Analyzability and semantic associations in referring expressions : a study in comparative lexicology

Urban, M.

Chapter 1

Introduction

The seminal moment to which this thesis owes its existence came in the summer of 2006 when I was a M.A. student in linguistics at the University of Cologne, long before I learned about the (far-flung) relevant literature and that there is something called "lexical typology." In the hall in front of the seminar room were old yellowed working papers of the institute from the seventies, lying there on a table waiting for someone to come by and take them away, which I did (in spite of suffering from a dust mite allergy, and these papers were *very* dusty). Among them were several that dealt with "descriptivity," by which the researchers understand a certain type of analyzability of lexemes. I then read Seiler (1975), in which this research program was introduced and explained. In this paper, I stumbled upon the following "descriptive" term for 'stone' from Cahuilla, a Uto-Aztecan language of California on which Seiler did fieldwork:

(1.) *qáw-iš*
harden-PFV (adapted from Seiler 1975: 24)

Cahuilla *qáwiš* is deverbal, formed by addition of a perfective aspect suffix to the verb stem: 'that which has hardened' (although, according to Seiler, it is conventionalized with the meaning 'stone' only nowadays). A couple of weeks later, I learned about the etymology of modern Germanic words for 'stone' completely by chance in a seminar. They can convincingly be shown to go back to the following Proto-Indo-European structure:

(2.) **stāi-no*
harden-PFV (adapted from Pfeifer 1993: 1353)

I was simultaneously intrigued by two facts: first, that languages actually do exist which have morphologically complex terms for such basic vocabulary items as 'stone,' and secondly, that two geographically, culturally, and temporally as remote languages as Cahuilla and Proto-Indo-European should have chosen to conceptualize the same meaning in precisely the same manner.

This anecdote serves to summarize the main concerns of the present thesis. First, it seeks to assess differences between languages in the number of morphologically complex lexical items. Scattered in the literature, one finds unsystematic statements about geographically and genetically widely dispersed particular languages and a felt preponderance for morphologically complex terms. At times, these statements make particular reference to the nominal domain, which will also be the focus of the present work (see chapter 3 for the reasons). Thus, Seiler (1976: 6) writes about Cahuilla (Uto-Aztecan):

> The analysability nd morphological transparency of a considerable portion of all nominal expressions ... is immediately recognizable.

Pawley (1993: 99), while mostly concerned with complex verbs, states that in Kalam (Trans-New-Guinea?),

> [t]here are no simple nouns for some conceptual categories which one might expect to be universal, such as those roughly translated by 'person, human being', 'parent', 'child', 'ancestor', 'enemy', 'rain', 'thirst'.

On Yélî Dnye, a language isolate spoken on Rossel Island, an island belonging to the Louisiade Archipelago located southeast of New Guinea, Levinson (2006b: 230) remarks that

> Yélî Dnye is a language where many important, commonly employed nominal concepts are expressed with compounds.

The clearest statement is made by O'Meara and Bohnemeyer (2008: 332-333) for Seri, an isolate of Mexico (sometimes affiliated with the putative Hokan family):

> Complex expressions ... are in fact pervasive in the Seri nominal lexicon. ... It is thus clear that paucity of monomorphemic lexicalization and compensatory use of complex descriptive terms is a general typological characteristic of the nominal lexicon of Seri.

More generally, Mithun (1999: 287), in a discussion of ritual speech registers, states that in North American languages

> [s]ingle words are often composed of many meaningful parts, and their literal meanings are in many cases still perceptible to speakers. Indeed, words in all domains are frequently coined from complex descriptions.

