



# LINGUISTIC DISCOVERY

DARTMOUTH COLLEGE

Volume 8  
Issue 1  
2010

## **Analyzing Semantic Maps: A Multifactorial Approach**

Andrej L. Malchukov

Max Planck Institute for Evolutionary  
Anthropology

doi: 10.1349/PS1.1537-0852.A.350

url: [http://journals.dartmouth.edu/cgi-bin/WebObjects/  
Journals.woa/1/xmlpage/1/article/350](http://journals.dartmouth.edu/cgi-bin/WebObjects/Journals.woa/1/xmlpage/1/article/350)

*Linguistic Discovery*

Published by the Dartmouth College Library  
Copyright to this article is held by the authors.

ISSN 1537-0852

[linguistic-discovery.dartmouth.edu](http://linguistic-discovery.dartmouth.edu)

# Analyzing Semantic Maps: A Multifactorial Approach

Andrej L. Malchukov

Max Planck Institute for Evolutionary Anthropology

*In this paper I argue that semantic similarity is not the only factor which motivates polysemy patterns cross-linguistically; I also show that these other factors (markedness, distinguishability, etc) may give rise to polysemies which are problematic for established semantic maps. Only when these other interfering factors, both functional and structural, are featured out, does a semantic network emerge and a “similarity map” reduce to a semantic map.*

## 1. Introduction<sup>1</sup>

There are two general approaches to semantic maps, which are also represented in this issue. These approaches are sometimes characterized as traditional vs. statistical approaches, as implicational vs. probabilistic maps (Wälchli, this issue), or as first- vs. second-generation maps (Sansò, this issue). In the traditional approach (conceptual) categories/functions are preselected, and a network of categories is constructed in such a way that the arrangement reflects formal similarities between adjacent categories. The connections between categories, represented in the form of lines, are usually taken as indicative of semantic overlap, which can be captured in terms of shared semantic components (e.g., features; cf. Zwarts, this issue). The traditional approach has been developed by Anderson, Croft, and Haspelmath, among others; the details are well known and need not be rehearsed here. In the alternative approach (maybe in its clearest form represented in the work by Wälchli, this issue) the categories are not preselected, and semantic maps are automatically generated from parallel corpora through the use of statistical scaling methods.

As noted by Cysouw (2007), the traditional approach faces a number of problems. First, it cannot represent frequencies of individual polysemy patterns.<sup>2</sup> A related problem is that as the amount of data increases vacuous maps become more and more widespread since frequent, rare, and exceptional patterns will all be represented on the map. Second, the traditional method is overgenerating since not all the predicted patterns are actually attested<sup>3</sup>. The most common response to the first problem by practitioners of the traditional approach is that semantic maps reflect most frequent polyfunctionality patterns in a certain domain; thus, exceptional patterns would be featured out in a larger sample. For example, Narrog & Ito (2007) do not represent polysemies of (instrumental) case markers found in less than 10% of cases. There may be more principled methods to distinguish between recurrent and exceptional patterns; for example, Rice and Kabata (2007) use Fisher’s exact test to determine which patterns of case polysemy

---

<sup>1</sup>Acknowledgement: I am grateful to Sonia Cristofaro for useful comments on the draft version of this article. The usual disclaimers apply.

<sup>2</sup>Note that although the traditional approach as it stands cannot represent *absolute* frequencies of individual polysemy patterns, it does capture *relative* frequencies of extended vs. restricted patterns (see Section 4 for further discussion).

<sup>3</sup>It is important to keep in mind that once the configuration of the semantic map is established it gives rise to predictions which go beyond the dataset on the basis of which the semantic map has been established. It is exactly this feature that makes (traditional) maps theoretically interesting and warrants the term “implicational maps” (Haspelmath 2003). Yet, this advantage turns into a disadvantage if certain patterns predicted to be possible (i.e. which are compatible with a semantic map) never turn up, even in extended samples.

(involving allative markers) are statistically significant. In other words, those polysemy patterns that are rare and have been overlooked in the early work on semantic maps, based on small scale comparison, are likely to be featured out in a larger sample as statistically insignificant. Indeed, maps proposed in earlier work are often confirmed in a larger sample. Thus, Narrog & Ito's (2007) study largely corroborated Haspelmath's (2003) map of instrumental and related functions, and Mauri's (this issue) map of coordination is "slightly different" from Malchukov's (2004) map of contrast markers as far as same categories are addressed. From this perspective, exceptional polysemies do not constitute a major problem for the semantic map approach because they would not make it into the map. While this is a legitimate approach, it involves data reduction; so the question arises as to what extent such data reduction is justified. This question can only be answered by analyzing motivations behind common vs. rare polysemy patterns. In this paper I will show that rare patterns are often not indicative of (immediate) semantic relatedness of respective categories, but are due to other factors. This does not mean that such minority patterns should be dismissed; on the contrary, they deserve to be analyzed in their own right as they can provide important insights into motivations behind polyfunctionality. Moreover, the problem of overgeneration, which has not been addressed in the literature so far, can arguably also be solved by invoking other factors inhibiting broader polysemies (see Section 4 for further discussion).

Another response to limitations of the traditional approach was the introduction of statistical scaling techniques (such as multidimensional scaling) that can represent relative frequencies of individual polysemy patterns more readily and involve less data reduction (Wälchli, this issue). Yet, as pointed out in the literature (Cysouw 2007; Zwarts, this issue, Narrog, this issue), the latter method has its own limitations. First, unlike traditional maps, the maps generated by multidimensional scaling (MDS) are not implicational, thus they cannot be used to constrain the data and therefore are "less interesting" (in this sense, they are more similar to vacuous maps in the traditional approach), nor can they readily represent a diachronic dimension (Narrog, this issue). Moreover, automatically generated maps are often difficult to interpret (which is somewhat ironic, given that they have been developed as a visualization tool) because the emerging picture is quite "messy" (this might be partly a consequence of the compression of multiple dimensions into the two-dimensional space; Cysouw 2007). Thus, it is often not clear whether there is a certain clustering pattern in the first two dimensions (cf. question-marked clusters in Sansò's contribution), and turning to further dimensions makes identification of clusters still more problematic (e.g., Wälchli, this issue, finds only the first three dimensions interpretable in his dataset). Thus, while MDS maps visualize general semantic dimensions or tendencies in a certain domain, they are less suited for capturing semantic connections between individual categories. Therefore, unlike traditional maps, MDS maps cannot directly feed semantic analysis. One reaction to this is Wälchli's (2007; this issue) proposal that conventional semantic analysis (including formal semantics) should give way to "similarity semantics". A less radical approach, advocated in this paper, is that semantic maps (whether constructed in a conventional way or automatically generated) can provide an important insight into semantics of the analyzed categories once the interfering factors that introduce "noise" are featured out. Only at this later stage can the semantic map feed the semantic analysis, aiming to account for the recurrent polysemy patterns in terms of shared semantic components. In this way semantic maps can become an important tool for semantic analysis and not just a visualization technique (a frequent criticism heard from opponents of the semantic map methodology).

In what follows I discuss a number of semantic maps proposed for different domains in the literature, including my own earlier work. The discussion will be intentionally programmatic as my goal is not to present the data, but rather to raise methodological questions relevant to the semantic map approach. In particular, I will argue that similar encoding may not reflect a semantic similarity but may be due to other factors, both functional (markedness, economy, distinguishability), structural, as well as diachronic. Only when these other interfering factors are featured out does a semantic network emerge. Thus, it might be useful to distinguish terminologically between a *similarity map* (e.g., obtained through MDS visualization technique) and a *semantic map*, which emerges once the “noise” is reduced. A similarity map is shaped by the cumulative effect of different factors contributing to polyfunctionality. A semantic map, on the other hand, represents, so to say, a semantic residue of the similarity map once the other factors are featured out.

## 2. Markedness Effects

Baerman et al. (2005) is a recent study of inflectional syncretism, in general, and of case syncretism, in particular. Since syncretism depends on a number of factors both functional and phonetic, it is not surprising that the patterns of syncretism are not consistent across languages. Thus, Baerman et al. (2005) conclude that only patterns involving syncretism of core cases are cross-linguistically consistent. Another generalization noted by Baerman et al. (2005) is that case syncretism shows strong dependency on number (in fact this is the only case of a strong interaction between inflectional categories noted for the nominal domain). Not surprisingly, case syncretism is found in non-singular numbers more frequently when compared to the singular. The following example from Sanskrit, where syncretism is more pervasive in the plural than in the singular and is more pervasive in the dual than in the plural, is representative in that respect.

