A Reinterpretation of Syntactic Alignment

Henk Zeevat
University of Amsterdam
henk.zeevat@hum.uva.nl

Gerhard Jäger
ZAS Berlin
jaeger@zas.gwz-berlin.de

1 Introduction: Harmonic Alignment in Syntax

Harmonic Alignment was proposed by Prince and Smolensky (1993) as a mechanism to establish a correspondence between different harmony scales within the overall framework of Optimality Theory ("OT" henceforth). They specifically address the combination of the phonological sonority hierarchy with the hierarchy of syllable positions. In recent work, Judith Aissen has taken up this idea as a mean to formulate insights from the functionally oriented markedness theory in morphology and syntax within OT syntax (cf. Aissen 1999, 2000). Though based on earlier work in typology like Silverstein (1981) , Aissen manages a formalization of a mechanism that promises an account of much that seems quaint and bizarre about natural languages when considered from the perspective of e.g. a designer of computer languages or logical formalisms.

Suppose a linguistic item can be classified according to two features, A and B. Suppose furthermore that A has two possible values, $A1$ and $A2$, while B has n possible values, $B1 \ldots Bn$, for some $n \geq 2$. Finally, the values of each both features are ranked according to their prominence. Lets say that $A1$ is more prominent than $A2$, and Bi is more prominent than Bj iff $i < j$. Formally, we thus have the prominence scales

(1) $A1 > A2$

and

(2) $B1 > B2 > \cdots > Bn$

Harmonic alignment means that these scales induce a partial ordering on combinations of these features. A combination of a prominent A with a prominent B is harmonic, and so is a combination of a non-prominent A with a non-prominent B. Combinations of a prominent A with a non-prominent B or vice versa are non-harmonic. More precise, the two prominence hierarchies induce the following harmony sub-hierarchies:

(3) a. $A1/B1 \succ A1/B2 \succ \cdots \succ A1/Bn$
b. $A1/Bn \succ A1/Bn-1 \succ \cdots \succ A1/B1$

Aissen uses this mechanism to align formal markedness hierarchies (esp. the hierarchy of grammatical roles) with substantive markedness hierarchies like the definiteness hierarchy or the person hierarchy. The fundamental observation pertaining to harmonic alignment in syntax is that a considerable variety of regularities across languages can be expressed by making reference just to some upper part of the harmony partial order. We give a few examples for the purpose of illustration; the interested reader is referred to Aissen's papers for more comprehensive discussion.

Differential Object Marking Many languages with overt case marking mark some objects, but not others. Bossong (1985) calls this phenomenon "Differential Object Marking" (DOM). According to Aissen (2000), DOM always applies to the top section of a markedness hierarchy that is obtained by multiplying the scale of grammatical functions with some substantive scale like definiteness. Object marking may be optional for this top section and obligatory for the bottom section, it may be prohibited at the top and optional at the bottom, or it is obligatory at the bottom and excluded at the top. Language particular forms of DOM furthermore differ insofar as different substantive scales may be used, and the split may occur at different positions. Let us consider some examples. The scale of grammatical functions and the definiteness hierarchy are given in (4a,b); harmonic alignment leads to the harmony scales in (c) and (d).

(4) a. Subj > Obj
b. pronoun > names > definite > specific indefinite > non-specific indefinite
c. Subj/pronoun > Subj/name > Subj/def > Subj/spec > Subj/non-spec
d. Obj/non-spec > Obj/spec > Obj/def > Obj/name > Obj/pronoun

Any split of the hierarchy in (4d) is attested in instances of DOM in certain languages.[1] Catalan, for instance, obligatorily marks object pronouns with *a*, while full NP objects are unmarked. In Pitjantjatjara (an Australian language), pronouns and proper nouns are case marked when they are objects while other NPs aren't. Hebrew marks only definite objects, and Turkish only specific ones. As borderline cases, one might add languages without any case marking like Kalkatungu (Pama-Nyungan) and languages with obligatory case marking like written Japanese, which select improper segments of the harmony hierarchy.

Similar observations can be made with regard to the animacy hierarchy and with regard to the Cartesian product of these two hierarchies.

Split ergativity The person specification of NPs induces another hierarchy. Simplifying somewhat, it says that the local persons (1st and 2nd) outrank 3rd person. Harmonic alignment thus yields the sub-hierarchies in (5).

(5) a. Subj/local $\succ$ Subj/3rd
b. Obj/3rd $\succ$ Obj/local

These patterns underly split ergative case marking in languages like Dyirbal where the choice between the nominative/accusative system and the ergative/absolutive system is based on person. The table in figure 1 (which is taken from Aissen 1999) shows the basic case marking pattern for Dyirbal.

	Unmarked	Marked
Local persons	Subject	Object
3rd person	Object	Subject (of transitive)
Case	Nominative/Absolutive	Accusative/Ergative

Figure 1: Case marking system of Dyirbal

Briefly put, Dyirbal only marks non-harmonic arguments, i.e. local objects and 3rd person subjects. It thus represents a combination of DOM with Differential Subject Marking.

[1] See Aissen (2000) for examples and references.

