY IS IMPORTANCE. In the final classification step, instances of this category were systematised further by tracing them back to more or less closely circumscribed semantic fields. The 12 classes emerging from this analysis are exemplified in (8):

(8) a. <i>tiefe Ergriffenheit</i> , deep emotion	<i>tiefe Empfindung</i> , deep sensation	<i>tiefes Gespür</i> deep intuition
b. <i>tiefe Freude</i> , deep joy	<i>tiefes Glück</i> , deep happiness	<i>tiefe Befriedigung</i> deep satisfaction
c. <i>tiefe Verzweiflung</i> , deep despair	<i>tiefe Depression</i> , deep depression	<i>tiefe Reue</i> deep regret
d. <i>tiefe Liebe</i> , deep love	<i>tiefe Zuneigung</i> , deep affection	<i>tiefe Bewunderung</i> deep admiration
e. <i>tiefer Hass</i> , deep hatred	<i>tiefe Abneigung</i> , deep aversion	<i>tiefe Verachtung</i> deep contempt
f. <i>tiefe Besorgnis</i> , deep anxiety	<i>tiefe Angst</i> , deep fear	<i>tiefe Beunruhigung</i> deep disturbance
g. <i>tiefer Wunsch</i> , deep wish	<i>tiefes Bedürfnis</i> , deep need	<i>tiefe Sehnsucht</i> deep longing
h. <i>tiefe Anteilnahme</i> , deep solicitousness	<i>tiefes Mitleid</i> , deep compassion	<i>tiefes Erbarmen</i> deep mercy
i. <i>tiefes Unverständnis</i> , deep lack of understanding	<i>tiefe Missachtung</i> , deep disregard	<i>tief Selbstmissachtung</i> deep self-disregard
j. <i>tiefe Verantwortung</i> , deep responsibility	<i>tiefe Loyalität</i> , deep loyalty	<i>tiefes Pflichtgefühl</i> deep sense of duty
k. <i>tiefe Überzeugung</i> , deep conviction	<i>tiefer Glaube</i> , deep faith	<i>tiefes Vertrauen</i> deep trust
l. <i>tiefe Zweifel</i> , deep doubts	<i>tiefe Skepsis</i> , deep scepticism	<i>tiefes Misstrauen</i> deep distrust

Interrater agreement for this classification was again less than perfect in the first round ($\kappa_{\text{initial classification}} = .68$). Here, the reason was that the analysis had now arrived at a fairly fine-grained level of semantic description: even with the 12 classes in (8) pre-established by the first coder prior to classification (thus suggesting a selection of relevant semantic dimensions in advance), there often remained more than one way of drawing the boundaries between classes in the data. On the other hand, since the respective other coder's judgments were often perfectly plausible, too, there was a strong tendency to align judgments when the criteria for controversial assignments were made explicit in the following discussion, which resulted in a markedly higher kappa for the final classification ($\kappa_{\text{final classification}} = .95$). Taking both tendencies into account, the comparably low initial kappa suggests that the postulated distinctions may not provide optimal fit for the data, whereas the high final kappa indicates that it is nonetheless possible to converge on these categories with substantial agreement once more detailed classification criteria are negotiated between coders. Taken together, this suggests that the postulated groupings are presumably not stable categories but constitute one (albeit plausible, see results of the second classification round) possibility of carving up the semantic space covered by the attested instances, which in reality form clouds that overlap in multiple ways and that are not easily pinned down to a set linearly separable categories (again not surprising from a Cognitive Semantic point of view).

Nevertheless, if the present category system is accepted as one such possible approximation, it is remarkable to see that only 8 out of 170 types (4.7%) could not be assigned to one of these 12 classes but had to be classified as 'other'. Moreover, most of the postulated classes are related to others in semantically transparent ways (notably via antonymy, but also through frame-based shifts and other semantic relations). Even though such intermediate-level patterning might constitute an interesting pointer to possibly recurrent 'pathways' of extension (cf. Zeschel 2007), this issue is beyond the scope of the present paper. For now, suffice it to acknowledge that the investigated collocation clusters show a remarkable degree of regular semantic patterning on different levels of schematicity. Table 3 on the opposite page reports the distribution of established and novel types across the 12+1 classes on this level. As in the case of the previous classification, a Correlation Analysis revealed a large positive correlation between the counts for novel and established types ($r_{\text{Pearson}} = .78; p < .01^{**}$).

5. Summary and conclusion

On all three levels of semantic analysis, large positive correlations were found between the semantic type frequency of recurrent variants ('established uses') of a given expression and the number of singleton types ('creative extensions') in the given cluster: even though the results of the first (sense-level) classification were only marginally significant, they were still pointing in the expected direction; the other two analyses produced significant results. In sum, the results thus support the hypothesis formulated in the introduction: there does seem to be a connection between the number of established variants of a given expression (of a given semantic type) and speakers' readiness to extend this expression to novel variants of the same semantic type.

Table 3. Dominant mapping: distribution of established and novel types across fields

Rank	Semantic field	Total	Est.	Novel
1	UNHAPPINESS: X feels bad	35	19	16
2	TRUST/CONVICTION: X has faith in something	18	10	8
3	STRONG IMPRESSION: X is moved by something	17	10	7
4	AFFECTION: X likes something	15	8	7
5	ANIMOSITY: X dislikes something	15	6	9
6	ANXIETY: X is worried	12	7	5
7	DISTRUST/UNCERTAINTY: X has no faith in sth.	12	7	5
8	EMPATHY: X shows concern for something	11	5	6
9	DESIRE: X wants something	10	5	5
10	HAPPINESS: X feels good	9	7	2
11	other	8	2	6
12	RESPONSIBILITY: X feels responsible for sth.	5	0	5
13	DISREGARD: X shows no regard for sth.	3	0	3
Total		170	86	84

Methodologically, the study has demonstrated how an interesting but speculative idea from the literature can be turned into a testable hypothesis and evaluated statistically. In addition, it has employed a procedure that seeks to counterbalance the (inevitable) idiosyncrasy of purely introspective semantic judgments, a notorious problem for semantic classifications within *any* framework. Even though a number of questions remain,³ I believe that an approach which combines careful and exhaustive semantic

analysis of naturally occurring data with validations across different coders (or else some sort of experimental evaluation of the obtained results) takes Cognitive Semantics into the right direction.

