

final version, January 2007

Creating economical morphosyntactic patterns in language change*

MARTIN HASPELMATH

(to appear in: Jeff Good (ed.) *Language universals and language change*. Oxford: Oxford University Press)

1. Overview

My starting point in this contribution is the observation that apparently the great majority of **universal morphosyntactic asymmetries** are economically motivated and thus exemplify the slogan "grammars code best what speakers do most" (Du Bois 1985:363). By *morphosyntactic asymmetries* I refer to coding differences that do not express a meaning difference, e.g. the contrast between *the book* and *(*the) my book* (where the definite article *the* is impossible although from the meaning it would be expected to occur).

In §3-5 I will provide a substantial number of examples of economically motivated asymmetries. On the basis of these, I hypothesize that the pattern is even more general and that in fact the strong claim in (1) is correct.

- (1) All universal morphosyntactic asymmetries can be explained on the basis of frequency asymmetries, i.e. they all show economic motivation: More frequent patterns are coded with less material.

In a second step, I want to examine the ways in which the economic motivation is implemented in languages through diachronic change. Economic motivation, like other types of functional motivation, needs to be interpreted in diachronic terms: From the point of view of the speakers' grammars, economical patterns are arbitrary (because speakers would be able to acquire non-economical patterns just as well), but the changes leading to them are not accidental; they are motivated by economy. So far, my study of the diachronic origins of economical patterns has not yielded very strong constraints, so the claim in (2) is perhaps not as surprising as the claim in (1).

- (2) Diachronically, economical patterns arise by
- (i) differential phonological reduction,
 - (ii) by differential inhibition of periphrasis/ grammaticalization, or
 - (iii) by analogical change.
- "Differential morphosyntactic reduction", while logically perfectly possible, does not occur.

I will proceed by first explaining what I mean by "universal asymmetrical morphosyntactic patterns" (§2). Then I discuss the notion of "economical coding" (§3) and list eleven cases of "complementary expected associations" (§4), and four cases of "non-complementary expected associations" (§5). All

* I am grateful for comments to the participants of the Berkeley workshop on Explaining Linguistic Universals, to four anonymous reviewers, and to Jeff Good. A version of this paper was also presented at a symposium on grammaticalization at the Freie Universität Berlin in November 2003 organized by Ekkehard König and Volker Gast.

these cases can be regarded as universal morphosyntactic asymmetries, and for all of them an economy-based explanation is plausible. In §6 I discuss the diachronic origins of the fifteen economical patterns of §§4-5.

2. Universal asymmetrical morphosyntactic patterns

By "asymmetrical patterns" I refer to situations in which one class of expressions behaves differently from another class of expressions without any apparent semantic reason.

- (3) a. *a book of mine*
b. ??*the book of mine*
- (4) a. *I saw you.*
b. **You saw you.*
- (5) a. *Who said what?*
b. **What did who say?*
- (6) a. *book/book-s*
b. **stair/stairs*
- (7) a. *sing/sang*
b. *bring/*brang*
- (8) a. *I'm interested in the picture.*
b. **I'm surprised in the picture.*

In all these cases, the (b) expressions are perfectly interpretable, yet definite and indefinite NPs behave differently in possessive constructions (cf. 3), disjoint and coreferential objects behave differently (cf. 4), subjects and objects behave differently in multiple *wh*-questions (cf. 5), *books* and *stairs* behave differently in the singular (cf. 6), and so on. So these behavior differences are initially surprising and require the linguist's attention. Other behavioral differences are unsurprising because they follow from the meaning of the expressions. For instance, *have to* is similar to *want to* in taking an infinitival complement (*I want to play/I have to play*), but behaves differently from it in not allowing a different-subject infinitival complement (*I want the kids to play vs. *I have the kids to play*). But this is expected, because the latter sentence is simply uninterpretable on the obligation reading of *have to*. So I consider only those cases in which the asymmetry is surprising in the sense that symmetry would be possible and hence expected. In fact, in many cases there are languages which do not show the asymmetry in corresponding structures. For instance, Italian has no asymmetry in the equivalents of (3a-b) (*un mio libro/il mio libro*), and German has no asymmetry in the equivalents of (5a-b) (*Wer sagte was?/Was sagte wer?*).