‘god’	singular	Dual	plural
NOM	<i>devas</i>	<i>Devāu</i>	<i>devās</i>
VOC	<i>deva</i>	<i>Devāu</i>	<i>devās</i>
ACC	<i>devam</i>	<i>Devāu</i>	<i>devān</i>
INS	<i>devena</i>	<i>devābhyām</i>	<i>devāis</i>
DAT	<i>devāya</i>	<i>devābhyām</i>	<i>devebhyas</i>
ABL	<i>devāt</i>	<i>devābhyām</i>	<i>devebhyas</i>
GEN	<i>devasya</i>	<i>devayos</i>	<i>devānām</i>
LOC	<i>deve</i>	<i>devayos</i>	<i>deveṣu</i>

Table 1: Sanskrit *a*-stem noun (Baerman 2009, citing Whitney)

Does this particular pattern of case syncretism reflect semantic similarities between individual cases? Not obviously so—while nominative and vocative forms may be regarded as constituting a natural class, there is no clear semantic motivation for the syncretism of the dative and instrumental, or dative and ablative cases. Moreover, semantic similarities cannot explain why syncretism is more pervasive in non-singular numbers than in the singular. Indeed, there is no obvious reason why semantic roles should be better distinguished for individuals rather than groups.

The main contributing factor to the syncretism pattern exemplified by Sanskrit is well-known and has been described in terms of markedness. It has long been noted in the literature that inflectional possibilities of the marked member of a category are reduced, compared to the unmarked member (Greenberg 1966:27; Croft 1990; Croft 2003:77-80). Croft, following Greenberg, even regards such skewing as one of markedness diagnostics, which he dubs “inflectional markedness”. Now, markedness is not an unproblematic notion and can probably be reduced to other factors such as frequency (as suggested by Haspelmath 2006<sup>4</sup>); yet effects of inflectional markedness are pervasive and can be found in different domains. To cite another example from the nominal domain: recent work on syncretism patterns in pronominal systems has revealed that neutralization of person categories (“vertical homonymies”) occur more frequently in the non-singular forms of both free pronouns and bound pronouns (Cysouw 2003; Siewierska 2004; Baerman et al. 2005). Similar examples abound in the verbal domain as well—for example, tense-aspect-mood distinctions as found in (unmarked affirmative) verbal forms are frequently neutralized in the negative forms (see Croft 1990, Aikhenvald & Dixon 1998 for more discussion and exemplification of markedness effects).

The question is what consequences do the effects of inflectional markedness have for semantic maps. It seems that if semantic maps are designed to capture similarities in semantic functions, such markedness effects should be featured out since syncretization of inflectional forms in the marked member is not driven by semantic similarity in the first place.

### 3. Other Functional Factors: Economy and Distinguishability

There is a tradition of representing alignment patterns (ergative, accusative, neutral) in the forms of diagrams reminiscent of semantic maps. The first linguist to use these schemes was apparently Charles Fillmore, but since the 1970’s, in typological studies (by Bernard Comrie, R.M.W. Dixon, Frans Plank and A.E. Kibrik, among others) it has become conventional to represent alignment patterns in such a way. Thus typology textbooks (e.g., Whaley 1997; Song 2001) include the following representations of different alignments: accusative ( $A=S \neq P$ ; Figure 1), ergative ( $A \neq S=P$ ; Figure 2), neutral ( $A=S=P$ ; Figure 3), and tripartite patterns ( $A \neq S \neq P$ ; Figure 4).

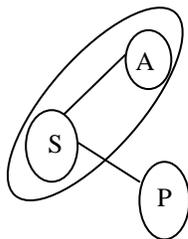


Figure 1

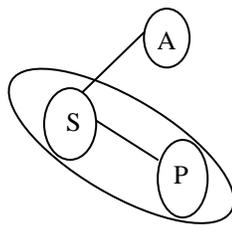


Figure 2

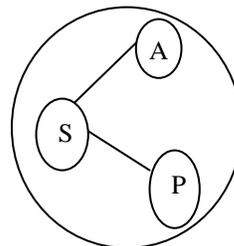


Figure 3

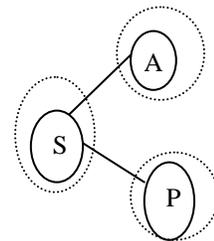


Figure 4

Originally, these schemes were not conceived of as semantic maps. For example, Kibrik (1985; cf. Plank 1985), discusses different alignment patterns from a multi-factorial perspective. While

<sup>4</sup>Thus one could argue that fewer distinctions are made in the marked category because marked categories are less frequent compared to unmarked and are therefore less resistant to analogical leveling leading to syncretism.

he admits that proper encoding of semantic roles is an important factor, he also acknowledges the role of other factors such as economy or distinguishability of arguments. Yet, more recently the alignment schemes have been explicitly reformulated in terms of semantic maps. Thus, Croft (2001:137, 147) proposed the following semantic map for the encoding of core arguments.

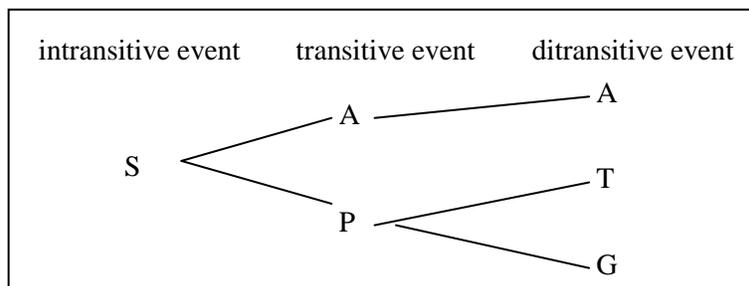


Figure 5: Croft's conceptual space for core arguments (participant roles)

This semantic map conveniently captures the variation of alignment patterns across languages, both in the monotransitive and the ditransitive domain. In the latter domain the distinction is made between languages with a direct vs. indirect object distinction (“indirective alignment”; schematically,  $T=P \neq R$ ), and languages with a primary vs. secondary object distinction (“secundative alignment”; schematically,  $T \neq P=R$ ) (Dryer 1986; Haspelmath 2004; Siewierska 2004). Given standard assumptions about the well-formedness of semantic maps, this approach correctly predicts marginality of alignment types which would display discontinuous segments on the map (such as the anomalous  $S \neq A=O$  pattern). Thus the following two maps represent two of the possible alignment types complying with the semantic map: the accusative “indirective” language such as German (see (1)), and an ergative language with a “secundative” alignment such as Eskimo (see (2)).

German

(2) *Ich gab ihm ein Buch*

‘I (NOM) gave him (DAT) a book’

Eskimo (West Greenlandic; Fortescue 1984:88)

(3) *Uuma Niisi aningaasa-nik tuni-vaa*

that.ERG Niisi money-INSTR.PL give-IND.3s->3s

‘(He) gave Nisi money’

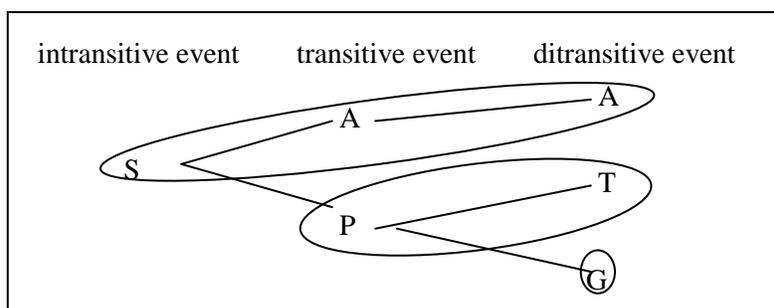


Figure 6: Alignment maps: German

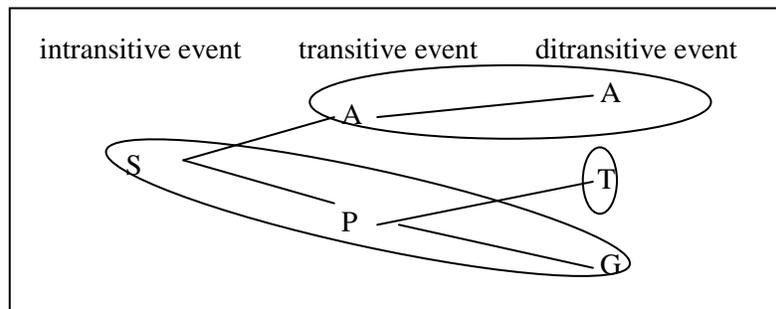


Figure 7: Alignment maps: Eskimo

More similar alignment schemes can be found in Dryer (2007), who, however, does not call them maps.