2 OT Formalization

Prince and Smolensky (1993) develop a simple trick to translate harmony scales into OT constraints: for each element x of a scale we have a constraint $*x$ ("Avoid x!"), and the ranking of these constraints is just the reversal of the harmony scale. For the person/grammatical function interaction discussed above, this looks schematically as follows (adapted from Bresnan *et al.* 2001):

(6)

Prominence scales	Harmonically aligned scales	OT constraint sub-hierarchies
Subj > Obj	Subj/local $\succ$ Subj/3rd	*Subj/3rd $\gg$ *Subj/local
local > 3rd	Obj/3rd $\succ$ Obj/local	*Obj/local $\gg$ *Obj/3rd

The idea is that the constraint rankings in the third column represent universal sub-hierarchies which are to be respected by any language particular total constraint ranking.

Bresnan *et al.* (2001) present an interesting application of these constraint sub-hierarchies pertaining to person/voice interaction in Lummi, a Salish language spoken in British Columbia. There passivization is obligatory iff the agent of a two-place relation is expressed by third person and the patient by a local person. To express the proposition *The man knows me*, only the Lummi counterpart of (7b) is possible, (7a) is excluded:

(7) a. *The man knows me
  b. I am known by the man

The alignment sub-hierarchy **Subj/Pat* $\gg$ **Subj/Ag*—which arises from harmonically aligning *Subj* > *Obj* with *Agent* > *Patient*—universally favors the active over the passive. On the other hand, the sub-hierarchy **Subj/3rd* $\gg$ **Subj/local* disfavors third person subjects. Languages differ as to how they resolve possible conflicts between these preferences. Lummi is characterized by the ranking **Subj/3rd* $\gg$ **Subj/Pat* $\gg$ **Subj/Ag*. This favors (7b) over (7a) and thus accounts for this grammaticality pattern. English, in comparison, ranks **Subj/3rd* lower than **Subj/Pat* and thus displays no categorical person/voice interaction of this kind. (Instead constraints referring to discourse features like topicality play a role that enforce passive under certain conditions.)

The applications of harmonic alignment that were discussed in the previous section are not covered yet by this OT treatment. Dyirbal, for instance, does not prohibit third person subjects, but it makes marking of those subjects obligatory. Generally, the common pattern of the examples is that non-harmonic combinations must be morphologically marked and harmonic combinations are unmarked. To formalize this idea in OT, Aissen employs a formal operation called "constraint conjunction" which she attributes to Paul Smolensky. If C_1 and C_2 are constraints, $C_1 \& C_2$ is another constraint which is violated iff both C_1 and C_2 are violated. Crucially, $C_1 \& C_2$ may outrank other constraints C_i that in turn outrank both C_1 and C_2. So the following constraint ranking is possible:

$$C_1 \& C_2 \gg C_3 \gg C_4 \gg C_1 \gg C_5 \gg C_2$$

Furthermore, two general constraints play a role:

- "*$\emptyset$" is violated if a morphological feature is not marked
- "*STRUC" is violated by any morphological marking

Each constraint resulting from harmonic alignment is conjoined with $*\emptyset$, and the ranking of the conjoined constraints is isomorphic to the ranking induced by alignment. (Also the conjoined constraints outrank each of their conjuncts.) The alignment of the person hierarchy with the scale of grammatical functions thus for instance leads to the following universal constraint sub-hierarchies:

(8) $*\emptyset$ & *Subj/3rd $\gg$ $*\emptyset$ & *Subj/local
$*\emptyset$ & *Obj/local $\gg$ $*\emptyset$ & *Obj/3rd

Interpolating the constraint *STRUC at any point in any linearization of these sub-hierarchies leads to a pattern where morphological marking indicates non-harmony. The choice of the threshold for morphological marking depends on the relative position of *STRUC. The Dyirbal pattern, for instance, corresponds to the following constraint ranking.

(9) $*\emptyset$ & *Subj/3rd $\gg$ $*\emptyset$ & *Obj/local $\gg$ *STRUC $\gg$ $*\emptyset$ & *Subj/local $\gg$ $*\emptyset$ & *Obj/3rd

3 Some problems

The basic idea of harmonic alignment is conceptually attractive, and it explains a variety of typological generalizations in an elegant way. It is also quite natural to employ OT to formalize the cross-linguistic parameterization of the relevant harmony hierarchies. Nevertheless we find some aspects of the particular OT implementation that Aissen uses conceptually not fully satisfactory. In this section we will point out some issues that strike us problematic. The remainder of the paper will suggest a solution to some of them, while others have to be left open for further research.

To start with, Harmonic Alignment as such is only defined if one of the two scales to be aligned is binary. However, there are natural configurations where both inputs have more elements. In the previous sections, we tacitly confined the hierarchy of grammatical functions to subject and object, but the full scale is much more articulated; it comprises at least the following elements:

(10) subject > direct object > indirect object

Suppose we want to align this hierarchy with the animacy hierarchy

(11) human > anim > non-anim

For the subject and the indirect object, we presumably get a copy and a mirror image of the animacy hierarchy:

(12) a. subject/human $\succ$ subject/anim $\succ$ subject/non-anim
b. i-object/non-anim $\succ$ i-object/anim $\succ$ i-object/human

It is unclear though what the harmony hierarchy for the direct object should be. Both (13a) and (b) can be justified

(13) a. d-object/human $\succ$ d-object/anim $\succ$ d-object/non-anim
b. d-object/non-anim $\succ$ d-object/anim $\succ$ d-object/human

At the present time, we have to leave this issue open.