Universal asymmetries are those that recur in language after language. Of the examples seen so far, (3) to (6) are of this type, but (7) and (8) are not, as far as we know: There is no cross-linguistic generalization under which the difference between *sing* and *bring*, and between *interested* and *surprised* could be subsumed. These seem to be language-particular idiosyncrasies, and no claims are made here about such non-universal, parochial patterns (which of course abound in all languages).

The universal patterns considered in this paper are all **implicational universals** in the classical Greenbergian sense (cf. Croft 2003:ch. 3). For instance, the difference between English and German with respect to multiple

wh-questions can be described as an implication: If a language allows multiple *wh*-questions with fronting of the object and the subject remaining in situ, it also allows multiple *wh*-questions with fronting of the subject and the object remaining in situ. The reverse is not true, as English shows. Such universals are "typological generalizations" in Kiparsky's sense, not "true universals" (see Kiparsky, this volume), because (i) they arise as a consequence of tendencies of diachronic change (rather than constraining change), (ii) they are not assumed to derive from the innate cognitive code for grammar (often called "Universal Grammar"),¹ and (iii) they are not necessarily exceptionless. Notice, however, that one of Kiparsky's examples of a "true universal", the "D-hierarchy", is here regarded as a typological generalization (see §4.4). In general, it seems that implicational universals are always "typological generalizations" in this sense, because proposals for Universal Grammar-derived implicational generalizations ("parametric effects") have been invariably unsuccessful (see Newmeyer 2004, 2005, Haspelmath 2007+b).

3. Economical coding

An expression shows **economical coding** compared to another expression if it is shorter (fewer words, fewer syllables, fewer segments) or otherwise requires less articulatory effort (e.g. less suprasegmental prominence). Such economical coding is functionally motivated if it occurs with frequently expressed meanings, while related rarer meanings are coded with more articulatory effort. Four different kinds of economical coding asymmetry can be distinguished, and linguists often treat these in very different ways. My point here is that they all instantiate economical coding, and when they occur systematically, they require a unified explanation.

3.1. Frequent: zero/Rare: overt²

In many cases, economical coding is manifested by a zero/nonzero contrast. Well-known examples are given in (9).

- | (9) frequent expression | rare expression |
|---|---|
| (i) a. singular: <i>book-∅</i> | b. plural: <i>book-s</i> |
| (ii) a. 3rd person:
Spanish <i>canta-∅</i> 'sings' | b. 2nd person:
<i>canta-s</i> 'you sing' |
| (iii) a. present: <i>I ∅ sing</i> | b. future: <i>I will sing</i> |

The overt nonzero element may be an affix (as in (9i-ii)) or a free word (as in (9iii)).

3.2. Frequent: shorter/Rare: longer

¹ See Haspelmath (2004b) for the term "cognitive code", referring to the innate cognitive prerequisites for acquiring language.

² The cases on the right-hand side are not rare in an absolute sense (whatever that sense might be), and neither are the cases on the left-hand side frequent in an absolute sense. All that matters here is that there is a significant frequency difference between them.

Economical coding may also be manifested by a long/short contrast.

- | | | |
|------|---|---------------------------------------|
| (10) | frequent expression | rare expression |
| | (i) a. Tamil inanimate locative <i>-il</i> | b. animate locative <i>-iṭam</i> |
| | (ii) a. Latin dative sg. <i>-ō/-ae/-ī</i> | b. dative plural <i>-īs/-īs/-ibus</i> |
| | (iii) a. Russian "middle" refl. <i>-sja</i> | b. ordinary reflexive <i>sebjā</i> |

In (10iii), the shorter expression is an affix, while the longer expression is a free form. This is another way in which short and long items often differ: Long forms tend to show greater freedom and behavioral possibilities, while short forms tend to obey tighter restrictions. Since this difference cannot be related directly to economy, I will not discuss it in this paper.³