What these assorted statements show is that the differential degree of morphologically complex terms clearly is a typological variable that has not received much attention by present-day comparative linguists so far (though it is sketched prominently in Saussure 1916/1967, see chapter 2 for review), and they suggest that there are clear differences between languages here waiting to be systematized. This has not been done so far in a principled approach, in spite of a general awareness by typologists of these differences, as revealed by remarks such as Aikhenvald's (2007: 21), who observes that "[l]anguages differ in how much derivational motivation (and hence derivational complexity) they allow for individual words. ... Decomposable terms in some languages can correspond to non-decomposable ones in others." Therefore, the questions to be addressed in this work include: are there significant differences between the languages of the world with respect to the degree of morphologically complex terms in the lexicon, possibly correlating with the affiliation to a particular language family or a linguistic area? Is the predominance of simplex lexical items in the better-known European languages an "atypical phenomenon"

(Sasse 2001: 503), going back to extensive language contact and concomitant lexical borrowing? What, if any, is the role that the language-specific means of word-formation have to play? Is it appropriate to postulate a typological trait, as O'Meara and Bohnemeyer (2007) suggest for Seri, for languages with a pronouncedly high degree of complex formations in the lexicon also for other languages, and how, considering also the grammatical features and their interaction with the lexicon, could this trait be reasonably delimited? Likewise, and perhaps even more importantly, possible causes for these differences also essentially remain in the dark, and an important aspect of the present study will be to give reasons for the behavior of individual languages by searching for typological correlates, paying particular attention, as suggested by Dixon (2010: 257), to structural properties of the languages.

As to the second main aspect of the present work, differences and similarities in the semantic structures found in morphologically complex terms, some work is available, at least for individual meanings and semantic domains (see chapter 2 and dispersed references in chapter 6 for a more extensive discussion of the state of research and cross-references). However, the lexicon is vast, and there are clearly many more recurring patterns to be discovered. Parallels like those in (1.) are remarkable in so far as the terms and their underlying structure have most likely arisen independently of one another. Canart (1979: 66), noting the frequent semantic extension of 'skin' to 'fur,' 'feathers,' and 'scales,' says that "[w]hat is perhaps curious is that ordinary people all over the world discovered a number of these fundamental truths quite independently and in the most unrelated languages and cultures" (again, see chapter 2 for a more thorough review of such statements).

Note that Canart's statement about similarities in semantic extension pertains to terms not characterized by morphological complexity, but rather to monomorphemic simplex lexical items. Indeed, from the semantic point of view, semantic associations not realized by morphologically complex terms but by semantic extension are just as interesting. Under this aspect it is not fruitful and even artificial to keep semantic associations by morphologically complex terms and by polysemy strictly apart, as is argued in chapter 3, which will expose the minimal theoretical framework of the study. Following Koch and Marzo (2007), a useful cover term that is used in this study for ties between words and the meanings they convey, regardless whether they are realized formally by word-formation relations or by polysemous or ambiguous conflation in a single monomorphemic lexical item is LEXICAL MOTIVATION,[1] and the property of such terms consequently is that they are LEXICALLY MOTIVATED. Although it also has a prehistory briefly touched upon in chapter 2, motivation as a current term in linguistics goes back to Saussure (1916/1967), who also includes onomatopoeia under this umbrella term, and the qualifier "lexical" hence is to highlight the fact that this type of motivation is established not directly by sound symbolism, but by a mirroring of a semantic relation on the level of linguistic expression, that is, the lexical item.

Armed with this general concept allowing for investigation of motivated words of both types, there are questions one can ask about the semantic side of motivated lexical items: can the universal tendencies in the (cross-)linguistic realization of certain mean-

[1] As a convention, technical terms will be printed in small caps when first introduced throughout this work.

ings that are beginning to show on the horizon be consolidated and can they also be found for other meanings than those already investigated (see chapter 2 for review). How strong are the tendencies in each case? How do recurrent conceptualizations look like, and is it perhaps even possible to learn something from them about human cognition? Conversely, which patterns are rare, only found in a few languages? Are there, next to universal trends, also patterns that are peculiar to a certain area (and if so, what is their history?) or peculiar to certain grammatical properties (lexicon-grammar-interaction)? Generally, the spirit of the approach adopted is that of Matisoff (2004: 385):

> Are human thought processes as reflected in language everywhere the same? The extremes of relativism and universalism are equally to be avoided, in favor of an empirical approach that appreciates in equal measure the nuances of how languages differ and resemble one another.