Yet, the alignment map in Figure 5 is not unproblematic since we find exceptions both in the monotransitive and ditransitive domain. Thus, the anomalous Double-Oblique pattern ( $S \neq A = O$ ), also called “horizontal alignment”, is cross-linguistically rare but is attested in a number of Iranian languages (see (6) from Vafsi below). Another type of violation, this time in the ditransitive domain, is found in Spanish. As is well-known, Spanish displays differential object marking (DOM) in the monotransitive domain—that is, P, if animate, has to be marked by the dative-accusative preposition *a*, but remains unmarked if inanimate. By contrast, the object-theme in the ditransitive domain always remains unmarked (as in (4)); note that marking an animate T by the preposition *a* makes the construction (5) unacceptable:

Spanish (Company 2003:234)

(4) *El maestro presentó ∅ su mujer a sus alumnos*  
 the teacher introduced his wife.ACC to his pupils.DAT  
 ‘The teacher introduced his wife to his pupils’

(5) *?El maestro presentó a su mujer a sus alum[n]o[s]*  
 the teacher introduced to his wife.ACC to his pupils.DAT  
 ‘The teacher introduced his wife to his pupils’

Since the intransitive subject is equally unmarked, the resultant pattern violates contiguity because T shares the same encoding with S, but not with P. How can one account for such a counterexample? One feature of the ill-formed pattern in Spanish is that it involves a zero marking. Now, distribution of zero markers is arguably governed by considerations of economy and distinguishability and need not reflect a semantic affinity. Note that marking an (animate) T with the “dative” preposition results in a pattern where both R and T are introduced by the same preposition (see the unacceptable pattern in (5)) thus violating distinguishability (see Malchukov 2008 for an optimality-theoretic account of the Spanish pattern along these lines). Economy considerations are equally important: arguably, absence of marking is not as strong an indicator of semantic relatedness as overt marking, and thus zero markers incur violations on the map. Another instructive example of the role of economy concerns the distribution of the direct case in languages with a two-term case system, recently discussed by Arkadjev (2005; 2009). Arkadjev observes that in some of these languages (languages of the “distributing” type, in his terms) distribution of the direct case (as opposed to the general oblique case) is problematic for the

semantic map approach. In these languages (e.g., some Pamir languages) the direct case is found, apart from the subject, also on nouns in peripheral functions such as temporal or locative, which are not connected to the subject function on his map. Given that the direct case is typically unmarked, economy provides a ready explanation for such cases. Indeed, in many languages nouns denoting temporal notions or place names remain unmarked, but this does not implicate semantic similarity with the subject; rather it indicates that the respective roles are recoverable in the absence of marking (see Malchukov & de Swart 2009 for further discussion of economy and distinguishability effects in case marking; cf. Aristar 1997). Thus, distribution of zero markers need not reflect a semantic similarity but may be rather a matter of economy/recoverability of respective functions.<sup>5</sup>

#### 4. Distinguishability as an Inhibiting Factor

Viewing alignment schemes as semantic maps may be problematic for another reason as well. In particular, it is not clear whether S is a unitary category in semantic terms or not, and thus it may be misplaced on the semantic map of core (macro)roles. Indeed, in semantic terms, it is more adequate to distinguish between agentive (or, unergative) subjects (Sa) and patientive (or, unaccusative) subjects (Sp). Given this distinction we can arrive at the following configuration A – Sa – Sp – P, which makes more sense semantically. In particular, this configuration allows us to represent different kinds of split intransitive systems which group together A with Sa and/or P with Sp (Dryer 2007; cf. Dixon 1994; Song 2001). A more radical solution is to exclude S from the semantic map altogether since it does not correspond to one role. Note that this exclusion is also justified for another reason as well. As noted above, S is usually unmarked, and thus its distribution will be less informative for the arrangement of functions on the semantic map. The latter approach is taken by Malchukov & Narrog (2009), who propose the following semantic map (conceptual space) for the case roles from the non-local domain.

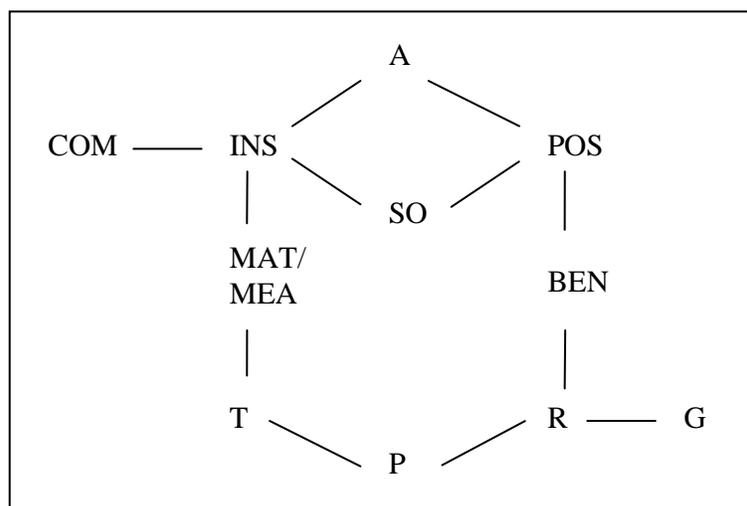


Figure 8: The semantic map (conceptual space) for major case relations (non-spatial domain)

In addition to agent (A), patient (P), theme (T), and recipient (R), as familiar from alignment maps, the map in Figure 8 introduces new case functions, mostly from a non-spatial domain:

<sup>5</sup>See also Wälchli, this issue, for a similar observation.

instrument (cf. *He hit the dog with a stick*), material/means (cf. *He built the house with bricks*), beneficiary (cf. *I built him a house*), possessor (cf. *I saw his house*), source (cf. *He returned from the store*), goal (cf. *He returned to the store*), comitative (cf. *He returned with a friend*). The upper part of the map in Figure 8 linking agents to instruments and causes, on the one hand, and to genitives, on the other, represents different reductions of the agentivity prototype (Grimm 2005). Thus, possessors and datives share sentience property with agents but lack the feature of instigation, while instruments are instigating entities but lack sentience. The lower part of the map is reminiscent of the semantic map for the ditransitive domain in Figure 5, even though the T role is not conceived as restricted to ditransitive themes but also pertains to monotransitive themes (they differ from patients in that they do not necessarily involve a change of state in the course of a verbal event). Note that in contrast to the alignment maps in Figures 1-5, the proposed map does not include a separate S function since S does not correspond to a single (macro)role. The upper and lower parts of the map are linked by two “routes”, one leading from Instrumental to Theme, and another from Possessive to Recipient marking. The former route is mediated by the means/material function (cf. *load the cart with bricks*), which shows similarities to instrument, on the one hand (it may be conceived of as an instigating entity), and to the theme, on the other (standing in an incremental relation to the verb). This accounts for the fact that the means/material function may pattern either with instruments or with themes across languages or indeed in the same language, resulting in a well-known “spray/load alternation” (cf. *load bricks on the cart* vs. *load the cart with bricks*). Finally, one of the conceptual underpinnings of the possessive-benefactive-recipient connection may be the “possession as goal” scheme, as suggested by Heine (1997; see 6.2 for further discussion).

The proposed map has both a semantic plausibility and also finds empirical support in the most frequent polysemy patterns as described in the contributions to Malchukov & Spencer (eds.) (2009) dealing with individual cases: the dative-allative polysemy, as familiar from English (recall the functions of *to*), is common across languages (Creissels 2008; Næss 2009), the dative-genitive polysemy is attested in many Australian and Austronesian languages but is also found elsewhere (Lander 2009; Næss 2009); the genitive-ablative polysemy is especially common in languages using adpositions for these functions (Heine 2009, Lander 2009); the dative-accusative polysemy is familiar from languages with differential object marking (Næss 2009; Malchukov & de Swart 2009); instrumental-accusative polysemy is typical for languages with secundative alignment (Kittilä & Malchukov 2009); the instrumental-comitative polysemy is the most frequent polysemy pattern involving both cases (Narrog 2009, Stolz et al. 2009); finally, the ergative-instrumental and ergative-genitive polysemies are identified as the two most frequent polysemy patterns involving ergative case (Palancar 2009). Thus the proposed map captures the major polysemy patterns of case markers in the non-spatial domain.

In light of the previous discussion, let us now consider the anomalous double-oblique pattern ( $S \neq A = O$ ), which is problematic for the alignment maps. The double-oblique pattern is cross-linguistically rare, yet it is attested in a number of Iranian languages (Payne 1980; Bossong 1985; Stilo 2004; Arkadjev 2005). This pattern is illustrated by an example from Vafsi, where A is marked by the oblique case in the past tense, and (prominent) P is likewise marked by the oblique.