3.3. Frequent: straightforward/Rare: roundabout. Often only the more frequently expressed meaning can be expressed in a straightforward way, while the rarer expression requires a roundabout construction:

- | | | |
|------|--|---|
| (11) | frequent expression | rare expression |
| | (i) a. <i>Gabriel's friend</i> | b. <i>a friend of Gabriel's</i> |
| | (ii) a. <i>I gave her it.</i> (British English) | b. <i>I gave it to Aisha.</i> |
| | (iii) a. German
<i>Ich will spielen.</i>
'I want to play.' | b. <i>Ich will, dass du spielst.</i>
'I want you to play.' |
| | (iv) a. Modern Greek
<i>Ton=iḗa.</i>
'I saw him.' | b. <i>Íḗa ton eafṭó=mu.</i>
'I saw myself.' |

In (11i/b), the possessor requires the extra preposition *of* and cannot occur in prenominal position. In (11ii/b), the full-NP recipient must occur with a preposition, so that the expression is longer and the two roles occur in different positions. In (11iii), the complement verb is finite and requires agreement with its subject, contrasting with the infinitive in the more frequent same-subject construction in (11iii/a). In (11iv), the reflexive pronoun does not occur in the preverbal pronominal slot, but in the postverbal slot of full-NP objects, i.e. despite its pronominal meaning it shows the behavior of a full NP.

While these cases are like those in (9) and (10) in that the rare expression is coded with more material, there are further differences in the constructions. But it is sometimes difficult to classify an asymmetric pattern, because an additional element may be regarded as itself constituting a constructional difference. Consider the pair in (12):

- | | | |
|------|---|--|
| (12) | frequent expression | rare expression |
| | a. <i>sing!</i> (imperative 2nd person) | b. <i>let her sing!</i> (imperative 3rd pers.) |

Here one would probably say that (12b) is a different construction and thus counts as "roundabout" like the (b) cases in (11), but it is also possible to argue

³ See Lehmann (1982 [1995]: ch. 4) for some discussion of the relation between shortness and lack of freedom.

that this is a simple case of a zero/overt contrast of the type seen in (9). Thus, the boundaries between §3.3 and §3.1 and §3.2 are not always clear, which is an additional motivation to treat these cases in the same way.

3.4. Frequent: existent/Rare: nonexistent. In many cases, a language simply lacks a way of expressing the rarer meaning.

- | (13) | frequent expression | rare expression |
|-------|--|--|
| | Tzutujil (Dayley 1985:145) | |
| (i) | a. <i>w-ati7t</i> 'my grandmother' | b. <i>*ati7t</i> 'grandmother' |
| (ii) | a. <i>juyu7</i> 'mountain' | b. <i>*w-juyu7</i> 'my mountain' |
| | Acehnese (Durie et al. 1994:177-8) | |
| (iii) | a. <i>Lôn-tém woe.</i>
I-want return
'I want to return.' | b. <i>*Lôn-tém droeneuh woe.</i>
I-want you return
'I want you to return.' |
| (iv) | a. <i>Who do you think that I met?</i> | b. <i>*Who do you wonder why I met?</i> |

At first glance, this case seems to be fundamentally different from the first three cases discussed in this section. However, it is clear that here, too, we observe "good" coding for frequently expressed meanings, in the sense that they can be coded straightforwardly, while the rarely expressed meanings are coded "badly" in an extreme sense: they cannot be expressed at all.

But when we look at this situation more closely, we realize that again, the dividing line between this case and the earlier clear cases of economical coding is not clear-cut. Languages are generally rich enough to render the meanings expressed by other languages, though sometimes significantly greater effort is required. So if a Tzutujil speaker is asked to translate 'grandmother' or 'my mountain' into Tzutujil (in an appropriate context), they would in all likelihood find a way of doing it. While Dayley's (1985) description is silent on this matter, one can easily imagine 'grandmother' being translated as 'a person's grandmother', and 'my mountain' as 'the mountain that belongs to me'. For Acehnese, Durie et al. (1994) suggest the sentence in (14) as a way of rendering 'I want you to return'.