The next points concern the nature of the OT constraints that implement Harmonic Alignment. It seems to be highly unnatural to assume constraints like "Avoid pronominal subjects" or "Avoid indefinite objects!" Technically this is harmless because they are always dominated by constraints that are effectively their negation. Nevertheless one rather does without constraints that exclude the least marked configurations one can imagine.

Likewise, the concept of constraint conjunction is technically compatible with the overall OT architecture, but it nonetheless does not fit in very naturally. It is one of the basic assumption of OT that one violation of a given constraint cannot be countered by arbitrarily many violations of lower constraints. Constraint conjunction undermines this. Consider the following constraint ranking:

(14) $C_1\&C_2 \gg C_3 \gg C_1 \gg C_2$

Effectively, this amounts to saying that violations of C_1 and C_2 each separately count less than a violation of C_3, but violations of C_1 and C_2 together sum up and are more severe than a single violation of C_3.

While this might be a marginal technical point, it appears to be *ad hoc* which constraints are conjoined with each other. The intuitive correlation of Harmonic Alignment and morphological marking is quite simple: Mark non-harmonic combinations! The OT formalization of this insight rests on the assumption that the constraints that are obtained from aligning markedness scales are conjoined with $*\emptyset$. It would be equally possible though to conjoin them with *STRUC instead. To take an example, if we exchange $*\emptyset$ and *STRUC in (9), we obtain the constraint hierarchy

(15) *STRUC & *Subj/3rd $\gg$ *STRUC & *Obj/local $\gg$ $*\emptyset$ $\gg$ *STRUC & *Subj/local $\gg$ *STRUC & *Obj/3rd

This constraint hierarchy describes the mirror image of Dyirbal, i.e. a language where only 3rd person objects and local person subjects are case marked. Briefly put, this hypothetical Anti-Dyirbal has case marking only on non-harmonic NPs. To our knowledge, no such language exists. Even stronger, the markedness regularities that Harmonic Alignment attempts to formalize in fact exclude such a language.

What is really at stake here is the status of constraints in OT. We are sympathetic with the hypothesis of Haspelmath (1999) that

> "the grammatical constraints are not innate, and are not part of Universal Grammar. They arise from general constraints on language use, which for the most part are in no way specific to language." (Haspelmath 1999:204)

As we will argue in the remainder of the paper, the markedness facts addressed by Aissen lend themselves in fact fairly naturally to the kind of functional explanation envisaged by Haspelmath.

4 Two Experiments

A way of explaining why morphology will appear on disharmonic elements (like human pronoun objects or non-specific subjects) is functional. The morphology marks the element as a subject or object and this helps the recognition of the elements as subjects and objects. Without the morphology, there would be a bias to interpret the elements as harmonic, i.e. recognize the human pronoun as a subject or recognize the non-specific NP as an object.

The bias would derive from the distribution in normal use of language. If human pronouns are normally interpreted as subject, interpreting the human pronoun as a subject is better than interpreting it as an object. We can see this as a conflict between two defeasible constraints,[2] one, **Generation** enforcing faithful interpretation of the morphology (adding a marker for a semantic property not in the input is bad), the other, **Bias** preferring the normal reading, where normal is defined as the reading that is available in most of the cases. There are two options for the interpretation of the second constraint. We could think of it as a question of yes and no: an interpretation is either normal or not, or it could be a question of preferences: the normal interpretation is preferred to the degree to which it is normal.

In either case, we would get a preference for normal interpretations. This means that when a semantic input is realized by means of disharmonic elements its preferred interpretation will be different from the input and by the weakest interpretation of bidirectionality that the realization is not available unless another and stronger constraint overrides **Bias**. Bidirectionality minimally requires that a good realization for an input is one that will (preferentially) be interpreted as that input and that would be the problem of disharmonic elements: they are syntactically allowed but their surface characteristic prefer an interpretation as an harmonic element. The combination of **Generation** and the extra morphology overrides this preference and allows the interpretation as a disharmonic element. Besides we take it that morphological marking should only be used if required by these constraints. Let us tentatively stipulate a constraint **Economy** (roughly corresponding to Aissen's *STRUC), that is violated by morphological marking. For the purposes of this paper, we will assume just these three constraints, ordered in the way indicated in (16).

(16) **Generation** $\gg$ **Bias** $\gg$ **Economy**

This explanation only works if in fact there is a bias towards harmonic elements in the natural distributions in language use. In this section, we present two corpus investigations which confirm that hypothesis and a third rather speculative argument to show that that sort of distribution is to be expected on the basis of three universal tendencies.

