

GRAMMATICAL CATEGORIES IN THEORY AND PRACTICE: HOW TO CAPTURE EVERYTHING FROM UNIVERSALS TO THE DIFFERENCE BETWEEN ZERO DETERMINERS IN DANISH AND ENGLISH

by
Kasper Boye & Peter Harder

This article discusses Carl Bache's descriptive approach to the study of grammatical categories as a continuation and improvement of the structural-functional tradition in European linguistics. Focusing on its achievement in integrating within a single format a level of maximal abstraction with a level of combinatorial multiplicity, we present what we see as the logical next step. At the most abstract end, we argue that the stipulation of metalinguistic conceptual distinctions needs grounding in semantic (in our terms functional-cognitive) substance on a cross-linguistic, typological basis. At the most concrete end, the multiplicity of semantic contrasts merges naturally into collocational and constructional variation.

1. Introduction

Over his career, Carl Bache has subjected the notion of 'grammatical category' to increasingly penetrating scrutiny. In so doing, he has contributed significantly to the continuing strength of a specifically European tradition in linguistics, which goes back all the way to the Alexandrian grammarians in antiquity and continues via European structuralism to various functionalist schools, including Hallidayan Systemic-Functional linguistics.

Modern American linguistics, in contrast, has focused on principles of syntagmatic orientation, leaving morphological categories in a more marginal position. This is true of both the formal and

the functional schools, although for different reasons. Generative grammar, as the mainstream formal approach, takes its point of departure in the foundational task of describing how an infinite set of possible sentences can be derived from a finite set of principles. This perspective means that focus will be on formal principles and the syntactic configurations that manifest them, rather than on the types of morphosyntactic markings that underlie the traditional concept of 'grammatical category'.

In direct contrast to this narrow focus on syntactic structure, American functionalists have focused on the roots of grammar in discourse and semantic functions. From that point of view, morphosyntactic markings are also marginal, in that they are more resistant than lexical and discourse phenomena to being explained by direct reference to function. Emphasis has been on showing how grammatical patterns are integral parts of an overall description of language in terms of function in linguistic interaction. Focus has been on individual expressions, not on categories, as illustrated in mainstream grammaticalization theory, where focus is on the diachronic trajectory of individual 'grams', going from a lexical etymon to a grammatical item, cf. the Romance adverbial ending *-mente* from the Latin ablative form of the lexical item *mens* ('mind'), or the Romance future forms from fusion with the verb *habeo*.

Synchronically, the most significant contribution of this focus on placing individual meanings within a larger framework has been the 'semantic mapping' approach. The essence of this approach consists in establishing an all-encompassing conceptual universe within which each meaning can be placed in a motivated manner, such that spatial closeness on the map reflects conceptual relatedness between individual meanings. Again, however, the notion of 'grammatical category' does not emerge automatically from such a map of related meanings, but has to be added as an independent afterthought.

One feature that has contributed to backgrounding categories in the functional discussion is their stubborn resistance to being

accounted for in a fully motivated manner – which is important in the polarized discussion between functionalists and formalists in the American context. Categories diverge strikingly even between related languages and their particular manner of organization within a specific language makes it obvious why they have historically constituted the core area for structural rather than (purely) functional analysis.

In the European structural tradition, however, there has been no such polarization between functional and structural interests. A structural approach automatically implies that linguistic features cannot be directly derived from non-linguistic features. However, once structural features have been identified, it is taken for granted that it is the task of linguistic description to capture the functions of grammatical elements in relation to other linguistic elements; such functional relations are the stuff of which structural description is made. Functional relations with non-linguistic elements are also implicitly recognized, in the sense that the 'substance domains' within which grammatical meanings belong are implicitly understood as *tertium comparationis* for comparing the different ways in which grammatical systems are organized. Thus the famous diagrams showing different borderlines between colour categories in languages imply that colours exist regardless of linguistic differences and are neutral between them, and one could think of colour terms as having the shared function of conveying colour categories that are as such not linguistic categories but categories of conceptual substance (strictly speaking *purport*, Da. *mening* in Hjelmslev's terms (1943/1953), cf. also Croft & Cruse (2002)).

Carl Bache's approach to the description of grammatical categories belongs within the European structural-functional tradition in a period that went beyond classic structuralism. While in the heyday of structuralism little explicit attention was paid to the anchoring of grammatical distinctions outside the language-specific system, Bache's work has unfolded in a period when their functional and conceptual underpinnings had become a central concern. In this

article, we pay tribute to his achievement in clarifying the relation between grammatical structures and functions, and try to take the process one step further.

2. *The architecture of grammatical categories as conceived by Bache*

Bache's descriptive approach is predicated on a distinction that has recognizable similarities with Hjelmslevian structuralism: the distinction between an abstract, theoretical level that is stipulated in advance by the linguist, and the outcome of applying categories at the theoretical level to the description of the empirical set of forms found in the object language under description.

The purely theoretical level serves a metalinguistic purpose: in order to have a criterion for subsuming particular linguistic forms under a particular category label, one has to take one's point of departure in an explicit and inspectable criterion, to which one remains accountable. This principle has the great merit of avoiding a situation in which one takes for granted the existence of a category because of its status as part of the traditional lore, and subsequently gets into trouble because the tradition turns out to have faults in its foundations.

Bache's theoretical categories have other merits. Like Max Weber's 'ideal types' they embody carefully honed abstractions designed to bring out the most typical and pervasive feature of the empirical phenomena they are designed to subsume. In the case of the category of aspect that formed the subject of his doctoral dissertation (Bache 1985), the abstract characterization involves a distinction between 'external focus' and 'internal focus'. Taking his point of departure in cases where both forms are possible, and where the meanings are only minimally distinct, Bache arrives at a description at the theoretical, definitional level that can apply to the whole range of cases, one of his staple English examples being *we celebrated Stephanie's*

birthday at my uncle's the other day vs. *we were celebrating Stephanie's birthday at my uncle's the other day*, where the only difference is the definitional one of focus on the ongoing process (*were celebrating*) as opposed to the whole event (*celebrated*).