- (14) Acehnese (Durie et al. 1994: 177-8)
- | | | |
|-----------------|-----------------|---------------------|
| <i>Lôn-lakè</i> | <i>droeneuh</i> | <i>beu-neu-woe.</i> |
| I-ask | you | HORT-you-return |
- 'I ask you to return.' (≈ I want you to return)

And if an English speaker wanted to say 'Who do you wonder why I met?' in acceptable English, they could resort to a non-standard structure involving a resumptive pronoun (Who do you wonder why I met them?), or to a highly stilted expression such as *For whom is it the case that you wonder why I met him/her?*

Thus, under the assumption that selective ineffability does not exist (i.e. any language can be translated into any other language), the existent/nonexistent contrast reduces to the straightforward/roundabout contrast of §3.3, though the degree of roundaboutness may be significantly higher in the examples in (13). While in the cases of §3.3, linguists perceive the (a) and (b) cases as corresponding members of an opposition, this is not the

case in §3.4, but of course we do not know whether such a difference exists for the speakers and how significant the difference is.

For this reason, I will consider cases like those in (13) as cases of economical coding, too. This goes beyond the common practice in the literature, but this is well-motivated because the cases of (13) all fall under the generalization in (1) (that is, if a meaning can be expressed directly in language A but only in a roundabout way in language B, then this meaning tends to be comparatively rare in language use).

Let us now look at a few of the asymmetries in somewhat greater detail, beginning with complementary expected associations.

4. Eleven complementary expected associations

Morphosyntactic asymmetries often arise when two grammatical or semantic properties cooccur in one expression, and a particular value of property 1 is **typically associated** with a particular value of property 2. An example of this is the cooccurrence of *person* (= property 1) and *mood* (= property 2) in verb forms. In this case, the value "2nd" of person is typically associated with the value "imperative" of mood, i.e. the association "2nd/imperative" occurs more frequently (and is hence more expected) than others such as "3rd/imperative"; conversely, value "3rd" is the expected value of person with the value "indicative" of mood, so that the association "3rd/indicative" is more frequent than "2nd/indicative". This is thus an example of a **complementary expected association** (what Croft 1990:§6.3 calls "complementary prototype").

In the following subsections, I will describe eleven such complementary expected associations, showing that they exhibit great similarities in their formal behavior across languages, and that this correlates in all cases with frequency asymmetries.

4.1. Person and mood

Our first example is the cooccurrence of person and mood. To simplify matters, we only look at 2nd and 3rd person, and at indicative and imperative mood, so that there are four possible associations of values, as shown in the four cells in Figure 1.

		person	
		3rd	2nd
mood	indicative	Spanish <i>canta-Ø</i> sing.IND-3SG 's/he sings'	Spanish <i>canta-s</i> sing.IND-2SG 'you sing'
	imperative	Latin <i>lauda-to</i> praise.IMPV-3SG 'let him praise'	Latin <i>lauda-Ø</i> praise.IMPV-2SG 'praise!'

Figure 1

In this figure (and in the following figures in this section), the boxed cells are frequent (and hence expected) associations and show zero/short coding, while the non-boxed cells are rare/unexpected and show overt/longer coding. The overt coding elements in the non-boxed cells are highlighted by boldface. The cells contain minimally contrasting examples from a single language or two languages, but the figure (as well as all the figures that follow below) embodies two universal claims:

- (i) **Universal Frequency Asymmetry:** In all languages, what is identified as "expected association" here is more frequent than the corresponding "unexpected associations".
- (ii) **Systematic Coding Asymmetry:** In all languages, the expected associations are at least as economically coded as the unexpected associations.

In order to make my claims water-tight, I would have to provide evidence for both of these claims for each expected association. This is too ambitious a task, and to the extent that I do not provide the evidence, my contribution must be seen as a providing a program for research rather than an empirical contribution. However, in most cases impressionistic observations give initial plausibility to my claims, and in some cases I will mention the results of very preliminary frequency counts that I did myself. These are mainly intended to show that the impressionistic observations do seem to have a basis in the facts and that doing systematic text counts of representative (i.e. colloquial) texts is a promising task.