The first corpus we looked at is a large annotated corpus, the Wall Street Journal corpus, consisting of text taken from the newspaper. Here we have about 250,000 NPs, divided by the annotators into subjects and non-subjects. There is a majority of non-subjects here since non-direct objects cannot be distinguished from direct objects. By looking at the head nouns of NPs, these can be divided into human and inanimate NPs (the Wall Street Journal does not discuss animals very frequently). We can also make an approximate division into pronouns, definite NPs, specific NPs and non-specific NPs by classifying formal characteristics like determiners, name or non-name. But this remains a bit of a black art of dubious reliability: bare NPs can be non-specific and specific indefinites as well as definite NPs (names of kinds or persons), and a proper classification would be very costly. There are only a couple of thousand pronouns. Another question that can well be raised about this corpus is its representativity for natural language use: it is monologue and the topic seems to be almost exclusively the state of the economy.

What we expect to find is that disharmonic combinations have a lower frequency than would be expected, i.e. than the frequency of the either element in the combination. For example, we expect

$$p(OBJ|HUM) < p(OBJ|NP)$$

and

$$p(HUM|OBJ) < p(HUM|NP)$$

[2]The idea of doing this in this particular way is due to Jason Mattausch

And this is borne out. $p(OBJ|HUM) = 42\%$ and $p(OBJ|NP) = 75\%$, while $p(HUM|OBJ) = 10\%$ and $p(HUM|NP) = 13\%$.

But since we are in the business of interpretation, we want to know whether we can predict the abstract category (the syntactic function) from the surface property (a feature like HUM is given with the recognition of the NP). And we can derive from the above that assuming that a human NP is a subject pays off: the probability that the human NP is is a subject is 58%.

82% of the pronouns are subjects while there is only 25% probability of a being a subject in the corpus.

Definites (without the pronouns) slightly increase the probability of being an object (88% vs. 75%), while the other NP objects have exactly the same frequency as the objects (75%).

Indefinites slightly raise the probability of the NP being an object: it increases from 75% to 90%.

We find here strong evidence for two rules: assume that pronouns are are subjects and assume that humans are subjects, especially if we make the assumption that the probabilities for being a subject and a non-subject should be corrected to 50-50 (a high frequency of non-direct objects comes from long sentences which are not expected in the natural spoken language environment). On the object side we find a tendency that indefinites are objects and the reflexes of the two rules that bias towards a assuming a subject. But pronouns are low-frequent in the corpus (5%) as are human NP (13%) which makes it hard to see effects from lexicality (non-pronouns) or inanimacy.

The following table gives the relevant results.

$p(subj\|np)$	$=$	25%
$p(obj\|np)$	$=$	75%
$\mathbf{p(subj\|X)}$	$\gg$	$\mathbf{p(subj\|np)}$
$p(subj\|pro)$	$=$	88%
$p(subj\|hum)$	$=$	58%
$\mathbf{p(obj\|X)}$	$\gg$	$\mathbf{p(obj\|np)}$
$p(obj\|-def)$	$=$	87%
$p(obj\|inan)$	$=$	90%

No effects for $p(obj|indef) = 75\%$.

In summary, we find strong effects in the subjects but less clear effects in the object. This may well be due to the relative scarcity of pronouns and human NPs in the corpus. Failure to find an effect for indefinites may be due to the difficulties of finding a good heuristics for that class. But we get confirmation of our expectation: that harmony in NPs is connected with frequency: harmony boosts frequency.

Our second experiment used a much more suitable corpus, *Samtal i Goeteborg* (conversations in Gothenburg) which is a collection of taped and transcribed conversations obtained by asking Gothenburgians to record some everyday conversation they were engaged in. Oesten Dahl used the corpus to obtain the data for his Dahl (2000) and in the course of that entered about 10% of the utterances into a database with annotations that were perfectly suited for our task.

The main difference with the WSJ corpus is that pronouns are highly frequent (72%) and that human NPs abound (54%). Another difference is the much smaller number of NPs (13692) and having only direct objects, so that now the subjects are in the majority.

We get the following data. I use *ego* for the egocentric pronouns *I, you, we* and their alternants, *3pro* for the other pronouns, *-def* for the non-definite NPs, *lexdef* for the non-pronominal NPs that are definite.

$p(subj\|NP)$	$=$	77%
$p(obj\|NP)$	$=$	23%
$\mathbf{p(A\|B)}$	$\gg$	$\mathbf{p(A\|NP)}$
$p(subj\|hum)$	$=$	97%
$p(subj\|ego)$	$=$	97%
$p(obj\|-def)$	$=$	87%
$p(obj\|lexdef)$	$=$	32%
$p(obj\|inan)$	$=$	46%
$\mathbf{p(A\|B)}$	$\approx$	$\mathbf{p(A\|NP)}$
$p(obj\|def)$	$=$	15%
$p(obj\|3pro)$	$=$	17%

The Aissen lattice is completely reconstructed by probabilities with which subjecthood is predicted from the category. We obtain the following linear order from those probabilities (we order by the value of p(subj—X)).