The procedure is a model example of the criteria that Hjelmslev (1943) espoused for evaluating the abstract theory: it must be arbitrarily stipulated in advance in order to be free of implicit assumptions – but it must be chosen to as to be *appropriate* in relation to the object of description.

Bache, however, wants the theory of grammatical categories to go beyond the description of specific languages. A linguistic theory must be capable of being applied to all languages, rather than being tailor-made to a specific language. It is not enough to perform abstractions over examples from one language. This significantly raises the stakes for the process of making definitions appropriate, and Bache is aware of the difficulties this creates (cf. Bache 1995: 34). Unlike Hjelmslev, however, Bache is fully explicit about the background for the actual stipulation of a definitional level for universal grammar: he speaks of a "source language" as one that has been used as substrate for the process of devising the theoretical, definitional level of a given category, reserving the term 'object language' for those languages to whose description the theory is applied.

The definitional level of grammatical categories may be regarded as an openly declared and explicitly motivated manifestation of the standard descriptive aim of European structural linguists: to find the abstract characterization that would capture the whole body of instantiations most adequately, while ignoring variant and inessential features. The so-called "functional" level is a considerably more innovative contribution to the understanding of grammatical categories. This is the level at which one may describe the semantic differences between minimal pairs without worrying about making sure that the description would be equally applicable to other minimal pairs. In the case of the contrast between progressive and

non-progressive forms, some cases involve a difference between complete and incomplete (*he swam/he was swimming across the river*) and some a difference between 'temporary' and 'permanent' (*he lived /was living in a coal cellar*).

This distinction allows Bache to include a theoretically adequate account of features that have tended to live separate lives in the literature about grammatical phenomena. While theorists have mainly argued about the merits of competing abstract generalizations (cf. e.g. Schousboe 2000, Durst-Andersen 2000 on rival theories of the progressive), traditional grammars and school textbooks have generally concentrated on finding illustrative examples of the panoply of usage variants, traditionally called 'uses' of the progressive, etc., without providing a theoretically satisfactory account of their status.

One of the merits of Bache's account is that he explicitly includes these variant cases as parts of the grammatical system. The familiar term 'uses' is misleading, both because it suggests that this level of description represents "raw empirical language use", rather than being well-established grammatical types, and also because an adequate description of this level of variation involves structural features of grammatical categories that have been overlooked. The key feature that Bache has pointed to is that the actual function of grammatical categories as a way to make distinctions between formally minimal pairs is 'non-monadic', i.e. when you choose a particular grammatical form, the choice has implications for several different semantic domains, not just for the definitional meaning of the single category you may be focussing on, cf. Bache (2008:107). A form like *spoke* is in structural contrast not only with present-tense *speaks*, although one can focus on this difference when investigating tense, but also in contrast to the progressive *was speaking*. Moreover, the difference between

- (1) *She speaks like a real professional.*
- (2) *She spoke like a real professional.*

is not captured by the temporal difference belonging at the definitional level of the tense category. The present-tense version is only naturally understood as involving a 'habitual' reading, while the past-tense version is ambiguous between a habitual and an event reading (cf. Bybee, Perkins & Pagliuca 1994: 55 for an example of how the lack of a theoretical distinction between the two levels may cause a misinterpretation of the structural meaning of the present tense as involving the aspectual meaning 'imperfective').

This type of phenomenon is not a consequence of accidental usage events, but is due to features of the way grammatical structures work: although each category may have a privileged invariant domain, it interacts with other structural choices to produce predictable effects that considerably enrich its functional role in the language. Including the "functional" level of description therefore also brings grammatical description much closer to the type of choices that are recognizable from a practical or teaching-oriented point of view.

Below, we are going to make some proposals of our own about the cline from 'most general and cross-linguistic' to 'most concrete and language-specific'. In so doing, we will argue that Bache's descriptive framework needs to be extended at both ends.

3. *Functional considerations and crosslinguistic grammatical categories*

3.1. What is a crosslinguistic grammatical category?

We begin at the most general end – with the nature of universal or cross-linguistic grammatical categories. From the perspective of Danish Functional Linguistics (e.g. Engberg-Pedersen et al. 2005; cf. Boye 2010a, 2010b, 2012 on crosslinguistic descriptive categories in general) crosslinguistic grammatical categories have the following properties.

1. *Crosslinguistic grammatical categories are descriptive rather than ontological categories.* Crosslinguistic grammatical categories are linguists' generalizations based on comparison of individual languages. While these generalizations may be assumed to reveal something about language-independent categories – notably, functional and conceptual categories – they cannot be identified with such categories, or assumed to reflect such categories in any straightforward manner.
2. *Crosslinguistic grammatical categories are semantically rather than morphosyntactically based.* In order to be a member of a grammatical category, a given item must of course be grammatical. Members of grammatical categories are affixes, auxiliaries, and particles, rather than verbs, nouns, and adjectives. Grammatical status is not sufficient to define particular grammatical categories, however. Obviously, for instance, tense cannot be defined solely with reference to the fact that all its members are grammatical. What grammatical categories share across languages is their semantics, not their morphosyntactic characteristics. Tense subsumes items which indicate (deictic) location in time, evidentiality covers items which indicate source of information, and so on and so forth.
3. *Crosslinguistic grammatical categories are substance-based rather than structural.* Danish Functional Linguistics subscribes to the structuralist view of linguistic structure as language-specific, and substance (Hjelmslev's *purport/mening*, cf. above) – i.e. functional or conceptual communication potential – as potentially universal. Thus, any descriptive generalization which is based on more than one language must inevitably be substance-based.

It follows that crosslinguistic grammatical categories are basically descriptive generalizations grounded in semantic substance.

Two types of crosslinguistic grammatical categories can be discerned. Following Whorf (1956: 113), we will refer to the two types

as respectively *specific* and *generic* categories. Specific crosslinguistic grammatical categories are descriptive generalizations over individual language-specific values like past, present and future (as in mainstream grammaticalization theory, cf. above). Generic crosslinguistic grammatical categories like tense are generalizations over groups of related individual values (tense, evidentiality, etc). Below, we only discuss generic categories. However, we believe that much of what we say applies also to specific categories.

