

On representing semantic maps

Ferdinand de Haan
University of Arizona
July 8, 2004

Introduction

This paper describes a technique for dealing with data which does not lend itself very well for a description with traditional terminology. There are areas of language for which there exist a wide variety of terminology. For instance, the area of modality offers a bewildering set of terms for what is essentially the same data. To give a small flavor, (1) below lists some of the terms used for what is traditionally termed *deontic modality* (from De Haan forthcoming):

- (1) deontic modality
root modality
'containing an element of will'
dynamic modality
agent-oriented modality
subject-oriented modality
participant-oriented modality
non-epistemic modality

Each of these terms was proposed to make small distinctions between it and other similar terms, but in actual practice, these terms are being used interchangeably in the literature. There are actual differences between these terms, but these differences are too small to cause shifts in terminology.

On the flip side, there are terms which are in general use, but for which there is no accepted set of data. One such example (the example used in this paper) is the area of *irrealis*, or the marking of unreal events. While that seems to be a relatively coherent area, in actual practice the use of the term *irrealis* is subject to what amounts to whim.

This paper is devoted to a discussion of the use of semantic maps to counteract the need for terminological multiplication and for a better representation of linguistic data. Semantic maps have been in use for quite some time (the first major use of maps can be found in Anderson 1982), but not until recently, with its inclusion in some functional theories of language, have they become more important. There is as yet no agreed upon architecture for semantic maps and this paper addresses that issue as well. Please note that this is work in progress and that the latest version of this paper can be found at <http://www.u.arizona.edu/~fdehaan/papers/semmap.pdf>.

Semantic maps


One of the most recent models that try to come to grips with the complex interactions of semantic meanings in the world's languages is the *semantic map* model: a representation that is the sum total of the semantic possibilities of the category under investigation. An exponent of this category in a given language can then be mapped onto this representation and thus be compared to similar means of expression in other languages. This model is also known under the terms *mental map*, *mental space*, or *semantic space*. While it has been claimed that semantic maps may be a direct representation of the way in which the mind classifies linguistic categories, it is best to view semantic maps as tools for linguistic representation, similar to, say, X-bar representations or predicate and propositional logic. Also, this paper uses semantic maps as a tool for descriptive purposes. Semantic maps can also be used for diachronic purposes, to predict language change. This is an important use of semantic maps, but not treated in this paper.

The literature on semantic maps is ever growing. The following is a small selection of relevant papers. An easy introduction to semantic maps is Haspelmath (2003), from which some examples have been taken in this paper. For a full discussion on the usefulness of semantic maps in typology see Croft (2003:133ff). Some areas of language for which semantic maps have been proposed are: the perfect (Anderson 1982), evidentiality (Anderson 1986), voice (Kemmer 1993), case (Croft 1991), coming and going (Lichtenberk 1991), modality (Van der Auwera and Plungian 1998), and indefinite pronouns (Haspelmath 1997). In addition, semantic maps play a prominent role in Radical Construction Grammar (Croft 2001).

Semantic maps are not treated the same way in these studies. There are significant differences in the geometry of semantic maps in the aforementioned studies.

Geometry of semantic maps

This section deals with the necessary components for semantic maps. In its simplest form, a semantic map consists of a number of grammatical functions plus a means to link these functions together, as appropriate. An example is shown in (2), where the grammatical functions are represented with letters, and the linking device is either a line (2a) or is an enclosing shape (2b). In this example, functions A and B are expressed by one and the same morpheme in this hypothetical example. Function C is expressed with a different morpheme. Note that the length of a line is not relevant, nor is the precise shape of the enclosing form.