The next two paragraphs describe the evidence that has been provided so far in the literature.

(i) Evidence for frequency asymmetry: Greenberg (1966:47) provides text counts from three languages showing that the indicative is by far the most frequent mood, but he does not take person into account, and his texts do not represent colloquial speech and may therefore not be truly representative. However, it seems clear that in all languages, imperatives are most often second person (or first person inclusive), because it is more efficient to address a command to the agent than to someone else.

(ii) Evidence for systematic coding asymmetry: Siewierska (2005) finds that out of 284 languages with verbal person marking, 103 have some zero-marked third person forms. However, she does not compare these figures with zeroes in first and second person forms. Bybee (1985:53) reports that 54% of her relevant languages (15 out of 28) use a zero for third person subject agreement, while only 14% (4) use a zero for first person, and only one language (Georgian) uses a zero for second person indicative. For the imperative, the papers in Xrakovskij (ed. 2001) show that there is a tendency for 2nd person imperatives to have zero-marking for person, and for 3rd person imperatives to have overt marking.

4.2. Number and gregariousness

The next example is the intersection of number and the lexical-semantic feature that I call here "gregariousness", i.e. this time we look at a grammatical-semantic property and a lexical-semantic property. "Individualist" nouns such as 'house' tend to be singular, and "gregarious" nouns such as 'leaf' tend to be plural. The effect on morphological coding can

be seen in languages such as Krongo (Kadugli; Sudan; Reh 1985:101ff.), which have both overt plural markers and singulative (=overtly coded singular) markers:

		number	
		singular	plural
gregariousness	individualist	Krongo <i>Ø-còorì</i> SG-house 'house'	Krongo <i>nóo-còorì</i> PL-house 'houses'
	gregarious	Krongo <i>ñ-táarù</i> SGT-wood 'leaf'	Krongo <i>Ø-táarù</i> PL-leaf 'leaves'

Figure 2

(i) Evidence for frequency asymmetry: Greenberg (1966:32) shows with counts from four languages that the singular is overall more frequent than the plural, and this seems to be due to the fact languages have more individualist nouns than gregarious nouns. Tiersma (1982) shows that some nouns such as 'arm', 'horn', 'tooth', 'stocking', 'thorn', 'tear', 'leaf' tend to occur more frequently in the plural. It is these nouns that are called "gregarious" here.

(ii) Evidence for systematic coding asymmetry: Greenberg's (1963) universal number 35 states: "There is no language in which the plural does not have some nonzero allomorphs, whereas there are languages in which the singular is expressed only by zero." There are many languages in which all nouns behave like individualist nouns, but whenever there are singulative markers, they seem to occur with gregarious nouns. However, I know of no cross-linguistic studies of the lexical semantics of singulatives.

4.3. (Co-)reference and "vertedness"

The frequency of coreference vs. disjoint reference of subject and object is affected by the lexical-semantic property of "vertedness": Some verbs such as 'see' are **extroverted**, i.e. the transitive action is typically directed toward another referent, whereas other verbs such as 'wash', 'shave', 'dress', 'defend' are **introverted**, i.e. the transitive action is typically directed toward the self (these terms are from Haiman 1983; see also Kemmer 1993, König and Siemund 1999). Examples come from standard Arabic and Hua (Trans-New Guinea; PNG; Haiman 1983:807):

		(co-)reference of subject and object	
		disjoint	coreferential
vertedness	extroverted	Arabic <i>raʔay-tu-ka</i> saw-1SG.SUBJ-2SG.OBJ 'I saw you'	Arabic <i>raʔay-tu nafs-ii</i> saw-1SG.SUBJ self-1SG 'I saw myself'

introverted	Hua <i>zoda k-toe</i> I.wash you-I.put 'I washed you'	Hua \emptyset - <i>zoe</i> REFL-I.wash 'I washed \emptyset '

Figure 3

(i) Evidence for frequency asymmetry: Already Faltz (1985:7) characterized introverted verbs as denoting actions "that are commonly performed reflexively by people". In Haspelmath (2007+a), I cite frequency data supporting this characterization.