(17) *human pronoun > inanimate pronoun > human lexical definite > inanimate lexical definite > human non-definite > inanimate non-definite.*

This is a linearization of Aissen's partial order and fully consistent with *human > inanimate and pronoun > definite > non-definite.*

It works out less neatly in the object ($p(obj|X)$). We get the ordering

(18) *inanimate pronoun > inanimate non-definite > inanimate lexical definite > human non-definite > human lexical definite > human pronoun*

This is consistent with *human > inanimate*, but does not respect *pronoun > non-definite* or *pronoun > lexical definite* on the inanimates. (It does on the human NPs). Whether we should be worried about this is debatable.

The data this time give robust effects both in the subject and the object though still stronger effects in the subject. The harmonic NPs are much more frequent than clearly disharmonic ones. And the corpus is clearly a natural one: it is the sort of language use that we engage in on a daily basis and that forms the basis for language learning. Is this enough to conclude that the frequencies are the same all over the world? We can adduce rather similar results (in another language, in another genre) of our previous experiment. Preliminary investigation of the SUSANNE corpus (a syntactically annotated collection of written English from different genres, cf. Sampson 1995), and the CHRISTINE corpus (transcriptions of spontaneous English dialogues which are annotated according to the SUSANNE scheme) show similar patterns, with CHRISTINE rather close to Samtal (with the exception of the indefinites which in Samtal have a much stronger preference for being objects) and SUSANNE like the WSJ corpus in having only minor effects in the object.

But as a case for universality it does not really add up to very much. There is however a way we can explain the data which does not seem to appeal to the peculiarities of Swedish or English.

It is generally accepted that subjects are the most agentive syntactic function. And the proto-agent properties of Dowty (1991) all have the tendency to make the referent more and more human. There is therefore a universal explanation of the fact that subjects tend to be human and humans tend to be subjects.

Similarly, objects tend to be foci or comments. This in turn makes it likely that new (both in the sense of material new to the context or material that is not in the current discourse topic). And these things are realized by lexical NPs and if they are fully new, by indefinite lexical NPs. How strong are these effects? Well, it seems we can assume the frequencies we found can be taken as representative. And can we then predict the other frequencies? Well, if we make some particular but not unreasonable assumptions.

The following picture is made by BayesBuilder a free software package that allows the modeling of dependencies.[3] It pictures statistical dependencies between our parameters and causal assumptions about how they influence each other. The bottom node (subject) is the part where the predictions are made: the NP is a subject or not with a certain probability. The picture shows the situation when an arbitrary NP is made: it has 28% probability of being egocentric, 16 percent of being indefinite, 18% of being definite, 38% of being a third person pronoun. Then there are dependencies between being lexical and human again determined by the corpus. The factors that push up the probability of being an object (indefiniteness and lexicality) are summated in the object box. The subject probabilities are computed from the object box and the human box: if they conflict they are put at 50-50, otherwise (for humans) at 97% for being subject, (for objects) at 100% for not being a subject and for the case when no factor applies at the overall probability for subjects (74%). BayesBuilder allows one to make more specific assumptions (e.g. lexical human) and then computes the resulting probability for the item being a subject. If one does that, the network gives the values we measured in the corpus within a couple of percentage points.

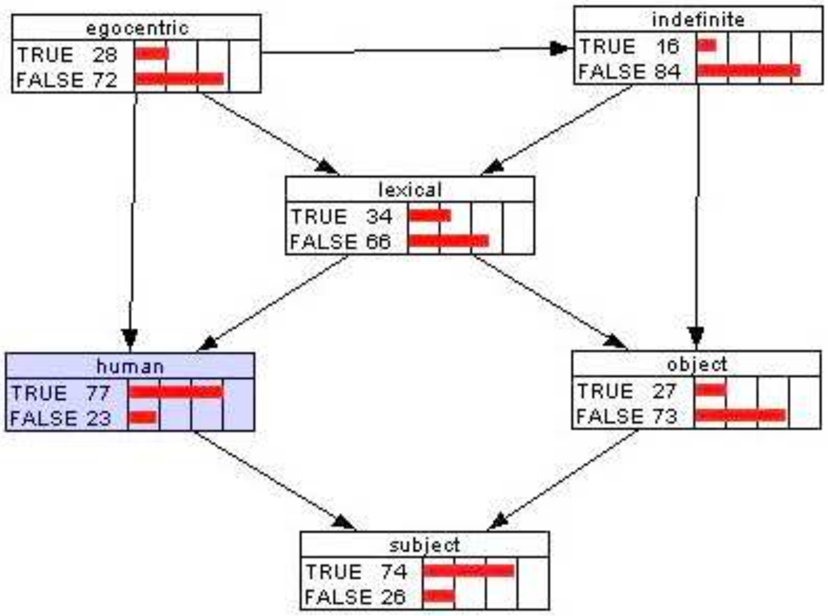

Figure 2: Bayesian Network

The assumptions and dependencies that we assume in the model are not unreason-

[3]Developed at Nijmegen University: http://www.mbfys.kun.nl/snn/Research/bayesbuilder/

able, but remain to some extent arbitrary. We show that a causal model is possible, based on the three factors indicated, not that we have found the true explanation.