Vafsi (Stilo 2004:244)

- (6) *luas-i*      *kærg-e = s*      *bæ-værdæ.*  
 fox-OBL.SG    chicken-OBL.SG=3SG    PFV-take.PST  
 ‘The fox carried off (the) chicken’

Given that both A and P are marked (by the oblique case) and S is unmarked, this pattern incurs a contiguity violation of the alignment maps in Figures 1-5. Indeed, this pattern seems to be puzzling as it violates all the functional motivations behind case marking—it does not reflect semantic similarity, nor does it distinguish between arguments and is not economical. However, this pattern is conceivable in light of diachronic data. As noted in the literature (Kerimova & Rastorgueva 1975; Arkadjev 2005), the double oblique pattern in these languages results from meaning extensions of the originally polyfunctional dative-genitive case. This syncretic case developed into the marker of a (prominent) object, on the one hand, and to the ergative marker in past tenses, on the other hand. Note that both developments are not unusual. Ergative markers of genitive origin are common in ergative languages (e.g. in Eskimo), while the DAT to ACC shift constitutes a well-known grammaticalization path, familiar from languages with differential object marking (such as Hindi; Bosson 1985; Lehmann 1995). Thus, the double-oblique pattern is due to a polysemy chain, with individual polysemy patterns well-attested elsewhere. Schematically:

If  $ERG = GEN$ ,  $GEN = DAT$ ,  $DAT = ACC$ , then  $ERG = ACC$ , resulting in a  $A=P \neq S$  pattern

Note that although this scenario cannot be represented on the alignment map in Figure 5, it can be easily captured on the semantic map for case functions in Figure 8. Thus, though a map violation in this case is apparent, it disappears once further functions (which are absent on the alignment map in Figure 5 but present in the case function map in Figure 8) are taken into account. This however also provides a (partial) explanation as to why this pattern is less frequent: it involves a “long-distance” polysemy. Note that in relative terms the languages displaying an “extended” polysemy should be rarer than languages displaying “restricted” polysemies contributing to the polysemy chain. Thus, a polysemy chain  $x=y=z$ , will be found in a subset of languages which display restricted polysemies  $x=y$  and  $y=z$ . Thus, while the traditional approach to semantic maps cannot (and is not intended to) capture absolute frequencies of individual polyfunctionality patterns, it can straightforwardly account for relative frequencies in extended vs. restricted polysemy chains.

In addition, there are other functional reasons why this polysemy pattern involving A and P should be infrequent. Note first that semantic similarity cannot account for such polysemy because semantic similarities support polysemies of adjacent (connected) categories on the map, while non-adjacent categories need not share any common semantic features. In our case, similar encoding of A and P does not reveal any semantic commonality between these (macro)roles. More importantly, the resultant pattern fails to distinguish between subjects and objects. Similar effects of distinguishability as an inhibiting factor for extended polysemies can be found elsewhere. Thus, while extension of instrumental case marker to As (resulting in an ergative pattern) and to Ts (resulting in a secundative pattern) is not uncommon, languages which show both polysemies are virtually unattested. Indeed, extension of the instrumental marker to subjects

and objects simultaneously would violate distinguishability. In fact, the few languages which combine ergative alignment for monotransitives with the secundative alignment for ditransitives use different cases for subjects and objects. Thus, Eskimo (see (2)) uses the Instrumental (“modal”) case for T arguments but uses genitive (“relative”) case for As, so no distinguishability violation arises. Instrumental polysemy involving the agent and comitative functions might provide another instructive example. Both polysemy patterns of instrumentals are widely attested (Stolz, 2009, Narrog, 2009), but the combined polysemy involving all three functions is not found (Stolz 1996; Palancar 2009) although it is perfectly compatible with the map and thus predicted to occur. Distinguishability or recoverability of functions is again a likely explanation for this restriction. Indeed, in a language showing such extended polysemy, an animate NP in the “instrumental” case would be regularly ambiguous, allowing both agentive and comitative interpretation. Note that such ambiguity is not found with instruments since the latter are typically inanimate. Similar distinguishability effects can be found in the domain of object encoding as well. Note that while extension of the dative case to P is common, further extension of the same marker to T is disfavored as it leads to ambiguity (recall the discussion of the Spanish examples (4)-(5)). It is to be expected that extended polysemies in other domains would also be disfavored due to other factors such as distinguishability<sup>6</sup>. Thus, invoking other functional factors can not only account for unexpected polysemy patterns, but also explain why certain (extended) patterns permitted by the map are not attested (recall the problem of overgeneration, raised by Cysouw).

## 5. Structural Factors and Pattern Inheritance

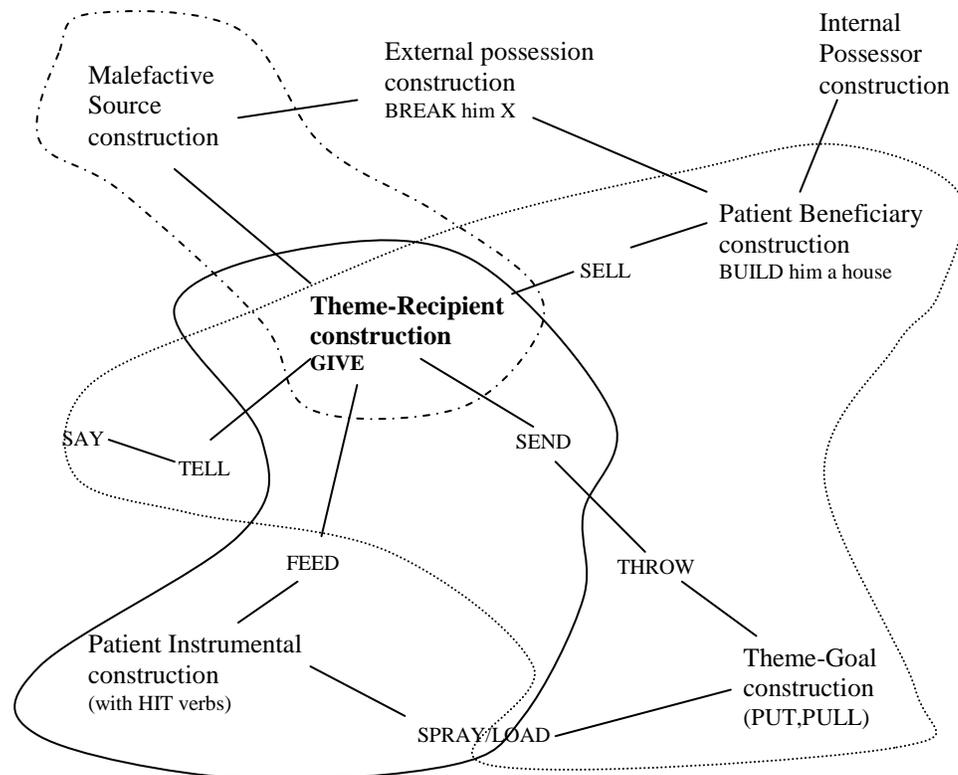
Semantic maps can be produced not only for grammatical markers but also for verbs types of different valency (cf. Wälchli, this issue, for motion verbs), or, more generally, for different constructions (cf. Cristofaro, this issue). Malchukov, Haspelmath, and Comrie (forthcoming, henceforth MHC) propose the following semantic map (conceptual space) for the domain of ditransitive constructions. In addition to the ditransitive recipient theme construction (cf. *I gave him a book*), it includes a number of other three-argument constructions (see Margetts & Austin 2007 for an overview): the malefactive theme construction (cf. *I stole a book from him*), patient beneficiary construction (cf. *I built him a house*), internal possessor construction (cf. *I saw his house*), external possessor construction (cf. Russian: *On mne slomal nogu* lit. ‘He broke me the leg’), theme goal construction (cf. *I put/threw a newspaper on the table*), and patient instrument construction (cf. *I hit the dog with a stick*), among others.

The claim embodied in the semantic map in Figure 9 is that categories adjacent on that map will share a similar construction (valency pattern), and the same map can constrain diachronic extensions of certain constructions. In Figure 9, some of the extensions of the basic ditransitive constructions in three languages (Finnish, Eskimo, and Jamul Tiipay) are represented, in order to illustrate basic alignment types: (i) indirective alignment (cf. the extension of the allative case in

---

<sup>6</sup>The patterns of syncretism in person paradigms can provide another example of the role of distinguishability as an inhibiting factor. While in the person paradigms both “horizontal” syncretism patterns (involving number neutralization) and “vertical” syncretism patterns (involving person neutralization) are widely attested (Siewierska 2004; Cysouw 2003), these patterns are in complementary distribution. Siewierska (2004: 100) attributes complementarity of horizontal and vertical homophonies to the need to maintain a certain level of explicitness within person paradigms.