- (2) a. A ----- B C
b.  C

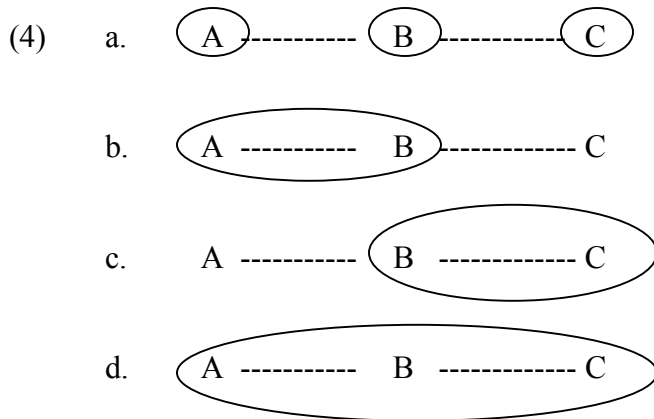
There are various possibilities for linking grammatical functions, including a combination of (2a) and (b), shading and coloring, and so on. In this paper, we will adopt the approach

by Haspelmath (1997, 2003), which uses *enclosing shapes* for functions expressed by one and the same morpheme in a given language, and *lines* to denote proximity of grammatical functions.

An example of how this works is shown in (3). Again, there are three grammatical functions to be mapped.

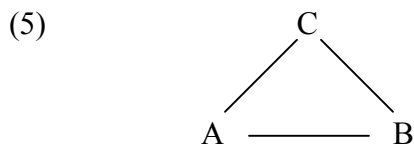
(3) A ----- B ----- C

The interpretation of this mini-map is that function A and B are closely related, as are B and C. Consequently, this map makes the prediction that the following types of languages are attested:



That is, there should only be languages in which either: all functions are expressed with different morphemes (4a), functions A and B are expressed by one and the same morpheme and C by a different one (4b), functions B and C are expressed by one and the same morpheme, and A by a different one (4c), or, all functions are expressed by one and the same morpheme (4d). The one possibility we should not find is one in which functions A and C are expressed by one and the same morpheme, and B by a different one. The map in (4) is a visual representation of the fact that functions A and C are further apart than either A and B, or B and C.

Should it transpire that a language is found in which A and C are expressed by one and the same morpheme, and B by a different one, we need to amend the semantic map to take into account that A and C are no longer further apart than any of the other possibilities. This is shown in (5):



This introduces a second dimension as it is no longer possible to draw a one-dimensional semantic map like the one in (4). The map in (5) makes the prediction that all combinations of functions are possible and attested in languages. Haspelmath (2003:217-

8) calls such maps *vacuous* maps, as they fail to eliminate any combinatorial possibilities. Nevertheless, such examples do occur in real life and they must be provided for.

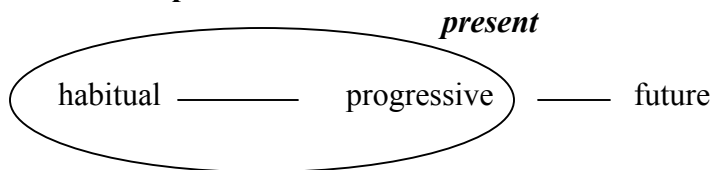
To give a concrete example of how semantic maps work, and the problems that can arise, we will discuss an example from Haspelmath (2003:236-7).

It has long been noted that there is a relationship between tense and aspect in that languages can use the same morpheme for certain tenses and aspects. For instance, habituais, progressives and futures can often be expressed with the same morpheme. This can be put in a simple grammatical map as follows (this map plots three grammatical functions and is therefore identical to (3) above):

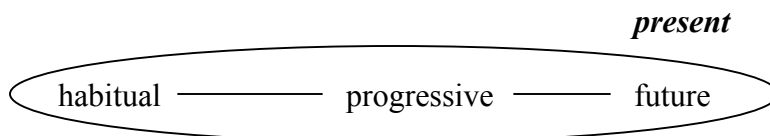
(6) habitual ——— progressive ——— future

This map makes the prediction, therefore, that there are four possibilities as far as combinations of functions is concerned, namely the ones shown in (4) above. To give just two attested languages:

(7) a. **Spanish**



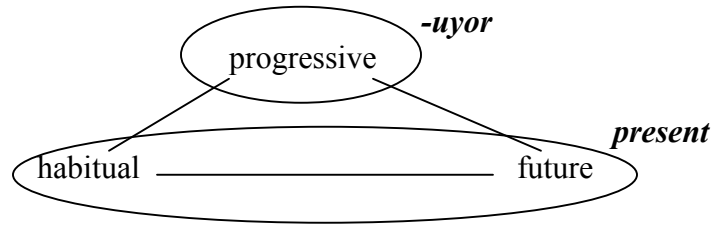
b. **German**



Thus, the Spanish Present tense is used for habitual and progressive aspect, but not for future tense, while the German Present tense is used for all three functions.

A problem arises with Turkish. Diachronic data show that the Present tense was originally used for all three functions (thus, identical to the German Present in (7b)), but a new Progressive morpheme was introduced, so that the original Present now refers to habitual and future, which is precisely what is not predicted by the map. At this stage we have a number of options to deal with the Turkish data, depending on how we wish to introduce the diachronic dimension, but for the present purposes we need to face up to the fact that synchronically we must abandon the one-dimensional map (6) and go to a two-dimensional map, like the one shown in (8)

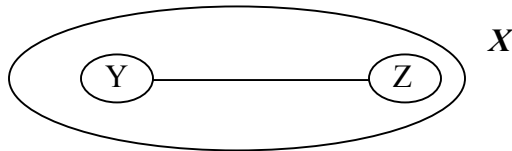
(8) **Turkish**



The semantic maps for German and Spanish need to be altered as well, in order to take the new geometry into account. Because of Turkish, the map is rendered vacuous but this does not need to be disastrous, given that this map is part of a larger map, which could well change things again (note, for instance, that the function of *present* tense has not been accounted for in this map!).

A further issue concerns the notion of the functions itself. We can ask: what constitutes a function to be represented on the map? It can well be argued that, to use the map of (6) again, whether the functions “progressive”, “habitual”, and “future”, are semantic maps in themselves. For instance, are there languages in which “progressive” is further subdivided, say into “present progressive” and “past progressive”? This is a valid point and we must ensure that any function on a semantic map is *primitive*. That is, it must conform to the following informal definition.

(9) A function X is primitive if it is **not** the case that



That is, a function X is not primitive if it can be subdivided into two (or more, not drawn) functions that are expressed by two separate morphemes in some language. Conversely, if we had postulated two categories Y and Z that are never expressed by two separate morphemes, they are not primitive and can be conflated into one category X.

One other possibility is the situation in which one function is expressed by two different morphemes. This is not a problem and can be accounted for as in the following example:

(10)



In this case there is one morpheme that expresses both functions A and B, and a second morpheme that expresses just function B. This can easily be extended to overlapping

situations, e.g., one morpheme for functions A and B, and a second for functions B and C.

Irrealis

In this section we will apply the semantic map theory to the notion of *irrealis*. This category is eminently suited for such an analysis, as is modality as a whole. Part of the problem in studying issues of modality is the fact that there is as yet no coherent framework suited for dealing with modality in natural language, especially from a cross-linguistic point of view (see De Haan, forthcoming, for some of the problems). It is especially difficult in that there is no agreed upon set of terminology for basic modal notions. While some scholars use terms such as *epistemic* and *deontic* as basic terms (which are terms used in logic), others use more linguistically oriented notions such as *agent-oriented* modality. There is a plethora of terminology and one of the downsides is that comparisons between languages or language families become cumbersome as it is very often the case that scholars who work on a specific language family have developed a set of terms which does not correspond with similar terms used by others.