3.2. How to identify crosslinguistic grammatical categories, step 1: conceptual generalization

As mentioned, crosslinguistic grammatical categories are basically descriptive generalizations in terms of semantic substance. Accordingly, in the process of identifying a crosslinguistic grammatical category the logical (though not necessarily the most practical) starting point is the establishment of a descriptive generalization in terms of substance, i.e. a substance-based *tertium comparationis* for the description of individual values and groups of individual values in distinct languages.

As discussed in Boye (2012: 8), a substance-based *tertium comparationis* may be arrived at in one of at least two ways. One way goes through elicitation of relevant substance values – relevant communicative functions. For instance, Majid and Bowerman (2007) elicit descriptions of cutting and breaking events by means of videoclips of a range of such events. This is well-suited for cases where the substance is representational. The other way is well-suited when the substance is more abstract, as in the case of the meaning of grammatical items. This consists in setting up a range of what Haspelmath (2010) calls "comparative concepts": conceptual constructs which are designed by the linguist to describe substance values, and thus to serve as a common ground for the description and comparison

of language-specific values and groups of values. For tense values, comparative concepts like 'past', 'present' and 'future' might be set up, and on a more general level a comparative concept like 'time' or 'deictic time location' might be introduced to encompass all specific concepts relevant to the description of the members of the generic crosslinguistic category of tense.

We will adopt Bache's term *category concept* for general-level concepts like 'time' or 'deictic time location'. Note, however, that there is a slight, but important difference between the way we conceive of category concepts as described above and the way Bache does. As described in section 1, for Bache, category concepts are not substance descriptions, but metalinguistic stipulations; although they are intended to serve for crosslinguistic comparison, Bache's category concepts are grounded in language-specific structural generalizations in that they derive from analysis of one or more "source-languages", in practice often a single language. They therefore, as explicitly stated by Bache (1995: 67), do not represent an ideal solution (we would say that they involve a risk of what Otto Jespersen called "squinting grammar") – but the descriptive practice of stipulation means that the linguist is quite open about the limitations of his choice. Our approach differs from Bache's also in another way: Bache (1997: 138) associates individual conceptual values like 'past', 'present' and 'future' with abstract morphosyntactic forms. There is an obvious practical reason to do so: it is convenient to be able to refer to comparable items in distinct languages as items of the same type. To spell it out, it is more convenient to say of item x in language X and item y in language Y that they are both, say, past tense *forms*, than to say that they both have a meaning that can be described in terms of the substance value (i.e. comparative concept) 'past'. However, abstract morphosyntactic forms are pure constructs rather than features of the object of description: they do not exist before the linguist decides to set them up. We think an approach based on semantic substance can provide a better grounding for crosslinguistic grammatical categories,

In identifying crosslinguistic grammatical categories, linguists traditionally tend to rest content with what are generally known as 'conceptual generalizations'. One symptom of this is the proliferation of terms like "conceptual categories" for describing grammar. As ways of 'carving up language at the joints', however, conceptual generalizations are theoretically problematic in that they necessarily incorporate a considerable element of subjective choice by the analyst. There is no guarantee that they actually describe coherent areas of cross-linguistic, semantic substance. To take an example, the term *modality* is used in semantics as a cover term for two sets of phenomena: 1) phenomena which – like Germanic modal verb semantics – can be described in terms of the concepts 'necessity' and 'possibility'; 2) phenomena which can be described in terms of concepts like 'speaker attitude'. The two sets are unquestionably related: many languages have items which express speaker judgments in terms of (epistemic) necessity and possibility. English is a case in point (data from the British National Corpus).

- (3) *For Jesus to have been accepted as the Messiah, he must necessarily have been prepared to wield the liberator's sword.* (BNC: EDY 886)
- (4) *It is possible that he may still be a candidate for the post of Speaker of the House of Commons.* (BNC: AK9 271)

Based on this, many linguists have tried to lump the two sets of phenomena into one. Most often, the concept of speaker attitude has been stretched to include necessity and possibility and the phenomena covered by these two concepts. The resulting conceptual generalization has found its way into even high-quality text books (e.g. Saeed 2008: 138-146). However, there are several arguments that the phenomena covered by the generalization do not make up a linguistically coherent area. For instance, 1) non-epistemic

necessity or possibility have different scope properties from epistemic necessity and possibility (e.g. Lyons 1977: 842-843; Palmer 1979: 35; Perkins 1983: 7-8; Boye 2001, 2005), and 2) expressions of non-epistemic necessity and possibility often have different distributional characteristics than do expressions of epistemic necessity and possibility (see e.g. Boye 2001 on the distribution of directional adverbs with Danish modal verbs). Wide-sense modality is a phantom based on a linguistically unanchored conceptual generalization.

In his discussion of comparative concepts, Haspelmath emphasizes the subjective character of conceptual generalizations. Comparative concepts, he writes, "are concepts created by comparative linguists for the purpose of formulating readily testable crosslinguistic generalizations" (Haspelmath 2010: 673). They are descriptive "constructs" (Haspelmath 2010: 666) that "cannot be right or wrong", but only "more or less productive, in that they allow the formulation of more or less interesting subdivisions and generalizations" (Haspelmath 2010: 678). Haspelmath leaves the impression that there is no way to constrain or validate comparative concepts. Since he takes comparative concepts to be the *sine qua non* of crosslinguistic comparison, he may be interpreted to suggest that the same applies here: crosslinguistic comparison is fundamentally untamed.

In a reply to Haspelmath's discussion note, Newmeyer (2010) criticizes this aspect of comparative concepts.