Finnish), (ii) secundative alignment (cf. the extension of the instrumental case in Eskimo), and (iii) neutral alignment (cf. the domain of the double object construction in Jamul Tiipay):



Jamul Tiipay: neutral (DOC)      - - - - -  
 Finnish: allative extensions      ·······  
 Eskimo: instrumental extensions      ————

Figure 9: Basic ditransitive constructions in Jamul Tiipay, Finnish, and Eskimo

In the present context it is interesting to consider extensions of the double object construction across languages. Kittilä (2006), which is one of the very few studies of lexical variation in ditransitive constructions, concluded that ‘give’ is one of the verbs which is most likely to be found in a double object construction. There are good reasons for this preference. As argued by Kittilä, ‘give’ scores high on Hopper & Thompson’s (1980) transitivity parameters. For example, unlike ‘send’, ‘give’ implies a successful transfer; hence, not only the theme but also the recipient is affected by the action. Furthermore, issues of distinguishability/recoverability seem to be relevant as well (as argued in MHC). Insofar as T is typically inanimate and R animate, the respective roles are easily recoverable and need not be distinctively marked. Note that verbs involving two animate participants like ‘introduce’ are less commonly used in a double object construction cross-linguistically.

Kittilä (2006) notes some counterexamples to the preferential use of ‘give’ in a double object construction. Some of these can be explained in terms of the same functional factors invoked above. Thus, verbs taking a malefactive source, like ‘steal’, arguably outrank ‘give’ verbs in affectedness (of the third participant); therefore, they are likely to be found in the double object construction (see the domain of double object constructions in Jamul Tiipay) and even may be preferably eligible for this construction, compared to ‘give’ verbs (as, for example, in Mandarin

Chinese; see MHC for discussion). Some other exceptions, however, are puzzling and defy a functional explanation. Thus, in Malayalam, ‘give’ takes a dative construction, while “less canonical ditransitives” (Asher & Kumari 1997:205) like ‘entrust’ and ‘feed’ take a double object construction:

Malayalam (Asher & Kumari 1997:205)

- (7) *kuTTi enikkə Peena tannu.*  
 child I.DAT pen give.PST  
 ‘The child gave me the pen.’

- (8) *’naan puuccakkə paal koTuttu*  
 I cow.PL.ACC grass eat.CAUS.PAST  
 ‘I fed the cows grass.’

One would be hard pressed to provide a functional explanation for the eligibility of these particular verbs for the double object construction in Malayalam, yet this can be easily explained in structural terms. As it turns out, both ‘entrust’ and ‘feed’ are causative verbs, and causatives (of transitives) take a double object construction in Malayalam. While cross-linguistic preference of causatives for the double object construction in its turn might need a separate explanation (see MHC for some proposals), it should be admitted that in the case of Malayalam the choice of the (double object) pattern is due to structural factors (morphological structure) and is not directly related to semantic similarities (between ‘feed’ and ‘entrust’ to the exclusion of ‘give’). So while evaluating semantic maps such structural factors should be featured out in order to account for apparent exceptions.

Another interfering factor discussed in MHC concerns verbal polysemy and pattern inheritance. Consider the case of Chechen (Nakh-Daghestanian), where ‘hit’ unexpectedly takes the dative-allative pattern, which it shares with caused motion verbs.

Chechen (Bickel & Nichols 2009)

- (9) *Daas ystaghna urs tyyxi-ra.*  
 father.ERG sheep.DAT knife.NOM strike.PST  
 ‘Father stabbed the sheep with the knife.’

Note that this pattern causes a discontinuity on the map in Figure 9 since the allative strategy is found both with ‘hit’-verbs and ‘throw/put’-verbs, but is not found with the intermediate type of ‘load’-verbs (which frequently align either with the former class or with the latter class, or with both, which leads to a familiar “spray-load alternation”). Now, as is clear from the glosses, the verb in the ‘hit’-construction is also used as a contact verb (‘strike at’) elsewhere. This latter use naturally provides a motivation for the use of the allative-dative pattern, which is frequent with contact and caused motion verbs. Similar cases where the pattern of ‘affect’-verbs (like ‘hit’) is modeled on caused motion verbs are attested elsewhere. For example, in Ewe, the verb *da*, which is used for ‘hit’, is used as a caused motion verb ‘throw’ elsewhere (Essegbey 1999:166). In one

interpretation,<sup>7</sup> this extension of the allative strategy into the instrumental domain is accompanied by a semantic shift (e.g., from ‘strike’/‘throw’ to ‘hit’; see Malchukov 2005 for a general discussion of pattern polysemy and pattern inheritance in case marking). Thus polysemy of individual verbal lexemes can introduce inconsistencies into the semantic map because the case frame would be motivated only by one of the meanings of the verbal lexeme. Such semantic shifts are problematic for semantic maps as they may incur a contiguity violation. In our case, either one needs to recognize a violation or postulate an extra link connecting ‘hit’ and ‘put’ verbs. While this is always possible (but undesirable since additional connections lead to less predictive maps) it also loses the generalization that this connection obtains when the verb itself is polysemous. Again, this violation is due to an interfering factor manifesting the pressure for an analogical syntactic behavior on the part of polysemous items. This factor can be factored out on the map by checking for polysemies of individual verbs.

## 6. Diachronic Factors

In this section we consider some diachronic factors which may cause violations of the contiguity of semantic maps. It should be noted that although the previous discussion has been mostly couched in synchronic terms, it is well known that semantic maps can be used to represent a diachronic dimension as well (for discussion of “dynamicized” semantic maps see Van der Auwera & Plungian 1998; Haspelmath 2003; Narrog, this issue). Viewed diachronically, polysemies arise from meaning extensions of individual categories. In most cases, such meaning extensions are gradual and proceed stepwise along the network of functions on a map without incurring a contiguity violation (Croft et al. 1987). However, there are more complex diachronic scenarios, to be discussed in this section, which might be problematic for semantic maps. We will discuss three such cases: polygrammaticalization of lexical items, reinterpretation through reanalysis, and gram replacement.

### 6.1 Polygrammaticalization effects on semantic maps

Klamer (fc.) discusses grammaticalization of motion verbs in the Papuan languages Teiwa and Kaera—these verbs are in the process of developing into the “oblique” case marker. Thus, in Teiwa, the marker *ma* (cf. Kaera *mi*) is used to mark goals (as in (10)) and instruments (as in (11)).

---

<sup>7</sup>In another interpretation (suggested by B. Bickel, p.c.) such cases still represent caused motion or contact verbs, like ‘strike’, rather than affect verbs, like ‘hit’. In this approach, no contiguity violation arises since the verb in (9) would not qualify as ‘hit’. Yet, it is not clear whether this latter interpretation can be upheld either in this particular case or in general as it seems to exclude the possibility of a semantic shift. In the case of Nakh-Daghestanian languages, for example, it may be relevant that in Chechen the sentence in (9) denotes a successful, not an attempted, stabbing (Z. Molochieva, p.c.). The same is true of the closely related Ingush (J. Nichols, p.c.).

Teiwa (Klamer fc.)

- (10) *Hala ta gi er-an gula' Ma haraba ma gad.*  
 people TOP go do-MOD finish CONJ stable OBL put

‘The people went to do (that) then put (it) in a stable.’

- (11) *Uy nuk ped ma tei taxar*  
 person one machete OBL wood cut

‘Someone cuts wood with a machete.’

Such polysemy is quite unusual typologically and is likely to incur a violation on a semantic map for case functions. Note that the goal and instrument are not contiguous on the map for syntactic case functions in Figure 8 above (or on similar maps of case functions proposed by Haspelmath 2003 and Narrog, this issue). The origin of this marker provides a straightforward explanation for this puzzling polysemy: as shown by Klamer, this marker represents early stages of grammaticalization of the verb *ma* ‘come’. Note that at early stages, grammaticalization is likely to be more context-dependent (Hopper & Traugott 1993:89), and lexical items involved in grammaticalization can develop into different markers in different constructions (a process sometimes referred to as “polygrammaticalization”; Craig 1991). In this particular case, the motion verb has grammaticalized into an oblique marker when combined with inanimates, while in a construction with animates it is rather used as a modal marker or a conjunction (cf. the first use of *ma* in (10) above). It is clearly futile to try to find a common denominator, or even shared components, for all these disparate functions. Thus, results of polygrammaticalization of lexical items are more likely to cover regions inconsistent with the map than those which represent further meaning extensions (“lateral” meaning shifts). One approach to handling these counterexamples would be to exclude from consideration cases where a grammatical item derives from a lexical one. At least in some cases this exclusion seems to be motivated as it is still possible to argue that we are dealing with different contextual uses of a lexical item. In fact, this is a position ultimately adopted by Klamer (fc.), who argues that these disparate functions of *ma* in Teiwa are better conceived of as different contextual uses of the single intransitive motion verb with an unspecified argument position.

Thus, early grammaticalization is likely to incur violations of the established semantic maps. A legitimate question is why we do not encounter such cases more frequently, which might eventually undermine the semantic map enterprise. The reason is that polygrammaticalization of a lexical item is construction sensitive and requires a number of preconditions. Thus, for Kaera and Teiwa, Klamer notes the following factors which allowed polygrammaticalization: productiveness of verb serialization; non-distinction of finite and non-finite verbs; lack of three-valent verbs. Given that such preconditions are likely to be areally/genetically restricted, polygrammaticalization effects would not produce a consistent pattern across languages.