A case in point is that of the *realis* – *irrealis* distinction. These are terms widely used by various scholars and in various grammars. The core of the distinction is a desire to distinguish between real events and unreal ones. That is, events that are or have happened versus events that did not or have not happened but which are possible, probably, hypothetically likely, or could have happened. As can be seen from the list (which is by no means extensive) is that irrealis notions cover a wide range of categories while realis is a relatively simple affair. The problem is that the term *irrealis* is used in grammatical descriptions in such a way that normally only a subset of irrealis notions is covered by presumed irrealis morphemes (see Bybee 1998 for a good description of the problem). The problem is not just limited to morphemes that are called irrealis; other types of morphemes, such as *subjunctive* or *optative* morphemes are affected in the same way. Palmer (2001) devotes large portions of his discussion to the problem of how to link subjunctive with irrealis, for instance. The immediate consequence is that it is *a priori* impossible to compare irrealis morphemes from one language to the next.

This has prompted Bybee (1998, among others) to effectively ban the term irrealis from grammatical description, as there is no one-to-one correspondence between irrealis and unreal events in languages. That is, a sentence marked for [-irrealis] is not by definition [+realis]. Instead, she advocates searching for more meaningful terms for an “irrealis” morpheme in a given language. That is easier said than done. Let us take a look at the distribution of the Irrealis morpheme *-ji* in the Australian language Maung (Capell and Hinch 1970:67, also discussed in Bybee 1998). The following chart shows the assumed distribution:

(11)

	PRESENT:	Indicative present, future
REALIS	IMPERATIVE:	Negative only (i.e., prohibitive)
	PAST:	Simple and complete past, imperfect

IRREALIS	PRESENT:	Potential, negative present and future
	PAST:	Negative past, Conditional, Imperative

A sample paradigm is shown in (12):

(12)	Realis		Irrealis
	ŋi-udba	<i>I put (pres.)</i>	ni-udba- ji <i>I can put</i>
	ŋi-wan-udba	<i>I shall put</i>	marig ni-udba- ji <i>I don't put</i> or <i>I won't put</i>
	juwunji g-udba	<i>don't put it!</i>	da ŋi-udba- nji <i>if I put</i>
	ŋi-udba-ŋ	<i>I put it (perfect)</i>	marig ŋi-udba- nji <i>I didn't put</i>
	ŋi-udba-ŋ-uŋ	<i>I was putting</i>	g-udba- nji <i>put it!</i>

There are several points to be made here. Firstly, there is no one-to-one correspondence between realis and real events on the one hand and irrealis and unreal events on the other. Several categories that are grouped as realis ones could just as easily been irrealis categories, most notably the future and the prohibitive. Future events are as yet unrealized and prohibitives are used to make sure events do not become realized. From this perspective, Bybee (1998) is correct to point out that calling the morpheme *-ji* an Irrealis morpheme is not the best way of analyzing it. However, it is equally hard to come up with a label that does cover all possible functions of *-ji* (and only those). What, for instance, do negation and imperative have in common, beside the fact that the actions are unrealized in some sense?

Secondly, there are several functions that need to be accounted for. Coming up with an all-encompassing label for these functions is at the very least impractical and at the very worst meaningless. After all, we need to repeat the exercise for every language that has an irrealis or irrealis-like morpheme and if the functions that make up the irrealis morpheme change, we need to come up with a new label (since the make-up of the category has changed). But this robs us, very likely, of cross-linguistic comparisons.

Therefore, we need to be able to account for both the uniqueness of languages and for the ability to make cross-linguistic comparisons. A semantic map is capable of handling both aspects.

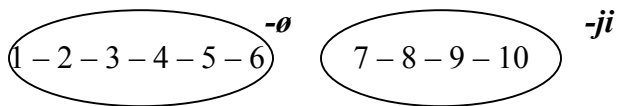
The first thing that needs to be done is to make a list of all functions that can possibly be expressed by an irrealis morpheme. From the Maung example above, we can glean the following categories:

(13)

- | | |
|----------------|----------------|
| 1. present | 7. potential |
| 2. past | 8. negation |
| 3. future | 9. conditional |
| 4. perfect | 10. imperative |
| 5. prohibitive | |
| 6. imperfect | |

Assuming that these functions are primitive, and some of them may very well not be, we can build a semantic map that looks like the one shown in (14), with numbers representing the categories in (13):

(14)