If the comparative concepts that enter into a typological generalization are 'productive', then we want to know WHY they are productive. The answer might possibly be due to an innate universal grammar (UG), or possibly be a consequence of a complex interplay of cognitive and functional factors. But there HAS TO BE an answer – productive hypotheses and lines of research derive their productivity from the fact that they shed light on the essential nature of the phenomena

being investigated, while nonproductive hypotheses and lines of research do not do so. (Newmeyer 2010: 688)

We endorse Newmeyer's stance: we want to be able to assess the adequacy of the common ground on which languages are described and compared, and since we do not endorse innate universal grammar, we go with the alternative: we want to ground our choice of crosslinguistic categories in what we know about functional-cognitive substance. In this way, a functional linguistics will be able to validate conceptual generalizations, rather than simply stipulating them.

3.3. How to identify crosslinguistic grammatical categories, step 2: structural validation

When we talk about tense, aspect, modality, and other categories as crosslinguistic grammatical categories, it follows from the rationale presented above that we should not merely talk about conceptual generalizations. Rather, the aim is to talk about generalizations that are **significant for the description of structural phenomena found in geographically and genetically distinct languages**. Descriptive significance, then, serves to validate and upgrade crosslinguistic conceptual generalizations as crosslinguistic categories (Boye 2012). Specifically, crosslinguistic *grammatical* categories can be defined as: **conceptual generalizations that are significant for the description of grammatical phenomena found in geographically and genetically distinct languages** (this means that identification of crosslinguistic grammatical categories presupposes the ability to tell grammatical from lexical phenomena; see Boye & Harder (2012) for a functional-theoretical basis of the distinction).

Boye (2012) proposes three ways in which conceptual generalizations can be significant for the description of structural phenomena in distinct languages. Two of these three ways will be outlined below

and exemplified by a tentative discussion of the category of tense (cf. Boye 2012 for a detailed validation of the three categories: epistemic modality, evidentiality, and epistemicity).

The first way has to do with distribution. A conceptual generalization qualifies as a crosslinguistic category if it is significant for the description of distributionally defined sets of linguistic items – henceforth, *morphosyntactic systems* – in distinct languages. A conceptual generalization is significant if each of the members of the morphosyntactic systems under analysis have (or can be assumed to have had; see Boye 2012: 51 for discussion) a meaning which can be described in terms of the generalization at hand. Tense, defined in terms of the category concept 'deictic time location', qualifies as crosslinguistic grammatical category in so far as a number of geographically and genetically distinct languages have morphosyntactic systems whose members all have meanings that can be described in terms of the category concept 'deictic time location'. For instance, Lithuanian has a morphosyntactic tense system of suffixes which is defined distributionally by the fact that all members of the system occur immediately after the verb stem. All members indicate 'deictic time location' (some additionally indicate person and number).

Lithuanian (Timberlake 2007; Kroeger 2005)

- (5) a. *Dirb-au.*
work-1SG-PAST
'I worked/was working'.
b. *Dirb-u.*
work-1SG.PRES
'I work/am working'.
c. *Dirb-s-iu.*
work-FUT-1SG
'I will work/be working'.

The existence of morphosyntactic tense systems which are conceptually homogenous and similar to the one described is a non-trivial empirical finding, which may serve to validate a conceptual generalization by making it significant for linguistic description.

The second way has to do with semantic maps (see Haspelmath 2003, and Boye 2012: 127-129 for introductions). Semantic maps, and semantic mapping, rest on the assumption that regular synchronic multifunctionality and regular diachronic change in meaning are non-accidental and constitute evidence of relations pertaining to semantic substance. Roughly, semantic maps consist of comparative concepts, and connecting lines between nodes. Connecting lines are established on empirical grounds. A connecting line between two comparative concepts can be established 1) if one or more language-specific items are found which have meanings the description of which includes both concepts, or 2) if one or more language-specific items are found which move diachronically between a meaning describable in terms of one of the concepts to a meaning describable in terms of the other one. Multifunctionality is taken to include ambiguity and polysemy as well as vagueness, and both semantic (coded, conventional) and pragmatic (context-dependent) meanings are taken into consideration. For obvious reasons, however, accidental homonymy, as in Danish *lam* ('lamb') vs. *lam* ('paralyzed') is excluded from the analysis.

A semantic map of tense items might look like the map in Figure 1. It should be noted that the map is low-granularity in that it covers only a rough three-way distinction. Moreover, it should be noted that the map depicts only temporal comparative concepts. Each of the concepts depicted may be linked to various other concepts; however, such other concepts are excluded from the map, as they are irrelevant for the present purpose.

Past ——— 1 ——— Present ——— 2 ——— Future

Figure 1: *A tentative semantic map of tense items*

The distinction (under the umbrella 'deictic time location') between the three comparative concepts 'Past', 'Present', and 'Future' would be based empirically on languages like Lithuanian in which a corresponding three-way distinction is made. Connecting line 1 would be based empirically on items with meanings the description of which includes both of the comparative concepts 'Past' and 'Present'. Examples of such items are found in Mapudungun and Lakhota (Timberlake 2007: 305). Finally, connecting line 2 would be based empirically on items with meanings the description of which includes both of the comparative concepts 'Present' and 'Future'. Danish *-er* in e.g. *svømmer* is an example: It can be used both to express a meaning describable in terms of 'Present' (*jeg svømmer nu* 'I am swimming now') and one describable in terms of 'Future' (*jeg svømmer i morgen* 'I will be swimming tomorrow').

A conceptual generalization qualifies as a crosslinguistic category if it covers a coherent area of a semantic map. If the map proposed above is correct, tense qualifies in so far as the category concept 'deictic time location' covers a coherent area of a map of tense items: Each of the comparative concepts covered by the category concept is linked to at least one other concept covered by the category concept. Like the existence of morphosyntactic systems that are conceptually coherent (see above), semantic-map coherence is a non-trivial finding. It is an empirical finding in so far as semantic maps are empirically-based descriptions of crosslinguistic variation pertaining to 1) synchronic multifunctionality, and 2) diachronic meaning change. Conceptual coherence may thus serve to validate a conceptual generalization by making it significant for linguistic description.