## 6.2 Diachronic factors: reanalysis

In a paper discussing the connections between benefactive and possessive domain, Daniel & Malchukov (forthcoming) propose the following map (simplified here), which can be seen as an elaboration of the map in Figure 8 from Malchukov & Narrog (2009) (cf. also Haspelmath 2003; Malchukov, Haspelmath & Comrie (forthcoming); Narrog, this issue for maps for the dative domain).

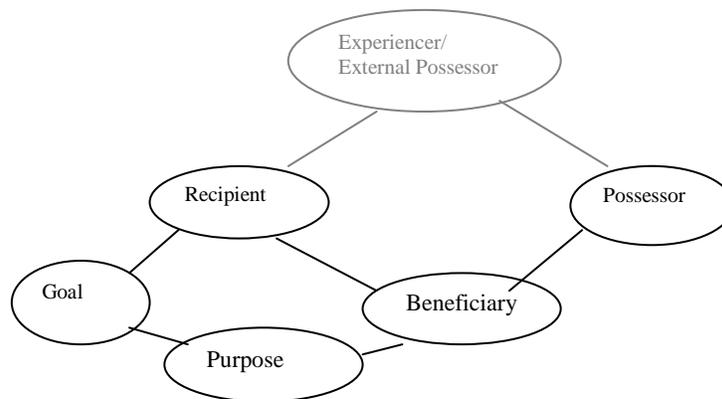


Figure 10: The possessive-benefactive connection

This map “zooms in” on the possessive-benefactive domain of case relations—additional functions, which do not appear on the map in Figure 8, include purpose (cf. *We went out for dinner*) and external possessor/experiencer (as in Figure 9); the latter functions will not concern us here. In the present context it is important that goal and possessor are not adjacent on that map; rather, the connections are mediated through the intermediate functions of recipient and beneficiary (the other route involving external possession is ignored here). The extension from goal marking through the recipient (or purpose) function to beneficiary and further into the domain of attributive possession is not exceptional across languages. Heine (1997) regards it as one of the major encoding patterns (“schemas”) for encoding possessors—the so-called goal schema (Y exists for/to X > X has/owns Y). The goal schema is more frequent in the domain of predicative possession (cf. French: *Ce chien est à moi* ‘This dog is mine (lit. to me)’), but is also occasionally extended to attributive possession as well (Heine 1997).

It is important to keep in mind that the term “goal schema” is potentially misleading since Heine’s goal, comprises recipients and beneficiaries as well. In accordance with the semantic map, partial polysemies between adjacent categories should be more widely attested. Indeed, we find many cases where the same marking is used for goal and recipient (cf. English *to*), or for recipient and beneficiary (e.g. Russian dative), or for beneficiary and possessor (e.g. *-paj* in Imbabura Quechua). Occasionally, we find “longer” polysemy chains, as when the allative marker in Finnish extends to recipients and beneficiaries (cf. Figure 9 for an extension of the allative strategy with different verbs types) or when possessive classifiers in Austronesian are extended to beneficiaries and recipients (Song 2005). In (Classical) Persian the purposive postposition *râ* (< *râdiy* ‘for the sake of’) spread to beneficiaries and possessors, on the one hand, and to recipients and further to patients, on the other hand (Bossong 1985:60 ff.; Hopper & Traugott 1993:158). Importantly, all these configurations are compatible with the semantic map

in Figure 10 (see Narrog, this issue, for more examples<sup>8</sup>). Occasionally, we find still broader polysemies which include goal, recipient, beneficiary, and possessor. One example is Budukh, a Daghestanian language where the “alienable genitive” is used to mark recipients and goals.

Budukh (Authier, fc.)

(12) *z-o*            *k'ant*  
 me.obl-GEN    knife  
 ‘my knife’

(13) *z-o*            *k'ant*            *yvaci*  
 me.obl-GEN    knife(NOM)    gave  
 ‘He gave me the knife (temporarily)’

Authier (fc.) argues that the “alienable genitive” originally had a locative function (better preserved in the genealogically related Kryz), and its extension to the adnominal domain followed from a reinterpretation of ditransitive constructions. The important point here is that semantic evolution/extension is gradual: allative extends first to dative function, then dative extends to genitive. As can be readily seen, these extensions are in accordance with the map in Figure 10.

In Khwarshi (another Daghestanian language) a seemingly similar pattern is attested (Khalilova, 2009). Also in Khwarshi the same case in *-lo* is found both on possessors and goals, yet the situation is more complex in two respects. First, the use of this case in the goal function is restricted to some contact verbs (like ‘touch’, ‘strike’), and is used with animate goals exclusively (with inanimate goals another case, concessive, is used). Second, the use of *-lo* in the genitive function is also restricted insofar as it is used to modify oblique rather than absolutive arguments (Khwarshi, as other Daghestanian languages, is ergative). A third puzzling point is that in Khwarshi, the same case is found on (some) goals and (some) possessors, but neither of the intermediate functions (recipient, beneficiary) seems to be involved. How does this polysemy come about? This is clearly not an extension of the allative marker, as proposed for Budukh. Indeed, its goal use is very restricted, so the possessive use is more likely to be the source function. One obvious explanation which accounts for peculiarities of this construction is that the “oblique” genitive is used in the allative function due to the omission of the head noun referring to a body-part (cf. *laga-qa* body-CONT in (15)).

Khwarshi (Zaira Khalilova and Raisat Karimova, p.c.)

(14) *De*    *ite-lo*            *kraska*    *l-ex-i*.  
 I.ERG    3SG.OBL-GEN2    paint    IV-touch-PST:W  
 ‘I colored her (with the paint).’

<sup>8</sup>This map is largely compatible with a diachronic map for dative-related functions in Narrog (this issue) as far as the same categories are concerned, but it is more restricted in two respects: it does not postulate either a direct recipient-purpose connection (which is not supported by the data) or a direct goal-benefactive connection (which seems to be mediated by the recipient function in languages cited as supporting evidence: English, Greek, Ik).

- (15) *De ite-lo k'ak'a-qa kraska l-ex-i.*  
 I.ERG 3SG.OBL-GEN2 leg-CONT paint IV-touch-PST:W  
 'I colored her leg (with the paint).'

Thus, the genitive-goal construction as in (14) developed from the construction in (15) where the goal is in the concessive case and the possessor predictably takes the oblique genitive (as it modifies an oblique argument). A similar explanation has been proposed by Comrie et al. (2007) for Tsez where the oblique genitive can also encode goals under similar conditions.

Thus, unlike what we observed in Budukh, the goal-possessor polysemy in Khwarshi is not due to a meaning extension but rather results from reanalysis. In such cases polyfunctionality need not comply with the map. Yet as in the case of polygrammaticalization discussed above, such changes are likely to be less common since they presuppose certain conditions. In the case of Daghestanian languages (Khwarshi and Tsez), availability of a special genitive case used for oblique arguments arguably made it easier for the body part to be omitted because its meaning is still recoverable. Note also that not all cases of reanalysis are problematic for semantic maps. Thus, reanalysis of the recipient-beneficiary to possessor involving “internalization” of beneficiaries into an NP (as posited for Budukh by Authier) or reanalysis in the opposite direction involving “externalization” of possessors (as posited for Austronesian by Song 2005), do not violate the map.

### 6.3 Semantic map violations due to gram replacements

Sadanobu and Malchukov (forthcoming) discuss the evolution of the aspecto-temporal forms in Japanese, in particular, the perfect-continuous form in *teiru* and the past form in *ta*. One puzzling point that has attracted much attention in the literature (see Sadanobu and Malchukov (forthcoming) for an overview) is that *ta* can perform two seemingly opposite functions: apart from the general past meaning as in (16), it can also be used for emphatic (or “mirative”) present, as in (17):

Japanese

- (16) *Kinou ame-ga sanjikan fu-tta.*  
 yesterday rain-NOM three\_hours fall-TA  
 'Yesterday it rained for three hours.'
- (17) *A, hora, anna tokoro-ni inoshishi-ga i-ta-yo!*  
 oh look such place-LOC wild\_hog-NOM exist-TA-AFF  
 'Oh, look, there is a wild boar over there!'  
 [uttered at the sudden sight of a wild hog during an excursion in the mountains]

While the polysemy confined to past tense and present mirative meanings is virtually unattested across languages, it is conceivable in light of the diachronic data. Historically *ta* derives from the perfect-resultative *tari*, which was subsequently reanalyzed as a general past form. Yet, it has retained the residual mirative functions related to its original resultative function. Note that although *ta* is related to perfect historically, synchronically, it does not qualify as a perfect marker. As shown in (16) it can freely combine with time adverbials (a context more

characteristic of imperfects cross-linguistically) and is furthermore banned from some contexts characteristic of perfects cross-linguistically<sup>9</sup>. The perfect function is currently performed by *teiru* (which also has continuous uses elsewhere). Thus the puzzling polysemy of *ta* in Modern Japanese can be attributed to the fact that the perfect meaning has been partially “taken-over” by *teiru*, which is a “young perfect” replacing the “old perfect” in *ta* in its core function. This scenario is depicted in the Figures 11 and 12 below representing the hypothesized semantic evolution of the *ta* form.