Of course, the map shown in (14) is just one of a large number of possible maps that can be drawn given the data from (13). Any possible permutation of the “realis” categories plus any possible permutation of the “irrealis” categories (plus their mirror images) would be a valid semantic map as each accounts for the given data. By examining more and more languages, we reduce the number of possible maps until, it is hoped, one map remains which accounts for the behavior of irrealis-type morphemes cross-linguistically. Individual languages then have their morphemes mapped onto the template. These mappings then represent contiguous areas on the basic template. That way, it is easy to compare “irrealis”-type morphemes across languages by comparing areas. Similarly, we can do away with a proliferation of distracting and prejudicing labels and can get on with the business of doing a proper linguistic analysis.

No one should underestimate the tremendous effort involved in getting from collecting primitive functions to producing a cross-linguistically valid map. It involves a painstaking analysis of individual grammatical descriptions and texts, especially for an heterogeneous category as modality or irrealis. To come up with an exhaustive list of primitive functions is far from easy. For instance, Bugenhagen (1994:5), in a study on irrealis morphemes in a variety of New Guinean languages, uses the following 15 functions:

(15)

Future	hypothetical conditional	counterfactual conditional
Purpose	wanting / intention	ability
Obligation	imperative	prohibitive
‘lest’	doubt	negation
habitual	present	past

This list adds to the number of categories of (13) by either adding new ones (such as *purpose*) or splitting old ones up into two or more categories (conditional is now *hypothetical* or *counterfactual*).

In order to show some of the advantages (as well as some of the problems), we will take a subset of these functions and use them to construct a semantic map of the irrealis in some New Guinean languages. The data and the functions that are used are shown in (16). The source is Bugenhagen (1994).

(16) **Languages**

Manam	Muyuw	Nakanai
Sursurunga	Sio	Sinaugoro
Mangseng		

(17) **Functions**

present	prohibitive	counterfactual
past	imperative	hypothetical
future		

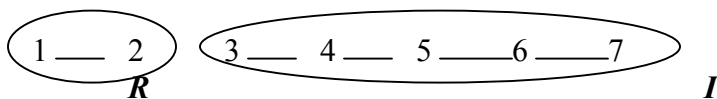
The specific data is shown in the following table. R = realis form, I = irrealis form, int. = intentive pronoun, data in italics are language-specific forms. Note that the data have been simplified here for explanatory purposes. It is not claimed that this is the definitive semantic map of irrealis. For instance, Manam actually makes use of more than one irrealis construction.

Table 1: irrealis data from selected New Guinean languages

	Man	Sin	Muy	Sur	Mang	Sio	Nak	code
Pres	R	R	unm/I	R	unm	unm	unm	1
Past	R	R	unm/R	R	unm	unm	unm	2
Fut	I	R	I	I	<i>aro</i>	I	<i>ge</i>	3
Imp	I	int.	unm	I	unm	I	<i>ge/unm</i>	4
Proh	I	int.	I	R	“don’t”	I	<i>ge</i>	5
Count.	I	I	I	I	<i>aro</i>	I	<i>ge</i>	6
Hyp	I	?	I	I	<i>aro</i>	I	<i>ge</i>	7

Since constructing semantic maps is a language-by-language endeavor, we will arbitrarily begin with Manam. The data from Manam allow for a very simple one-dimensional map, shown in (18).

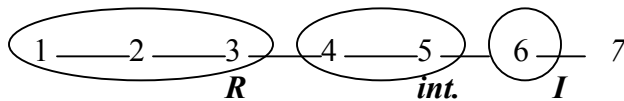
(18)



As discussed before, (18) represents just one of the possible semantic maps because any permutation of functions 1 and 2, and 3-7 is also a valid map. Also, the precise link between the realis and irrealis side needs to be worked out.