If our basic line of thinking is right though, we show that the frequencies depend on other and deeper alignments: the alignment between subject and agentivity and the alignment of object with focus and comment. The frequencies are a surface effect of those alignments.

5 Theoretical repercussions

The results and considerations from the last section suggest that Aissen's syntactic alignment patterns can completely be explained in a functional way. Syntactic harmony basically means that a linguistic item conforms to the expectations that can be derived from the statistical patterns of language use. To take an example, it is a good heuristic to assume that a pronoun is a subject. Subject pronouns are in line with the corresponding expectation. Unmarked object pronouns would constantly risk to be misinterpreted as subjects. It is thus a good idea for a speaker to mark a pronoun with object case if it is supposed to be an object. As indicated above, we can formalize this intuition by assuming two constraints, **Bias** and **Generation**. **Bias** is fulfilled if an NP has the grammatical function that its semantic characteristics indicate—it is fulfilled by pronominal, definite, human etc. subjects and by indefinite, lexical, non-animated etc. objects, and violated by indefinite subjects, first person objects and the like. This is only one of its functions: it basically makes sure everything means what it normally means. **Generation** favors an interpretation that is faithful to the morphological case of the NP—accusative NPs are object, absolutive NPs are subjects etc. In general, it makes sure that all the syntactic rules are followed in generating the sentence. Finally, by a simple economy consideration, morphological marking should only be used when necessary—this is Aissen's constraint ***STRUC**, which we rename **Economy**.

The universal ranking of these constraints is

$$\textbf{Generation} \gg \textbf{Bias} \gg \textbf{Economy}$$

This has to be paired with a particular version of bidirectional evaluation. Let us say that a input-output pair is *optimal* iff 1. the output is optimal for the constraints encapsulated in **Generation** 2. there is no alternative input that is preferred by **Bias** and 3. there is no output that is equally good for the same input but which is shorter. (Note that this way of optimization is non-standard; OT usually takes solely the speaker perspective, but we assume that the speaker has to take in the hearer's point of view in order to make sure she will be understood. This makes our notion a special case of bidirectionality in the sense of Blutner 2001, though a different one from the two notions considered there.)

Now suppose we are in a language that has an accusative morpheme at its disposal. Suppose you are the hearer and you have to interpret a pronoun. There are two scenarios: the pronoun may be unmarked or carry accusative case. The unmarked pronoun is preferably interpreted as subject and the marked one as object.

(19)

			Generation	**Bias**	**Economy**
Pron+$\emptyset$	☞	Subj			
		Obj		*!	
Pron+ACC		Subj	*!		*
	☞	Obj		*	*

In both cases, from a pure speaker's perspective, there is a preference for the unmarked case. But bringing in the hearer in the **Bias** constraint brings a preference for the marked accusative pronoun.

While this account is tempting and, we feel, in a sense correct, it cannot be adopted in an unqualified way. It fails in two respects. First, optimization is not restricted to single NPs. We always have to consider at least an entire clause, and there other devices (like word order) may suffice to disambiguate, if they are not occupied for other expressive tasks. Even if this may seem to undermine our case for a language like English, it does not really undermine the functional account that we are advocating because the robust parsing required for speech recognition is impossible without redundancies and has to cope with sentence fragments. But it is an argument against the OT formalization just sketched.

Second, and more severely, the above account by itself has nothing to say about the typological pattern of DOM and DSM ("Differential Subject Marking") that Aissen discusses. It predicts that DOM applies to all non-harmonic combinations. There is no space for cross-linguistic parameterization like the fact that Hebrew marks all definite objects and Turkish all specifics. Assuming separate constraints for each alignment cell like Aissen seems to be in fact inevitable.

Nonetheless the discussion above gives a clue for a Haspelmath style functional motivation of the constraints involved. Starting from the proposal above, we suggest to split up **Bias** into separate interpretative constraints like "Pronouns are subjects", "Definites are subjects", "Indefinites are objects" etc. These constraints still express parsing heuristics that are founded in the statistical patterns of language use. The problem is however that we have nothing to say about those languages where the marking has to occur on perfectly harmonic elements, such as non-specific object NPs in accusative languages or subject pronouns as in pure ergative languages.

The next modification concerns the optimization algorithm as such. Above we used a version of bidirectional OT in which we derive why in certain circumstances subject and object markers appear on subjects and objects. This in essence gives us an explanation of optional subject and object marking. The constraint **Bias** is not just taking in regularities about the kinds of NPs that are subject and object, but also other kind of preferences. E.g. an inanimate and a human NP as arguments of the verb "to please" will almost certainly have the inanimate NP as the subject and the human NP as the object and would not have to be marked since the two effects of **Bias** obliterate each other.

Now it is rather clear that while optional case marking for NPs with a fixed set of features exists these are far less frequent than obligatory systems: an object must be case marked if it has certain features. Most case marking is obligatory. This can only be explained as a grammaticalization process: an optional marking possibility becomes required by the grammar.