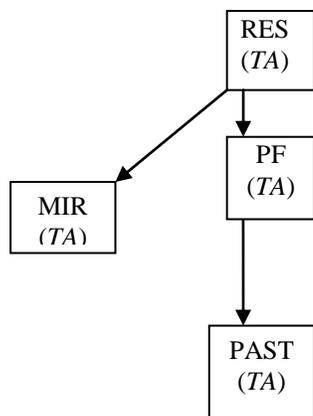


Figure 11: Perfect-resultative meaning of *ta* and its extensions

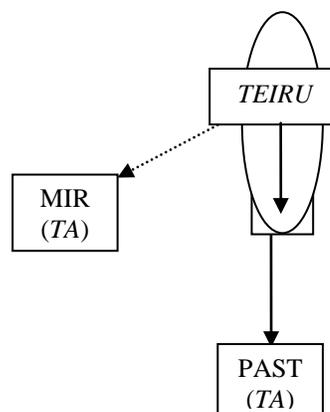


Figure 12: *Ta* displaced in the central functions by *teiru*

These figures may be viewed as a representation of a grammaticalization path, as familiar from the grammaticalization literature (see Bybee et al. 1994 for grammaticalization of perfects-antérieurs) but also as a semantic map incorporating a diachronic dimension. Note that the map in Figure 12 is ill-formed since the domain of *ta*-marking is no longer contiguous on the map. This illustrates a well known case of contiguity violation conditioned by gram replacement. Similar examples have been cited in the literature for different grammatical domains, including voice (Croft et al. 1987), mood (Van der Auwera & Plungian 1998), tense (Haspelmath 2003), and clause combining (Malchukov 2004).

As noted in the literature, such processes of gram replacement present a challenge for semantic maps because, viewed synchronically, they violate contiguity of semantic maps (see also Zwarts, this issue). In my view, such exceptions can be dealt with on the assumption that the new grams replacing the old ones are structurally “heavier”, that is are morphologically more complex. This is obviously true for Japanese but seems to hold for most other cases discussed in the literature (e.g. renewal of reflexive markers in the middle domain discussed in Croft et al. 1987, renewal of the present case in Turkish, discussed by Haspelmath 2003:236, and renewal of the concessive conditional function in Russian, discussed in Malchukov 2004). Thus even in the absence of historical evidence, we can account for such exceptions by a simple metric comparing structural complexity of the “offending” category to the adjacent discontinuous grams (the discussion of complexity metrics is discussed in another connection by Mauri, this issue).

<sup>9</sup>Thus, unlike perfect forms in other languages *ta* is not compatible with the adverb *mada* ‘yet’ (Ogihara 1999: 332).

Again a question arises as to why such violations are not more common cross-linguistically given that grammaticalization is a pervasive process, and grams are often driven through the same grammaticalization paths. Yet it is important to keep in mind that such renewal processes do not necessarily lead to a contiguity violation. First, such violations can occur only if a grammaticalization path is “branching” in a certain domain. Second, at early stages, extension of a new gram usually yields an overlap rather than total replacement. On the other hand, at a later stage of grammaticalization, the “new” gram will replace the “old” one in the extended functions as well so that no contiguity violations arise.

Thus, in general, semantic maps are better suited to represent the most common scenarios involving meaning extensions of particular grams, while more complex scenarios, involving gram interaction are more likely to yield contiguity violations. Yet, as has also been observed with regard to polygrammaticalization and reanalysis, the latter scenarios are more complex, because they require more preconditions. Therefore such violations are predicted to be rare and can be featured out on a large scale semantic map.

## 7. Conclusions

In this paper I have argued that semantic similarity is not the only factor which motivates polysemy patterns cross-linguistically; I have also shown that these other factors (markedness, distinguishability, etc.) might give rise to polysemies problematic for the established semantic maps (i.e. induce contiguity violations). It should be noted (as pointed out by S. Cristofaro, p.c.) that it is not the case that these other factors always yield contiguity violations on the established semantic maps. Also, I am not claiming that patterns originating from these other factors should always contradict a particular semantic map. In some cases a contiguity violation can be remedied by introducing further functions and/or connections (cf. the discussion of the agent-patient polysemy pattern found in some Iranian languages in Section 4). Yet, although the correlation is not absolute, it is certainly true that ill-formed (discontinuous) patterns on semantic maps are most often determined by factors other than semantic affinity. Note that an extended polysemy in itself challenges the general assumptions behind the semantic map approach insofar as (non-adjacent) categories involved in extended polysemies need not share a common semantic component (recall again the discussion of the double oblique pattern in Iranian languages).<sup>10</sup> On the other hand, although the effects of other factors need not always be problematic for an established semantic map, they do not consistently support the configuration established on the basis of semantic affinities either.

Thus, semantic affinity is one of the factors which conditions cross-linguistically recurrent polysemies, but is not the only factor. The other factors which may give rise to identical encoding include markedness, economy, distinguishability, as well as structural and diachronic factors. Yet, compared to semantic affinity, the effects of other factors are less pervasive and less consistent for a number of reasons. Thus, markedness effects, by definition, are confined to neutralization in the marked category, while economy considerations regulate distribution of the zero-marked grams. The role of distinguishability as a factor contributing to polyfunctionality is also limited to specific contexts (recall the discussion of Spanish ditransitives in Section 3), while elsewhere it rather favors different encoding of categories which might be otherwise

---

<sup>10</sup>Joost Zwarts (this issue) suggests that common semantic features may be lacking only in polysemy chains on semantic maps with “cycles”. In my view, however, any (non-adjacent) categories involved in a polysemy chain may lack common semantic components, irrespective of the configuration of the semantic map.

confused (recall the inhibiting function of distinguishability discussed in Section 4). The effects of structural factors and polysemy patterns, as observed in individual languages, are not totally random, yet they do not easily generalize cross-linguistically and therefore do not yield a consistent pattern. Finally, violations of the semantic maps can occur under certain diachronic scenarios (polygrammaticalization of a lexical item, reanalysis, gram replacement), but these scenarios are more complex—they require certain preconditions and are therefore expected to be more rare as compared to the gradual meaning extensions consistent with the map.

Thus, if we conceive of a semantic map approach as a method for uncovering semantic relationships between categories, these interfering factors should be considered as “noise” which should be featured out before the semantic map can feed the semantic analysis. Below we have outlined certain heuristics as to how this can be achieved:

- Consider polysemy patterns of the unmarked member to check for effects of “inflectional markedness”;
- Semantic maps should be restricted to cases of overt marking; distribution of zero-markers may be driven by economy/recoverability rather than by similarity;
- Check (code) for structural factors to reduce “noise”;
- Check for polysemy patterns of verbs because a similar syntactic pattern can be due to analogical extension (pattern inheritance);
- Exclude items having both grammatical and lexical functions to check for polygrammaticalization effects which can lead to discontinuities;
- Check for “heaviness” (structural complexity) of “offending” markers, compared to the adjacent discontinuous grams to check for possibility of gram replacement.

Once the effects of interfering factors are featured out, a *similarity map* (as obtained through multidimensional scaling, for example) reduces to a *semantic map*, representing a semantic residue of the similarity map. At the next stage, the semantic map can feed the semantic analysis, aiming at identifying common semantic components of (adjacent) related functions. On the other hand, minority patterns introducing “noise” should be also analyzed in their own right as they can provide important insight into the different factors contributing to polyfunctionality. To conclude: further advances in the semantic map approach crucially depend not only on developments of new representation (visualization) techniques but also on advances in the analysis of the established polyfunctionality patterns (see also Zwarts, this issue, Cristofaro, this issue, for related proposals).

## References

- Aikhenvald, A.Y. and R.M.W. Dixon. 1998. Dependencies between grammatical systems. *Language* 74.56-80.
- Aristar, A.R. 1997. Marking and hierarchy types, and the grammaticalization of case-markers. *Studies in Language* 21.313-68.
- Arkadjev, P. 2005. Dvuxpadezhnye sistemy v indoiranskix jazykax: tipologicheskaja perspektiva (Two-term case systems in Iranian languages in a typological perspective). *Indoiranskije jazyki i tipologija jazykovyx situacij*, ed. by N. Kazanskij. St. Petersburg.
- 2009. Poor (two-term) case systems: limits of neutralization. In Malchukov & Spencer (eds.) *Handbook of Case*: OUP, 686-700.