The data from Sinaugoro do a lot to flesh things out. For one, there are three constructions to be mapped, a realis, an irrealis, and one called intensive pronouns. The first thing that is noticeable is that the Future is a Realis category. Second, the Imperative and Prohibitive have a different marking and must therefore be grouped together. Note that data from the hypothetical are missing and so can't be used for mapping. This data leads us to the map shown in (19)

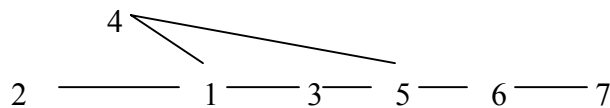
(19)



Compared to (18), the map in (19) is more accurate as it is now known that 3 is linked to 1 and 2 (with the precise order of 1 and 2 still outstanding). Also, 4 and 5 form a group, as does 6. The links between 3 and 4, and 5 and 6 are carried over from the previous map, as they are linked there. We can still represent the irrealis as a one-dimensional map.

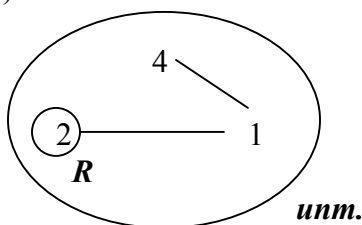
Data from Muyuw call for a drastic revision. The main opposition is between irrealis-marked functions and unmarked functions. Only the Past can be marked with a special Realis morpheme. The present can be either unmarked or marked with an Irrealis morpheme. The imperative (but not the prohibitive) is always unmarked. Because we must now link the imperative with the past and present, we must use a two-dimensional structure, as shown in (20)

(20)

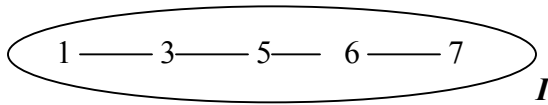


This map can then be used for realis/unmarked (21) and irrealis (22) [NB: this should be one map, but Word is not too crazy about overlapping circles, so I have used two maps – FdH]

(21)



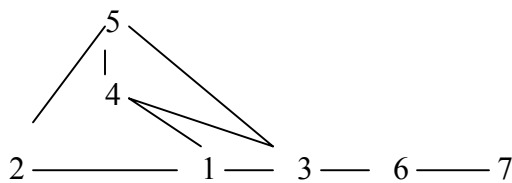
(22)



This map fixes the relative order of past and present, as only the present can be marked as Irrealis (and must therefore directly link to other irrealis categories). Note that 4 must link to both 1 (because of the fact that they are both unmarked in Muyuw) and 5 (due to the fact that they are marked similarly in Manam). The ambiguous position of 1, being both unmarked and irrealis, can serve as a point of further study. It might be the case that 1 is not a primitive function and that there is a further division to be made, or it could be that there is a choice whether to mark 1 as irrealis or to leave it unmarked.

We will forgo a detailed description of the remaining languages and present in (23) the final semantic map, based on this sample of languages.

(23)



This two-dimensional structure will account for the data in Table 1. Each language and its morphemes can be represented as a contiguous part of the map. It must be stressed that a semantic map is a tool, and should not be confused with an explanation. The advantage is that we can now compare morphemes across languages independent of possibly confusing and contradicting terminology.

Implementation

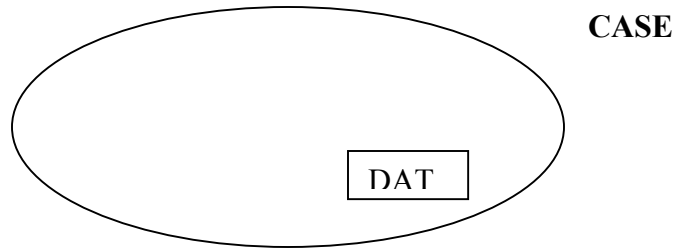
This section suggests a way to implement semantic maps. The first thing to note is that, unlike other representations, a semantic map is non-hierarchical. That is, there is no relation of the ISA-type (from *is a*). Examples of these would be:

- (24) a. Clyde is an elephant.
 b. Dative is a case.