In terms of linguistic model-building, the results of the study support the usage-based assumption that speakers store linguistic units of varying sizes (including fully predictable collocations) and re-use these representations for later linguistic categorisations. In addition, they also support the assumption that speakers sort and cluster their input using both formal *and* semantic criteria, thus arguing against purely form based accounts of type frequency and the way speakers extract productive schemas from concrete expressions.

Put differently, the main implication of this finding for studies of productivity in general is that for the vast majority of items that can fulfil a particular linguistic function (here: intensification), it is actually misleading to speak of the productivity of this element *in aggregate* (here: to quantify the productivity of intensifying *tief* at large, as compared to other such intensifying adjectives): as long as the interplay of analogy and schematisation has not yet produced a completely generalised formative (i.e. something like *very* in the domain of adjective intensification), calculating overall scores will simply obscure the fact that productive use is in fact restricted to more or less closely circumscribed semantic 'islands' within the more extensive total usage spectrum of the item.

Notes

1. I thank Bettina Rust for coding part of the data, the audiences at ICLC 10, Kraków, and DGKL/GCLA 3, Leipzig, for discussion and two anonymous reviewers for helpful comments.
2. While it is problematic to categorically rule out *any* such combination a priori, there are nevertheless clear differences in degree of conventionalisation. Hence, expressions marked '?' do not have a single instance in the 100m word British National Corpus, and expressions marked '??' do not even yield a single hit on the web with Google (query: *went X*, last checked: 30-9-08).
3. *Duden: Deutsches Universalwörterbuch*. Mannheim: Bibliographisches Institut, 2006. Only two observations (0.04%) could not be assigned to any of the sense groups in (4) due to missing or unrecoverable context.
4. A reviewer remarks that an expression which occurs only once in the corpus is not necessarily unconventional and hence creatively assembled or 'novel'.

Conversely, an expression which occurs more than once in the corpus is not necessarily firmly 'established' in the speech community and hence cognitively entrenched. I agree that these are certainly simplifications, but the operationalisation of creative formations as hapax legomena in a sufficiently large corpus is neither unusual nor implausible (cf. e.g. Baayen & Lieber [1991] and Bybee and Eddington [2006] for discussion and application of the same strategy). Also, "more than one token in a large corpus" may not be an ideal operationalisation of "sufficiently frequently experienced to be familiar", but there is no evident alternative cut-off frequency that would be less questionable in this respect. Ultimately, the precise relationship between textual frequency counts in corpora on the one hand and degrees of cognitive entrenchment in an idealised speaker-hearer (whose linguistic experience the corpus is assumed to represent) remains unclear and probabilistic, not least because speakers' linguistic inventories are constantly changing, and no two linguistic inventories of members of one and the same speech community are in fact fully identical.

5. Instead of distinguishing spatial source meanings from non-spatial figurative meanings, the second coder mistakenly drew the boundary between the Duden's *prototypical* spatial meanings (e.g. 'reaching far (vertically) downward': *ein tiefes Loch*, 'a deep hole') on the one hand and *all other* meanings/subsenses on the other, which also included shifted though still spatial variants of literal meanings 1-3 (e.g. 'low, not far from the lower boundary of something': *tiefe Wolken*, 'deep clouds').
6. For instance, other coders may arrive at a different category scheme to begin with if they were not already confronted with a set of pre-established classes by the first coder. Also, it is not clear *how many* categories there should be in the first place (which is problematic since agreement will of course be higher if there are just two or three rather general classes than if there are, say, 20 categories involving very subtle distinctions).

References

- Abbot-Smith, Kirsten and Michael Tomasello
 2006 Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review* 23: 275–290.
- Baayen, Harald and Rochelle Lieber
 1991 Productivity and English derivation: A corpus-based study. *Linguistics* 29: 801–843.

Brennan, Robert L. and Dale J. Prediger

- 1981 Coefficient Kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement* 41: 687–699.

Bybee, Joan L.

- 1985 *Morphology: A study of the relation between meaning and form*. Amsterdam / Philadelphia: John Benjamins.
- 1995 Regular morphology and the lexicon. *Language and Cognitive Processes* 10: 425–455.
- 2006 From usage to grammar: The mind's response to repetition. *Language* 82(4): 711–733.

Bybee, Joan L. and David Eddington

- 2006 A usage-based approach to Spanish verbs of 'becoming'. *Language* 82(2): 323–355.

Erman, Britt and Beatrice Warren

- 2000 The Idiom Principle and the Open Choice Principle. *Text* 1: 29–62.

Goldberg, Adele

- 2006 *Constructions at Work. The nature of generalization in language*. Oxford: Oxford University Press.

Israel, Michael

- 1996 The *Way* constructions grow. In *Conceptual Structure, Discourse and Language*, Adele Goldberg (ed.), 217–230. Stanford: CSLI.

Langacker, Ronald W.

- 2000 A dynamic usage based model. In *Usage-based Models of Language*, Michael Barlow and Suzanne Kemmer (eds.), 1–63. Stanford: CSLI Publications.

Zeschel, Arne

- 2007 Delexicalisation patterns: A corpus-based approach to incipient productivity in 'fixed expressions'. PhD dissertation, Universität Bremen.