Another perspective one can assume is to depart from a certain meaning and to ask what structure the terms designating it have cross-linguistically. How are certain semantic fields, such as meanings revolving around 'fire' and 'water,' organized cross-linguistically?

Matisoff (2004: 384-385) calls for a "massive international effort ... to create a master database of semantic associations in the world's languages." The present work surely cannot offer such an effort, but the results are initial steps in this direction.

Throughout, and concerning both main aspects just mentioned, the present work also seeks to shed light on the typological, historical, socio-cultural and, occasionally and very cautiously, also cognitive background of the similarities and also of the differences if they are found.

When attempting to situate the present work into the context of related typological research, it is clear that it is part of what has come to be called "lexical typology." This field has only in the past few years, with a workshop at the conference of the Association of Linguistic Typology in 2007 in Paris and publications such as Vanhove (2008), come to be recognized as a subfield of typological investigations of languages on its own (maybe, as Haspelmath 2003: 211 says, because "many linguists regard the study of grammar as more interesting and prestigious" when contrasted with the lexicon, perhaps an aftereffect of Bloomfield's 1933: 274 (in-)famous dictum that "[t]he lexicon is really an appendix of the grammar, a list of basic irregularities"). Luckily, the volume has received attention in the community, as evidenced by reviews such as Citarrella (2010), Newman (2010), and Traugott (2010), who in particular notes that previously available suggestions and hypothesis were based on either evidence from one language only or from a small set of languages. This thesis, having an explicitly cross-linguistic orientation, is another step towards remedying this situation. However, as with any young field of research, "lexical typology" is presently characterized by a multitude of coexisting methodologies which are based on varying background assumptions about the nature of meaning, the lexicon, and the relation between the two, with at times decidedly heterogeneous research goals (contrast the conceptions of Lehmann 1990 with that of Behrens and Sasse 1997, Koch 2001,

and with that in the work of Talmy, whose overview article on lexicalization patterns, Talmy 2007, notably bears the title "lexical typologies").

Given this orientational pluralism and the exploratory nature of the thesis, it is primarily data-driven, not theory-driven. This means that the theoretical framework to be developed in chapter 3 is intended to categorize formal and semantic relations typologically, but in a way that does not distort the data from a particular point of view (such as a particular theory of morphology, word-formation, or lexical semantics) and to allow the data to speak for themselves. It also means that the generalization and results to be presented in chapters 5 and 6 are to a large extent generated out of the data themselves, rather than through the application of general preconceived assumptions about the nature of "language" (in the sense of *langage*) or the cognitive infrastructure that renders it possible.

If lexicology is concerned with the study of the structure of the lexicon of a particular language, the topics dealt with here could be labeled for the time being and to use a term coined by Tappolet (1895: 2), as questions of COMPARATIVE LEXICOLOGY (in contrast to research programs concerned with semantics proper, such as Levinson and Meira 2003): the cross-linguistic comparative investigation of the structuring of the lexicon, both of the formal structure of the words it contains, the semantic fields they belong to, as well as of the ties these structures betray to other elements of the lexicon, and, in a second step, to ask why the lexicon is organized in this or that way in different languages and to motivate this behavior, be it on grounds of language-internal, areal-typological, or extra-linguistic reasons. Eventually, once the field of lexical typology is more consolidated, comparative lexicology might become one of its subbranches.

To be sure, this introductory discussion is merely meant as a brief panoramic vision of the topics to be discussed and to provide an attempt to roughly situate them in the context of existing research. As such, it does not do full justice to the copious amount of literature that exists on questions related to lexical motivation, and therefore, before setting out the framework of the present study in more detail and presenting its results, it is appropriate to delve further into the (pre-)history of the topic. This is undertaken in chapter 2. Readers not interested in such a discussion can skip this chapter and continue on to chapter 3 immediately without losing crucial information for the understanding of the framework presented there and hence for the rest of this book.