- Asher, R.E. and T.C. Kumari. 1997. Malayalam. London: Routledge.
- Authier, G. to appear. *Eléments de la langue kryz, dialecte d'Alik*. Paris: Société de Linguistique de Paris.
- Baerman, Matthew. 2009. Case syncretism. In Malchukov & Spencer (eds.), 219-231
- Baerman, M., D. Brown and G. Corbett. 2005. The syntax-morphology interface: A study of syncretism. Cambridge: Cambridge University Press. (Cambridge Studies in Linguistics 109).
- Bickel, B. and J. Nichols. 2009. Case marking and alignment. Malchukov and Spencer (eds.), 304-322.
- Bossong, G. 1985. *Differenzielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Narr.
- Bybee, J., R.D. Perkins and W. Pagliuca. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.
- Company-Company, C. 2003. Transitivity and grammaticalization of object: The struggle of direct and indirect objects in Spanish. *Romance objects: Transitivity in Romance languages*, ed. by G. Fiorentino, 217-261. Berlin: Mouton.
- Comrie B., M. Polinsky and R. Rajabov. 2007. Tsezian Languages. Unpublished manuscript, Max Planck Institute for Evolutionary Anthropology.
- Craig, C. 1991. Ways to go in Rama: A case study in polygrammaticalization. *Approaches to grammaticalization*, ed. by E.C. Traugott and B. Heine, vol. 2, 455-493. Amsterdam: Benjamins.
- Creissels, D. 2009. Spatial cases. Malchukov and Spencer (eds.), 609-626.
- Croft, W., H. Shyldkrot and S. Kemmer. 1987. Diachronic semantic processes in the middle voice. *Papers from the 7th international conference on historical linguistics*, ed. by A.G. Ramat, O. Carruba and G. Bernini, 179-192. Amsterdam: Benjamins
- Croft, W. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.
- Cysouw, M. 2003. *The paradigmatic structure of person marking*. Oxford: Oxford University Press
- , 2007. Building semantic maps: The case of person marking. *New challenges in typology: Broadening the horizons and redefining the foundations*, ed. by M. Miestamo and B. Wälchli, 225-247. Berlin: de Gruyter.
- Daniel, M. & A. Malchukov. in preparation. Possessive encoding of beneficiaries: A cross-linguistic study.
- Dixon, R.M.W. 1994. *Ergativity*. Cambridge: Cambridge University Press.
- Dryer, M. 1986. Primary objects, secondary objects, and antitativity. *Language* 62.808-845.
- , 2007. Clause types. *Language typology and syntactic description*, 2nd edition, ed. by T. Shopen, vol. 1, 224-275. Cambridge: Cambridge University Press.
- Fortescue, M. D. 1984. *West Greenlandic*. London. Croom Helm.
- Grimm, S. 2005. *The lattice of case and agentivity*. MA Thesis. University of Amsterdam.
- Haspelmath, M. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. *The new psychology of language*, ed. by M. Tomasello, vol. 2, 211-243. Mahwah, NJ: Erlbaum.
- , 2004. Explaining the ditransitive person-role constraint: A usage-based account. *Constructions 2*. Available online at <http://www.constructions-online.de/articles>
- , M. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42/1.25-70

- Heine, B. 1997. *Possession: Cognitive sources, forces and grammaticalization*. Cambridge: Cambridge University Press.
- 2009. Grammaticalization of cases. Malchukov and Spencer (eds.), 458-470.
- Hopper, P.J. and E.C. Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Kerimova, A. A. and V.S. Rastorgueva. 1975. Kategorija padezha [The category of case]. *Opyt istoriko-tipologicheskogo issledovanija iranskix jazykov [A Diachronic-Typological Study of the Iranian Languages]*, ed. by V.S. Rastorgueva, vol. 2, 117-199. Moscow: Nauka.
- Kibrik, A.E. 1985. Toward a typology of ergativity. *Grammar inside and outside the clause*, ed. by J. Nichols and A. Woodbury, 268-324. Cambridge: Cambridge University Press.
- Khalilova, Z. 2009. *Grammar of Khwarshi*. PhD dissertation.
- Kittilä, S. 2006. The anomaly of the verb ‘give’ explained by its high (formal and semantic) transitivity. *Linguistics* 44/3.569-612.
- Kittilä, S and A. Malchukov. 2009. Varieties of accusative. Malchukov and Spencer (eds.), 549-562.
- Klamer, M. forthcoming. One item, many faces: ‘Come’ in Teiwa and Kaera. *Typological and Areal Analysis: Contributions from East Nusantara*, ed. by M. Ewing and M. Klamer. Leiden: KITLV Press.
- Lander, Y. 2009. Varieties of genitive. Malchukov and Spencer (eds.), 581-593.
- Lehmann, C. 1995 [1982]. *Thoughts on grammaticalization*. München: Lincom.
- Malchukov, A. L. 2004. Towards a semantic typology of adversative and contrast marking. *Journal of Semantics* 21/2.177-198.
- 2005. Case pattern splits, verb types, and construction competition. *Competition and variation in natural languages: The case for case*, ed. by M. Amberber and H. de Hoop, 73-117. Oxford: Elsevier.
- 2008. Animacy and asymmetries in differential case marking. *Lingua* 118.203-221.
- Malchukov, A. and P. de Swart. 2009. Differential case marking and actancy variation. Malchukov and Spencer (eds.), 339-356.
- Malchukov, A. and H. Narrog. 2009. Case polysemy. Malchukov & Spencer (eds.), 518-535.
- Malchukov, A., M. Haspelmath and B. Comrie. forthcoming. Ditransitive constructions: A typological overview. *Studies in ditransitive constructions*, ed. by A. Malchukov, M. Haspelmath and B. Comrie. Berlin: Mouton.
- Malchukov, A. and A. Spencer (eds.). 2009. *The Oxford handbook of case*. Oxford: Oxford University Press.
- Margetts, A. and P. Austin. 2007. Three-participant events in the languages of the world: Towards a crosslinguistic typology. *Linguistics* 45/3.393-451.
- Næss, Å. 2009. Varieties of dative. Malchukov and Spencer (eds.), 572-581.
- Narrog, H. 2009. Varieties of instrumental. Malchukov and Spencer (eds.), 593-601.
- Narrog, H. and S. Ito. 2007. Reconstructing semantic maps: The comitative-instrumental area. *Sprachtypologie und Universalienforschung* 60/4.273-292.
- Ogihara, T. 1999. The semantics of tense and aspect in Japanese. *The Handbook of Japanese Linguistics*, ed. by Natsuko Tsujimura, 326-348. Oxford: Blackwell.
- Palancar, E. L. 2002. The origin of Agent markers. Berlin: Akademie-Verlag. (*Studia Typologica* 5).
- 2009. Varieties of ergative. Malchukov and Spencer (eds.), 562-572.
- Payne, J.R. 1980. The decay of ergativity in Pamir languages. *Lingua* 51.147-186.

- Plank, F. 1985. The extended accusative/restricted nominative in perspective. *Relational typology*, ed. by F. Plank, 269-311. Berlin: Mouton.
- Rice, S. and K. Kabata. 2007. Crosslinguistic grammaticalization patterns of the allative. *Linguistic Typology* 11.451-514
- Sadanobu, T. and A. Malchukov. forthcoming. Evidential extensions of aspecto-temporal forms in Japanese from a typological perspective. *Cahiers Chronos (Proceedings of the Chronos conference, Antwerp, 2006)*.
- Siewierska, A. 2004. *Person*. Cambridge: Cambridge University Press.
- Song, J.J. 2001. *Linguistic typology: Morphology and syntax*. London: Longman.
- . 2005. Grammaticalization and structural scope increase: Possessive-classifier-based benefactive marking in Oceanic languages. *Linguistics* 43/4.795-838.
- Stilo, D. 2004. *Vafsi folk tales*. Wiesbaden: Reichert.
- Stolz, T. 1996. Some instruments are really good companions: On syncretism and the typology of instrumentals and comitatives. *Theoretical Linguistics* 23/1-2.113-200.
- Stolz, T., C. Stroh and A. Urdze. 2009. Varieties of comitative. Malchukov and Spencer (eds.), 601-609.
- van der Auwera, J. and V.A. Plungian. 1998. Modality's semantic map. *Linguistic Typology* 2/1.125-139.
- Whaley, L. 1997. *Introduction to typology: The unity and diversity of language*. Thousand Oaks, CA: Sage Publications.
- Wälchli, B. 2006/7. Constructing semantic maps from parallel text data. Unpublished manuscript.

Author's contact information:

Andrej L. Malchukov

Department of Linguistics

Max Planck Institute for Evolutionary Anthropology

Deutscher Platz 6

04103 Leipzig

[andrej\\_malchukov@eva.mpg.de](mailto:andrej_malchukov@eva.mpg.de)