Summary: Many linguists would subscribe to crosslinguistic grammatical categories of tense and epistemic modality. Most of them would do so for pretheoretical reasons, for instance based on intuitions formed in the course of acquiring a basic linguistic competence. In the discussion above we have outlined a notion of

crosslinguistic grammatical category. Armed with this notion, and an adequate crosslinguistic database, we may confront claims about categories made in the literature. In some cases, surely, intuitions will be confirmed. In other cases, however, they will have to be rejected.

As discussed above, crosslinguistic grammatical categories are descriptions of semantic substance which are significant for the description and comparison of structural phenomena in distinct languages. Crosslinguistic grammatical categories thus represent a link between potentially universal substance and language-specific structure. In one direction, they provide an empirical and theoretical basis for descriptions and comparisons of individual language. In the opposite direction, they reveal something about substance, in so far as substance is structurally significant.

3.4. Comparison between stipulation and cross-linguistic validation

There is no necessary empirical conflict between stipulated and cross-linguistically validated categories. In a sense, the validation procedure merely constitutes what we see as the key prerequisite for optimal stipulation; we do not wish to claim that descriptive categories emerge directly out of linguistic reality (if you merely look hard enough).

From this it follows that we are not arguing that all traditional generic categories have to be rejected. In fact, one would expect there to be something robust behind most of those categories which have survived the test of time. However, as repeatedly demonstrated by Bache, e.g. in his criticism of Systemic-Functional Linguistics (Bache 2008: 18), and of Harder (1996) (Bache 2008: 87-88), the precise rationale for the use of traditional generic categories is often unclear or absent. Stipulating, in Bache's manner, a precisely defined semantic-conceptual criterion thus remains a step forward compared to implicit reliance on the tradition.

The position we have argued for is essentially a claim about the work that remains to be done when linguists have offered their proposals for precise criteria for the cross-linguistic identification of generic grammatical categories. Our claim may, in that sense, be regarded as a working hypothesis: we have conjectured that there exist properties of the functional-cognitive semantic **substance** (or purport) that makes it possible to evaluate different **metalinguistic** ways of 'cutting the pie' by reference to the ways in which **languages** cut the pie. In this, we continue the structuralist tradition of taking seriously the way specific language organize semantic substance – but we anchor the structural investigation in a universe of functional-cognitive meaning potentials, with non-random properties, which constitute the input to linguistic structuring. Just as the structure of a textbook can be expected to be influenced by the structure of the content (one would expect a reader in anatomy to have headlines referring to body parts, for example), we expect grammatical structures to reflect the functional-cognitive content of the substance domains that they encode. In both cases, however, it is impossible to derive the structure unambiguously from the content: arbitrariness and motivation are not opposites, but invariably go hand in hand.

4. *Abstract generalization vs. concrete usage in linguistic theory*

4.1. The same abstract principle – but where to draw the line?

But the problem of abstract vs. more concrete description has not yet been fully dealt with at this point. The discussion above is central in that it addresses the issue of how to provide universal grammar with a dimension of empirical accountability, but the problem recurs in other forms. In general, it is a matter of empirical investigation

what categories are best suited to capture the generalizations that are at work in actual language use. Assuming that the goal is to carve the object of description at the joints, universally as well as in specific languages, we need to pursue the same inquiry also at the other end of the scale spanning from universal and abstract to language-specific and concrete

This issue may be illustrated with the case of zero determination in English, which is rather different from the classic morphosyntactic case of tense. The general rule is that zero determination occurs with plural or non-count nouns, when only the category meaning, i.e. no quantification, is intended (Quirk et al: 1985:275). This is also known as 'generic' or 'non-specific' reference. Prototype cases are

(6) *Boys will be boys.*

(7) *Oil floats on water.*

The phenomenon is well-known cross-linguistically (on Danish, cf. Hansen & Heltoft 2011: 547). Thus, the zero forms in (6) and (7) have direct counterparts in Danish:

(8) *Drenge er nu engang drenge.*

(9) *Olie flyder oven på vand.*

The distribution, moreover, is iconically motivated: the category meaning is encoded by the noun alone, and so there is no extra semantic work for a determiner to do in case of non-specific/generic meaning. It would appear that here we have another case of a universal and very abstract distinction that can be applied so as to capture what goes on in concrete cases in a specific language like English.

However, this underestimates the complexity of usage level facts. Consider (10)-(12):

(10) *Unemployment is rising.*

(11) *Communism is dead.*

(12) *Tigers are becoming extinct.*

The interesting fact is that (10)-(12) cannot be translated into the corresponding zero forms in Danish:

(13) **Arbejdsløshed er på vej op* (but *arbejdsløshed er af det onde* 'unemployment is a bad thing').

(14) **Kommunisme er død* (but *kommunisme er noget urealistisk vås* 'communism is a load of unrealistic nonsense').

(15) **Tigre er ved at uddø* (but *tigre er kattedyr* 'tigers are felines').

Both forms, however, come under the heading of generic (non-specific) reference: it is not a particular segment of unemployment that is rising, or a particular quantity of tigers that are becoming extinct. We see here that although the zero form can be characterized by the same abstract and universal concept, this description does not suffice to capture the semantic territory of the forms in the two languages.

The examples used above can be characterized partly by saying that Danish is more restrictive; a further step might be to suggest that Danish tends only to use the zero form when it is the pure abstract category that is in question, while zero forms are found in English also when used about the category as manifested in the world (e.g., it is not the abstract category of unemployment that is rising, but its

total occurrence in the real world). In Danish, the cases that involve the category as instantiated in the real world have to be rendered as definite forms, in other words as involving definite reference:

(16) *Kommunismen er død.*

(17) *Arbejdsløsheden er på vej op.*

(18) *Tigeren/tigre er ved at uddø.*

Some numbers drawn from corpus studies may be used to support this intuition-based account: Of 50 occurrences of *inflation* (as a result of random selection as offered by the BNC system), 31 were zero-article cases, of which 21 could not naturally be rendered by zero-article forms in Danish. For *unemployment*, the figures were 35 zero-article occasions in English, of which 15 could not naturally be translated into zero-article forms in Danish. Out of 4,484 total occurrences of *inflation* in BNC, only 165 involved the combination *the inflation*.¹

In contrast to this, in Korpus.dk, there were 1,336 occurrences of *inflation* in all forms, including the inflected definite form *inflationen* (= 'the inflation'), which alone accounted for almost half (619) of the occurrences.