ISA relations are well-known in the AI literature, and they form the basis for such systems as WordNet and, of course, various ontologies of language. An example such as (24b) would be represented as in (25) or (26):



(26)



These hierarchical considerations are not represented in semantic maps. While it is true that a semantic map is always based upon some domain (be it case, voice, or tense) which means that hierarchical relations could conceivably be built in, it is important to realize that in semantic map models nothing really follows from hierarchical relations. To give an example: from the fact that *dative* is a CASE, nothing follows from this fact as far as the semantic range of “dative morphemes” in individual languages is concerned. As the semantic map of the dative in Haspelmath (2003) shows, the only relevant factor is the semantic map of the dative itself, not the fact that a dative-map belongs to a larger case-map. Figure (23) above is the semantic map for the realis – irrealis domain. From this map it is not predictable which function is realis and which is irrealis. Indeed, in the languages surveyed, only one function can never be irrealis (Function 2) and only one can never be Realis (Function 6, this number will probably increase to 2 if Function 7 turns out to be irrealis in Sursurunga). No hierarchy can predict this.

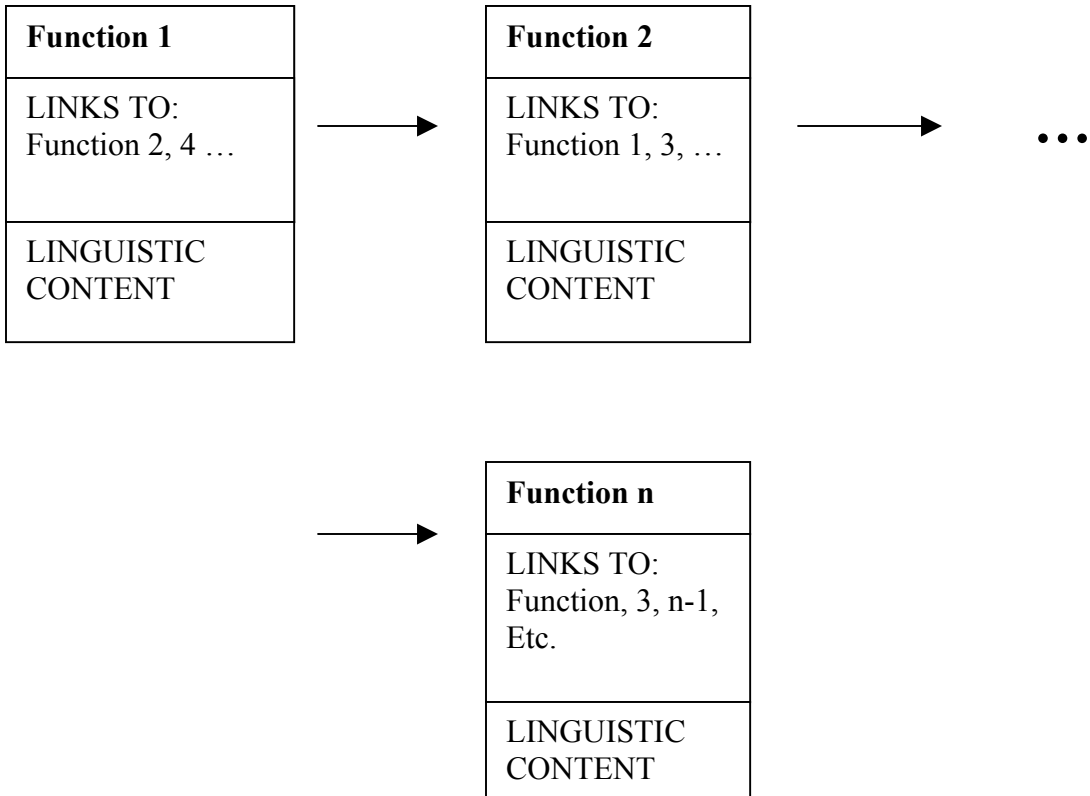
That does not mean, obviously, that the question of hierarchy is wholly unimportant. The relations between various domains can be of crucial importance for the interpretation of semantic maps, but, again, that does not mean that we must think that hierarchy is an important consideration in the drawing of semantic maps.