(ii) Evidence for systematic coding asymmetry: Haiman (1983) and Kemmer (1993:24-28) observe that introverted verbs are typically coded with a shorter markers, and some languages have zero-coding for reflexive use with these verbs (e.g. English). Occasionally the disjoint (non-reflexive) use of these verbs requires a special overt marker, as in Hua (cf. Figure 3). The ordinary reflexive pronoun used with extroverted verbs is usually at least as long as the non-reflexive pronouns (cf. Faltz 1985, Haspelmath 2007+a), and the situation of Arabic (cf. Figure 3), where disjoint pronouns are affixes but coreferential pronouns are full NPs, is quite typical.

4.4. Role and animacy

It is widely known that there is an interesting relationship between role and animacy: agents tend to be animate, and patients tend to be inanimate. The following examples are from Spanish and from Dyirbal (Pama-Nyungan; Queensland, Australia; Dixon 1972:42).

		role	
		agent (=A)	patient (=P)
animacy	human	Spanish <i>Juan-\emptyset me ve.</i> Juan me sees 'Juan sees me.'	Spanish <i>Veo a Juan.</i> I.see ACC Juan 'I see Juan.'
	non-human	Dyirbal <i>yamani-gu</i> rainbow-ERG	Dyirbal <i>yamani-\emptyset</i> rainbow-ACC

Figure 4

(i) Evidence for frequency asymmetry: Comrie (1989:128) claims that "...the most natural kind of transitive construction is one where the A is high in animacy and definiteness, and the P is lower in animacy and definiteness", but what does "natural" mean here? It must mean the same as "expected", and the expectation derives from the frequency asymmetry. Jäger (2004) provides frequency data that fully support this.

(ii) Evidence for systematic coding asymmetry: The tendency for animate objects to be marked with more material ("Differential Object Marking") has often been observed (Blansitt 1973, Silverstein 1976, Comrie 1989: ch. 6, Bossong 1985, 1998, Lazard 2001, Aissen 2003) and need not be discussed further. The tendency for inanimate subjects to be marked in a special way

has been less widely discussed and is less widely observed, but to the extent that "differential subject marking" is observed, it also shows a systematic coding asymmetry (see e.g. Dixon 1994:83-97).

4.5. Possessedness and alienability

In many languages, "inalienable" nouns (mainly body-part and kinship terms) and "alienable" nouns (all others) behave differently in possessive constructions. It used to be thought that this difference has to do with the different conceptualization of types of possession in these languages, but Nichols (1988:579) pointed out that the different behavior of inalienables simply has to do with the fact that they occur more often as possessed nouns than the alienables. The examples in Table 5 are from Koyukon (Athabaskan; Alaska; Thompson 1996).

		possessedness	
		unpossessed	possessed
alienability	alienable	Koyukon <i>Ø-tet</i> UNPOSS -sock 'socks'	Koyukon <i>se-tel-e'</i> 1SG-socks-POSS 'my socks'
	inalienable	Koyukon <i>k'e-tlee'</i> UNPOSS-head 'head'	Koyukon <i>se-tlee'-Ø</i> 1SG-head-POSS 'my head'

Figure 5

(i) Evidence for frequency asymmetry: I know of no frequency counts in the literature, but again it is quite easy to do such counts on texts. The figures in Table 1 are based on the German Goethe-Corpus of the COSMAS database⁴ and show a few person-denoting nouns, occupational terms and kinship terms.

		unpossessed	possessed
alienable	<i>Gärtner</i> 'gardener'	24	0
	<i>Jäger</i> 'hunter'	48	2
	<i>Pfarrer</i> 'priest'	12	0
inalienable	<i>Schwester</i> 'sister'	32	58
	<i>Tante</i> 'aunt'	47	22
	<i>Tochter</i> 'daughter'	46	53

Table 1

As expected, occupational terms are hardly ever possessed, while kinship terms are very often possessed. (There are relatively many cases of overtly unpossessed kinship terms because possessive pronouns may be left implicit in German, unlike in English.)