For *unemployment*, the English figures were a total of 6,442, with 394 occurrences of *the unemployment*, while the Danish figures were 2,828 total occurrences of *arbejdsløshed* (all forms), out of which almost half (1,334) were the definite version *arbejdsløsheden*.

We are thus left with a problem of descriptive strategy: how do we capture differences of this kind? The problem is situated at the same descriptive interface as the distinction between Bache's "definitional" and "functional" levels: on the one hand we have a very general distinction designed to carve universal linguistic nature at

the joints; on the other we have a set of messy facts that cannot be exhaustively captured by the abstract distribution (in spite of the fact that it does apply to both languages, up to a point). A 'functional' level of description, however, is not readily imaginable as a solution in the cases described above: no other categories offer themselves as sources of explanation in cases (10)-(18) (i.e. by interacting with the zero article, analogously to the combination of verbal categories discussed above).

For purposes of illustration, we may begin with a type of case which does involve a functional level that includes categorical interplay. Thus *inflation* and *recession* can both occur as non-count nouns with the zero article. However, *recession* differs from *inflation* in terms of its ability to occur as a count noun. This can be illustrated by comparing figures for co-occurrence with the indefinite article: BNC found 234 occurrences of *a recession*, and only 28 of *an inflation*. Of the latter, 17 were *an inflation rate*, one where *inflation* = 'inflation rate' (*an inflation of 5%*), + 5 similar adnominal constructions (*an inflation target/volume/proof/allowance/problem*, 3 metaphorical uses (*an inflation of pride/of the share price/in employers' demands*). Only two typical countable cases were found: *To see how destructive such an inflation can be / it swallows up an inflation*. The results for the corresponding constructions in Danish in Korpus.dk are parallel: 1,336 cases of *inflation* (in all forms), only 25 for the combination 'en inflation', of which 24 were translatable as 'inflation rate' and only one was understandable as 'an instance of inflation'. *Recession* had 52 occurrences (in all forms), including 11 cases of *en recession* (= 'instantiation of recession').

The categorical interplay is at work in the status of the definite form *the recession*. Here, two choices are involved: the choice of definiteness and the choice of countable vs. non-count construal of the category. When a recession is construed as a unit, the choice between definiteness and indefiniteness is different from the non-count construal: the indefinite version designates 'an unidentifiable

unit among others'. Hence, a substance meaning that includes not only the abstract category, but its 'total occurrence in the world' (as in *unemployment*, cf. above) is not compatible with the count construal signalled by the indefinite article. It follows that only the definite choice will match up with a substance meaning that includes the 'total occurrence'.

In the case of *inflation* the countable construal is a very marginal option. This means that at the 'function level', the two contrasts (*inflation/the inflation* and *recession/the recession*) are different, because the two grammatical choices 'definite/indefinite' and 'countable/non-count' interact in different ways (compare the analogous complexity of the choice of progressive vs. simple aspect in the cases discussed above, such as *he swam/was swimming across the river* vs. *he lived/was living in a coal cellar*).

However, this does not solve the key problem of capturing the difference in where English and Danish draw their boundaries between zero and definite article for non-count noun phrases. Yet the semantic map approach will offer an option for capturing the classic structural notion of having the net of structure throwing its shadow upon substance (as charted in the semantic map). When generic meaning is involved, one can establish a continuum from wholly abstract meaning, where only conceptual content is involved (as in dictionary definitions, such as "inflation" = economic process where the amount of goods and services purchased for a given sum of money is going down), towards more and more concretely identifiable referents. In this substance domain, the case where external reference is intended but non-delimited (i.e. encompasses everything that currently instantiates the concept) belongs in the middle, with a delimited, identifiable instantiation further away from purely conceptual reference. On such a map, one can therefore strive to make a description of substance meaning differentiated enough to allow the linguist to put the boundary in marginally different places, depending on the individual language.

4.2. Item-specific differences

But the differentiation can go even further. Construction grammar, as a radical reflection of the general bottom-up orientation of usage based cognitive-functional grammar, embodies the claim that the process whereby particular forms of expression gain currency operates basically in a bottom-up manner. This entails that the individual combination has a life of its own that is not captured by general rules ('partial compositionality' is one of the ways of capturing this phenomenon). Thus for each linguistic item, one needs to find out how strong its 'preference' is for a particular collocation – including the choice between zero and definite article.

It is illustrative in this regard to compare the figures for *unemployment* and *inflation* with the figures for *discrimination* and *racism*. Where above we saw that in Danish, the definite version accounted for close to half the total occurrences, as opposed to a vanishing figure for English, the figures below are much closer in the two languages:

(19) Danish:

Diskrimination 434
Diskriminationen 31

(20) *Racisme* 737

racismen 107

(21) English:

Discrimination 1,990
The discrimination 70

(22) *Racism* 1,086

The racism 45

Based on these raw figures, however, we cannot determine to what extent such differences are due to other factors, including a different distribution of generic vs. specific occurrences, or purely coinci-

dental usage reasons. A speculative possibility might be that this is an area where the abstract principle of the issue has a particularly strong appeal ('the repulsive idea in itself'), rendering the choice of definite reference to the messy realities of the phenomenon 'as manifested' less striking also in Danish. (If that is assumed, the necessary corollary is that the 'pure abstract ideas' of inflation and unemployment must be assumed to have less clout as compared with their manifestations in reality.)