Bias itself offers a way in which this process can be explained. Let us assume that the language marks 50% of its human object NPs counterbalancing frequency statistics which without the marking would make the human NP a subject with a probability of 60%. But the marking has an effect on these statistics: unmarked NPs are now subjects with not 60% probability but with 75% and marking becomes more necessary as a result, thus further reducing the probability that an unmarked NP is an object. Once an optional marking strategy becomes non-exceptional and if it is functional, the marking makes itself more necessary and will normally become obligatory. It is then for the language learner at some point not distinguishable from a generation rule that requires marking certain combinations of features. As its original functional motivation and the process of self-reinforcement are not transparent to new language learners, learning it as a generation rule becomes the only option for new language learners.

This explains why we normally find —next to bias-driven optional systems— obligatory rules of case marking. We predict —but have not investigated this empirically— that optional case marking only occurs where the frequency of the case-marked NPs is low and there is no functional pressure.

Another prediction of our functional explanation is that one finds marking in the disharmonic cases only. A language that marks low prominent subjects and high prominent objects seems to be the best way out of the predicament caused by **Bias**.

But this is only so if the language has both subject markers -ERG and object markers -ACC. Assume an input meaning “the apple hits John”. In the following tableau we rank the relevant possibilities.

(20)

		GENERATION	**BIAS**	**ECONOMY**
☞	apple John hit: hit(j,a)			
	apple John hit: hit(a,j)		*	
	apple-ERG John hit: hit(j,a)	*		
☞	apple-ERG John hit: hit(a,j)			
	apple John-ACC hit: hit(j,a)	*		
☞	apple John-ACC hit: hit(a,j)			
	apple-ERG John-ACC hit: hit(j,a)	*		*
	apple-ERG John-ACC hit: hit(a,j)			*

This is the optional system: accusative marking on John, or ergative marking on apple are the preferred options. (This changes when the marking is grammaticalized. In that case only the option from the last line of the tableau is open. The others do not pass **Generation**.)

In the optional system with both subject and object markers, we can see that all possibilities can be dealt with. High prominent subject and high prominent object is disambiguated by case on the object. Low prominent subject and low prominent object receives case on the subject. And high prominent subject and low prominent object does not need case-marking. Split ergative languages have grammaticalized the situation described here.

But if there is no subject marker, this changes. If we want to distinguish two low prominent arguments, we can only do this by case marking the object, even though low prominent objects are harmonic. Accusative languages are grammaticalizations of this situation.

If there is no object marker, the language has to distinguish two high-prominent arguments and grammaticalization of this results in a pure ergative language.

Now there is no need to put the events leading to this at the beginning of time. Case morphemes phonetically erode. We predict that if a subject marker disappears object marking will be extended and vice versa. The Spanish object marking of human NPs offers an example of how a preposition is reemployed as an object marker. (The lack of functional pressure that keeps it restricted to that class is due to subject marking by agreement and additional object marking by clitic doubling.)

6 Change

Very schematically, we can distinguish languages according to their behavior with subject and object marking. There are languages which mark all subjects ("Coherent Subject Marking": CSM), some subjects (DSM), all objects (COM) and some objects (DOM). And these can be combined. In addition a language can mark nothing. Assuming no other marking strategies, the language then is more or less able to distinguish subjects and objects, when these have high or low prominence. The following table shows the possibilities for the eight types to make clear which are the subjects and the objects given **Bias**.

(21)

subj-obj	high-high	high-low	low-high	low-low
∅	−	+	−	−
DOM	+	+	+	−
DSM	−	+	+	+
DOM+DSM	+	+	+	+
COM	+	+	+	+
CSM	+	+	+	+
COM+CSM	+	+	+	+

The empty system and the only DOM and only DSM systems are defective: they not mark certain situations. Other dimensions can be used (passivization or word order) but that also means that these are no longer free for other purposes like topic marking. Leaving aside those possibilities, such systems will be likely to develop marking.

Systems like CSM+COM overdo their marking on the other hand and can be said to be uneconomical: they force more marking effort on their users than is required by the functionality of the marking. The only three systems that have a good balance between the functional needs and economy are DOM+DSM, CSM and COM. But even such stable systems may change under the influence of phonetic erosion.

Five types of linguistic change are important in this perspective. The first is the process of *annexation*. If a lexical device can sometimes make a distinction which is useful for the interpretation of the utterance, a usage of that item may arise in which its purpose is to mark that distinction and not to convey its lexical meaning. Examples of such annexation are the use of *already* in Singapore English as a marker of perfectivity (in which its meaning "earlier than expected" is obliterated) or the use of the Latin preposition *ad* to mark objects in Spanish (obliterating its locative meaning).

There is something marked about this use of the marker: the lexical meaning of the marker is still available and this is rightly seen as analogous to epenthesis: the addition of extra phonetic material to the sentence. Semantic epenthesis can be formulated as follows. Do not put material in the sentence which has a meaning that is not in the input.

And epenthesis only occurs for a reason. In this case, we can identify that reason with our constraint **Bias**: non-use of the marker would mean that the default reading (subject or in the case of Singapore English Fong (2001) the non-perfective or futurate reading) will be generated.