The same thing can be said for *prototype* analyses. There is no built-in mechanism to determine prototypes for given domains. That does not mean that it is an unimportant matter question. However, a prototype should ideally be determined by examining the data before determining the prototype. After all, in order to determine that a robin is a more prototypical bird than a penguin, we need to have examined most, if not all, bird species.

In our domain and sample, it would appear that Function 6 is the most prototypical “Irrealis” category and Function 2 the most prototypical “Realis” category. This is based on the fact that in all seven languages they are marked as such. Function 1 is less prototypically realis than Function 1, et cetera. This is useful information in typological studies (in order to answer questions like *how prototypically irrealis is the morpheme aro in Mangseng?*) but for the purposes of terminological considerations hardly necessary. Therefore, no mechanism needs to be built in.

The best way for representing semantic maps appears to be a linked list data structure in which the individual functions are represented as nodes and the connecting arcs as separate data entries to that node. A semantic map, like the one in (23), can then be represented as follows:

(27)



In order to represent the semantic map of (23), we can replace the schema of (27) with the actual situation and we can even add a field with the preferred name (label) of the node, if we so desire. The node for Function 1 of semantic map (23) is then as follows:

(28)

Function 1
PRESENT
Function 1 Function 3 Function 4
CONTENT

There is one other item of note in this representation. Note that the arcs connecting the various nodes are *bi-directional*. That is, if there is a field in Function 1 that states that there is an arc going to Function 2, then there is a corresponding field in Function 2 to show that there is an arc going from Function 2 to Function 1. strictly speaking, this is not necessary but it is done for reasons of clarity, symmetry and to give the ability to start at a random node in the representation (if the actual implementation allows us to do so).

Bibliography

- Anderson, Lloyd B. (1982). The 'perfect' as a universal and as a language-particular category. In Hopper (ed.), 227-64.
- Anderson, Lloyd B. 1986. Evidentials, Paths Of Change, And Mental Maps: Typologically Regular Asymmetries. In Chafe and Nichols, 273-312.
- Bugenhagen, Robert D. (1994). The semantics of irrealis in the Austronesian languages of Papua New Guinea. In Ger P. Reesink (ed.) *Topics in descriptive Austronesian linguistics*. Leiden: Rijksuniversiteit Leiden.
- Bybee, Joan L. (1998). "Irrealis" as a grammatical category. *Anthropological Linguistics* 40, 257-71.
- Capell, A. and H. E. Hinch (1970). *Maung Grammar: texts and vocabulary*. The Hague: Mouton.
- Croft, William (2001). *Radical Construction Grammar. Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Croft, William (2003). *Typology and Universals, second edition*. Cambridge: Cambridge University Press.
- De Haan, Ferdinand (forthcoming). Typological approaches to modality. In William Frawley (ed.) *Modality*. Berlin: Mouton de Gruyter. Draft on the web at <http://www.u.arizona.edu/~fdehaan/papers/typmod.pdf>.
- Gärdenfors, Peter (2000). *Conceptual spaces: the geometry of thought*. Cambridge, MA: MIT Press.
- Haspelmath, Martin (1997). *Indefinite Pronouns*. Oxford: Oxford University Press.
- Haspelmath, Martin (2003). The geometry of grammatical meaning: semantic maps and cross-linguistic comparison. In M. Tomasello (ed.) *The new psychology of language: cognitive and functional approaches to language structure, vol. 2*. Mahwah, NJ: Lawrence Erlbaum, 211-42.
- Hopper, Paul, ed. (1982). *Tense – Aspect: between semantics and pragmatics*. Amsterdam: Benjamins.
- Kemmer, Suzanne (1993). *The middle voice*. Amsterdam: Benjamins.
- Lichtenberk, F. (1991). Semantic change and heterosemy in grammaticalization. *Language* 67, 475-509.
- Van der Auwera, Johan; Vladimir Plungian. 1998. "Modality's semantic map". *Linguistic Typology* 2, 79-124.