At the level of the collocational preferences of individual words, idiosyncratic factors may play an important role. In the case of zero vs. definite article combined with non-count nouns, an example would be *the Enlightenment*, which has become established as a brand name for the historical period, thus skewing collocational preferences. The abstract principle has not been put out of business, however: It would still be possible to talk of *enlightenment* in the abstract, also about what was happening in the era in question – but it is less natural than if the collocation *the Enlightenment* had not become entrenched as a preferred way to sum up what was characteristic of the period.

An analogous example for the progressive would be the McDonald slogan *I'm loving it*: this example is marginal from the point of view of the abstract contrast (*love* is a stative verb that only awkwardly lends itself to a 'mid-process' interpretation), but its usage status is skewed because of the massive exposure that is the result of the use in advertising. For that reason, both *I'm loving it* and *the Enlightenment* must be accounted for partly at the level of the individual construction.

The logical ending point of this trajectory is the point at which explanation has to give way to variability as a fundamental fact about the way the world works. As pointed out by Croft (2009), when a number of alternatives are available to encode a given situation (his illustration case being "the pear stories"), it is a fallacy to assume that there is always a precise reason for the choice of one alternative

among others. Whether we say *the unemployment* or *unemployment* is partly a matter of chance. As we move into the level of granularity where randomness acquires an increasing role, precise explanation should bow out gracefully.

5. *Final remarks*

Part of what this discussion has tried to show is that differences in terms of preferred theoretical approach can sometimes be profitably reinterpreted as differences in terms of the empirical phenomena that are at the centre of attention. European structuralism focussed on abstract and general contrasts, because the orientation of the framework was based on an ontology that considered abstract constancy as the ideal level for true description of linguistic fact, while usage variation was epiphenomenal and of little inherent interest. At the other end of the spectrum, construction grammar (in the cognitive-functional version) focuses on the unique characteristics of a particular usage combination of linguistic choices, backgrounding the generalizations that are shared between such constructions. But neither approach would deny that both phenomena are part of linguistic reality.

For this reason, Fillmore's 'ordinary working grammarian' cannot afford to choose between such theoretical approaches. The only option left is to consider the respective advantages of each such preferred focal area. This is even more true of the ordinary working grammarian's cousin, the ordinary working language teacher, who needs to refer to grammar from time to time (cf. Lightbown & Spada 2006 on why the tantalizing dream of teaching language without grammar has been given up). It is necessary to maintain an awareness of the different levels at which linguistic conventions operate, so as to be able to choose the level that is relevant in the case of the phenomenon that one wishes to grasp.

At the present time, the concrete end of the spectrum is mostly in vogue among functionally oriented linguists – justifiably so, since in the past couple of generations an unconscionable amount of time has been invested in purely theoretical arguments with questionable empirical content. But it follows from the argument above that this bad choice of linguistic preference should not be confused with the empirical issue of how abstract the *real* distinctions are in the linguistic conventions of a particular language. In English, there is a progressive aspect that has become highly generalized, while in Danish the translational equivalents are spread across a cluster of constructions, each of which is much less abstract and general. Pursuing an unrelenting preference for either abstract generalization or for individual item-specific collocations would blind the grammarian to the empirical differences between the two languages.

The error of the structuralist approach was that it took for granted the existence of a small set of highly abstract elements and contrasts between them. This led to theoretical descriptions that left empirical facts almost out of sight. The converse error would be to stay content with describing constructions one at a time, giving up the search for those abstractions that are actually embodied in linguistic conventions.²

This article has addressed the issue of how to describe paradigmatic contrasts in a way that integrates accountability towards concrete usage with a search for structural contrast, and is solidly anchored both in cross-linguistic, typological facts, and in respect for the integrity of the conventions of the specific language that constitutes the object of description. We see this as a contribution to the same linguistic goal, based on the same ideals, that Carl Bache has pursued, when for instance he subjects the structural pattern embodied in Systemic-Functional Linguistics to a critique that demonstrates that the system has been abstracted too far away from actual usage patterns to capture them optimally, cf. Bache (2008).

Bache's distinction between a level of maximal abstraction and a 'functional' level is an example of how the grammarian can and should include the co-existence of different levels of abstraction within the same account. In suggesting how the most revealing maximal abstractions can be anchored in typological facts rather than solely in the linguist's own 'definitions', we hope to underpin rather than undermine this project – and similarly, in continuing down the scale of abstraction, beyond the functional level, down to the level of usage granularity where accidental variation begins to take over, we hope to be more explicit about the whole cline within which linguistic abstractions belong. Our proposals are therefore congenial to Carl Bache's, even if they suggest that his proposals could gain by being anchored in types of facts and descriptive procedures that go beyond his own basic system.

Kasper Boye
 Dept. of Scandinavian Studies
 and Linguistics
 University of Copenhagen
 Njalsgade 120
 DK-2300 Copenhagen S

Peter Harder
 Dept. of English, Germanic
 and Romance Studies
 University of Copenhagen
 Njalsgade 120
 DK-2300 Copenhagen S

Notes

1. The corpus figures cited here and below are rough-and-ready in that they involve only the 'bare' noun + determiner combinations, i.e. they ignore combinations with a premodifier between the definite article and the noun (cases like *the soaring inflation*). However, a search on *the inflation* with one word in between *the* and *inflation* yielded only 127 cases, of which half involved the collocation *inflation rate*. Similarly, *the* (□) *unemployment* occurred 178 times, of which more than a third involved

collocations like *unemployment rate/figure/statistics/problem*. We therefore believe that figures that include premodifiers would not materially alter the pattern.

2. Let us give an example from Bache's work, involving an insightful correction, based on a very abstract analysis, of previous work on English tense by one of us. Harder (1996) suggested a description of the contrast between the primary tenses past and present, which analysed that contrast as equipollent (each having a substantial semantic value), while in the case of future and perfect the contrasts were privative, i.e. a question of presence vs. absence of a substantial semantic value. As pointed out by Bache (2008), however, the 'equipollent' element can be understood as involving 'finiteness' rather than tense. The property that the English primary tenses share is that of deictic anchoring – and 'anchoring' in itself has no contrast built in. If that is abstracted out of both past and present tense in English, the only difference that remains between them is the presence vs. absence of a marked element, namely 'past location'. Analysing the past as marked is of course nothing new, but this analysis has traditionally been based on the fallacious understanding of the present tense as having no privileged link to the deictic now, cf. Harder (1996:341). Bache's analysis maintains the link between present tense and 'now', but because finiteness is abstracted out, remaining at the deictic centre becomes an unmarked, non-substantial choice on a par with non-choice of the future.