Annexation is a process that turns a lexical item into an optional grammatical marker.

Optional marking reinforces **Bias**: the marked occurrences do not count and need to be subtracted. To use the example sketched above again, if 60% of the unmarked human NPs are subject, then 40 percent are objects. But if object human NPs are marked in 50% of the cases that means that unmarked human NPs are subjects in 75% of the cases and that the bias against object interpretation grows. This leads to further marking and even stronger bias against unmarked object human NPs. Unless the marking is exceptional or there is a choice between different markers, optional marking is not very stable.

Necessary marking can be seen as the extreme case where there is no possibility of non-biased interpretation. This is proper *grammaticalization*: the rule to use the marker when some condition applies becomes part of **Generation**. The complex semantics of these conditions is no doubt due to the categorization process that has to be built into the generation process itself.

Spread is the extension of necessary marking to a weaker condition. It can again be seen as a result of changing biases due to grammaticalization of markers or the annexation of a marker: unmarked NPs become more probable subjects the more often objects are marked, unless there are opposite biases or factors that guarantee that the proper readings are reached all the same. Spread needs optional marking and optional marking needs **Bias**-related misinterpretations.

Reinterpretation is a process by which optional marking becomes a method for expressing the trigger of marking rather than the case. If you have to be prominent to be marked for object, the speaker can make you prominent by object marking you. Object-marking is an economy transgression unless there is **Bias** against the intended interpretation. But violating economy is a good way of drawing attention to the referent of the expression that violates economy. That in turn can be used for different reasons, so that assigning prominence can be politeness, expression of respect, elevation of status for literary purposes, but can also become a conventional indication of semantic properties like specificity. (Again grammaticalization: **Bias** rigidly assigns the specificity to the object marked NP, the absence of the marker starts meaning non-specificity, the appearance of the object marker may become obligatory.)

Phonetic and semantic erosion is the last relevant changing process, but it is due to other factors. Grammatical markers can lose stress and unstressed elements can lose their phonetic profile to the point of obliteration. Grammaticalized markers can become more ambiguous and vague under spread. While we have discussed a number of stable situations, stability can be threatened by these phonetic and semantic obliteration processes. CSM may entirely disappear forcing new annexations or the spread of DSM, DOM may disappear forcing DSM to become CSM or new annexations to occur. The history of languages is cyclic.

While the Aissen system offers a good and concise way of describing the different grammatical systems with the necessary finegrained-ness, a proper explanation of DSM and DOM must take **Bias** and history into account.

7 Conclusion

The above is fairly speculative, but offers the beginning of an account of the typological observations of Aissen. We start out with a functional reinterpretation of subject and object marking and then explain how the grammars of particular languages come to contain certain marking rules, with application conditions as predicted by Aissen. The grammaticalization process has to live on an Aissen application condition (top or bottom of the prominence partial order), since such application conditions define areas where the frequency of subjects or objects is higher than in its complement.

We will follow up this work with a more thorough investigation of the historical processes that we assumed in the last section and try to evaluate them on real typological data. It is clear that much more needs to be done than we have been able to do here.

Acknowledgment We wish to express our gratitude to Marie Nilsenová, Reinhard Blutner, Hans-Martin Gärtner, the participants of Szklarska Poreba 2001 and of the Borjomi Conference for discussions which helped to shape the paper, Khalil Sima'an for help with the WSJ Corpus, Jennifer Spenader and Sofia Gustafson-Çapková for help with the corpus work and Östen Dahl for putting his annotations on *Samtal i Göteborg* at our disposal.

References

Aissen, J. (1999). Markedness and subject choice in Optimality Theory. *Natural Language and Linguistic Theory*, **17**, 673–711.

Aissen, J. (2000). Differential object marking: Iconicity vs. markedness. Manuscript, UCSC.

Blutner, R. (2001). Some aspects of optimality in natural language interpretation. *Journal of Semantics*, **17**(3), 189–216.

Bossong, G. (1985). *Differentielle Objektmarkierung in den neuiranischen Sprachen*. Günter Narr Verlag, Tübingen.

Bresnan, J., Dingare, S., and Manning, C. (2001). Soft constraints mirror hard constraints: Voice and person in English and Lummi. In M. Butt and T. H. King, editors, *Proceedings of the LFG01 Conference*. CSLI Publications, Stanford. to appear.

Dahl, Ö. (2000). Egophoricity in discourse and syntax. *Functions of Language*, **7**(1), 37–77.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, **67**, 547–619.

Fong, V. (2001). "already" in singapore english. Abstract, Aspect Conference, Utrecht, December 12-14.

Haspelmath, M. (1999). Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft*, **18**(2), 180–205.

Prince, A. and Smolensky, P. (1993). Optimality theory: Constraint interaction in generative grammar. Technical Report TR-2, Rutgers University Cognitive Science Center, New Brunswick, NJ.

Sampson, G. (1995). *English for the Computer: The SUSANNE Corpus and Analytic Scheme*. Clarendon Press, Oxford.

Silverstein, M. (1981). Case marking and the nature of language. *Australian Journal of Linguistics*, **1**, 227–244.