References

- Bache, Carl. 1985. *Verbal Aspect: A General Theory and its Application to Present-Day English*. Odense: Syddansk Universitetsforlag.
- Bache, Carl. 1994. Verbal categories, form-meaning relationships and the English perfect. In C. Bache, H. Basbøll & C.E. Lindberg (eds.) *Tense, Aspect and Action: Empirical and Theoretical Contributions to Language Typology*. Empirical Approaches to Language Typology 12. Berlin and New York: Mouton de Gruyter. 43-60.
- Bache, Carl. 1995. *The Study of Aspect, Tense and Action: Towards a theory of the Semantics of Grammatical Categories*. Frankfurt: Peter Lang.

- Bache, Carl. 1997. *The Study of Aspect, Tense and Action: Towards a theory of the Semantics of Grammatical Categories*, 2nd ed. Frankfurt: Peter Lang.
- Bache, Carl. 2008. *English Tense and Aspect in Halliday's Systemic Functional Grammar. A Critical Appraisal and an Alternative*. London: Equinox.
- Boye, Kasper. 2001. The force-dynamic core meaning of Danish modal verbs. *Acta Linguistica Hafniensia* 33. 19-66.
- Boye, Kasper. 2005. Modality and the concept of force-dynamic potential. Alex Klinge & Henrik H. Müller (eds.). *Modality. Studies in form and function*. London: Equinox. 49-80.
- Boye, Kasper. 2010a. Semantic maps and the identification of cross-linguistic generic categories: Evidentiality and its relation to epistemic modality. Michael Cysouw, Martin Haspelmath & Andrej L. Malchukov (eds.). *Semantic maps: Methods and applications. Linguistic Discovery* 8.1. 4-22.
- Boye, Kasper. 2010b. Cross-linguistic generic categories are linguists' generalizations: Author's reply to 'Semantic Maps as Helpers in the Quest for Generic Categories' (Veselinova 2010) and 'Positing Grammatical Categories: Linguists' vs. Speakers' Generalizations' (Cristofaro 2010). Michael Cysouw, Martin Haspelmath & Andrej L. Malchukov (eds.). *Semantic maps: Methods and applications. Linguistic Discovery* 8.1. 31-34.
- Boye, Kasper. 2012. Epistemic meaning. A cross-linguistic and cognitive study. Berlin: Mouton de Gruyter.
- Boye, Kasper & Peter Harder. 2012. A usage-based theory of grammatical status and grammaticalization. *Language* 88, 1. 1-44.
- Bybee, Joan, Revere Perkins and William Pagliuca. 1994. *The evolution of grammar. Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- Croft, William A. 2009. *Toward a Social Cognitive Linguistics*. In V. Evans & S. Pourcel (eds.). *New Directions in Cognitive Linguistics*. Amsterdam/Philadelphia: John Benjamins. 395-420.
- Croft, William A. & D. Alan Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Durst-Andersen, Per. 2000. The English progressive as picture description. *Acta Linguistica Hafniensia*, vol. 32. 45-103.
- Engberg-Pedersen, Elisabeth, Michael Fortescue, Peter Harder, Lars Heltoft, Michael Herslund & Lisbeth Falster Jakobsen (eds.). 2005. *Dansk Funktionel Lingvistik: En helhedsforståelse af forholdet mellem sprogstruktur, sprogbrug og kognition*. Copenhagen: University of Copenhagen.

- Hansen, Erik & Lars Heltoft. 2011. *Grammatik over det danske sprog*. København: Det Danske Sprog- og Litteraturselskab.
- Harder, Peter. 1996. *Functional Semantics. A Theory of Meaning, Structure and Tense in English*. Trends in Linguistics. Studies and Monographs 87. Berlin: Mouton de Gruyter.
- Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and crosslinguistic comparison. In Michael Tomasello (ed.). *The new psychology of language: Cognitive and functional approaches to language structure 2*. London: Lawrence Erlbaum. 211-242.
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86 (3). 663-687.
- Hjelmslev, Louis. 1943. *Omkring sprogteoriens grundlæggelse*. København: Københavns Universitet. (English Translation, 1953: *Prolegomena to a Theory of Language*. Bloomington: Indiana University).
- Kroeger, Paul R. 2005. *Analyzing grammar: An introduction*. Cambridge: Cambridge University Press.
- Lightbown, Patsy M. & Nina Spada. 2006. *How Languages are Learned*. Oxford: Oxford University Press.
- Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Majid, Asifa & Melissa Bowerman (eds.). 2007. Cutting and breaking events: A crosslinguistic perspective. Special issue of: *Cognitive Linguistics* 18 (2).
- Newmeyer, Frederick. 2010. On comparative concepts and descriptive categories: A reply to Haspelmath. *Language* 86.3. 688-695.
- Palmer, Frank R. 1979. *Modality and the English modals*. London: Longman.
- Perkins, Michael. 1983. *Modal expressions in English*. London: Frances Pinter.
- Quirk, Randolph et al. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Saeed, John I. 2008. *Semantics*, 3rd ed. Oxford: Blackwell.
- Schousboe, Steen. 2000. The Endless progressive. *Acta Linguistica Hafniensia*, vol. 32. 105-119.
- Timberlake, Alan. 2007. Aspect, tense, mood. In Timothy Shopen (ed.) *Language typology and syntactic description, Vol. III: Grammatical categories and the lexicon*. Cambridge: Cambridge University Press. 280-333.
- Whorf, Benjamin L. 1956 [1938]. Some verbal categories of Hopi. In John B. Carroll (ed.). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Chicago: MIT Press. 112-124.