(ii) Evidence for systematic coding asymmetry: Nichols (1988:579) observes that "the possessive affixes used on the closed ('inalienable') set of nouns are

⁴ Institut für deutsche Sprache, Mannheim (<http://corpora.ids-mannheim.de/~cosmas/>).

typically shorter, [and] involve fewer morphemes than the open class...". Some languages simply lack unpossessed forms of inalienable nouns, and in some languages some alienable nouns are unpossessible (see the discussion of Tzutujil in §3), but all languages allow inalienables to be possessed, and all languages allow alienables to be unpossessed.

4.6. Pragmatic function and semantic class

Most people would readily agree that nouns are predisposed for referring and verbs for predicating, but of course it is also possible to refer with verb-based expressions (e.g. nominalizations), and to predicate nouns (by using a copula). The predisposition again must be interpreted in terms of frequency (cf. Croft 1991: ch. 2): Thing-denoting words more often have the pragmatic function of reference, and action-denoting words more often have the pragmatic function of predication.

		pragmatic function	
		reference	predication
semantic class	thing	English <i>book-Ø</i> (no nominalizer)	English <i>is a book</i>
	action	English <i>buy-ing</i>	English <i>Ø buys</i> (no copula)

Figure 6

(i) Evidence for frequency asymmetry: Croft (1991:§2.5) presents the results from text counts in four languages (Quiché, Nguna, Soddo, Ute), showing that thing-words are most often used for referring, and action-words are most often used for predicating.

(ii) Evidence for systematic coding asymmetry: Croft (1991:§2.3) shows that across languages, overt "function-indicating morphosyntax" (such as copulas and nominalizers) tends to occur on the less preferred associations of semantic class (action/thing) and pragmatic function (reference/predication). Thing-words hardly ever require an overt nominalizer, and action-words rarely require an overt copula.

4.7. (Co-)reference and complement-taking verb class

Different complement-taking verbs have different preferences with regard to coreference of the main-clause subject with the subordinate subject. Verbs like 'want' have a strong preference for same-subject complements, while verbs like 'say' may have the opposite preference, for different-subject complements. The examples in Table 7 are from Fongbe (Kwa, Niger-Congo; Togo; Lefebvre and Brousseau 2002: 78-82) and German.

		(co-)reference of main and subordinate subject	
		disjoint	coreferential
verb class	'say'	Fongbe <i>K`ɔkú d`ɔ é m`ɔ X</i> Koku said he saw X 'Koku _i said he _j saw X.'	Fongbe <i>K`ɔkú d`ɔ émi m`ɔ X</i> Koku said he saw X 'Koku _i said he _i saw X.'
	'want'	German <i>Ich will, dass du komm-st.</i> I want that you come-2sg 'I want you to come.'	German <i>Ich will Ø kommen.</i> I WANT come 'I want to come.'

Figure 7

(i) Evidence for frequency asymmetry: I know of no evidence for complements of 'say' or other utterance predicates, but I did text counts for 'want' complements for two languages, Italian (Table 2) and Modern Greek (Table 3) (from Haspelmath 1999c):

text frequency: (Italian)	forms of <i>volere</i> 'want'	509	100%
	same-subject	444	87%
	different-subject	65	13%
Source:	Alessandro Manzoni, <i>I promessi sposi</i> , 1840-42. (Letteratura Italiana Zanichelli (LIZ) on CD-ROM)		

Table 2

text frequency: (M. Greek)	forms of <i>thélo</i> 'want'	43	100%
	same-subject	38	88%
	different-subject	5	12%
Source:	Kóstas Tzamális, <i>Stin Athína tu Periklí</i> , Athen: Estía/Kollaru, 22-122.		

Table 3

So clearly, same-subject complements are far more frequent than different-subject complements with 'want'. The Modern Greek evidence is particularly interesting because unlike German and Italian, Greek does not have two different constructions, i.e. it does not show any coding asymmetry. Thus, the frequency asymmetry cannot be due to the coding asymmetry.

(ii) Evidence for systematic coding asymmetry: In Haspelmath (1999c), I studied same-subject and different-subject complements of 'want' in 50 languages, and concluded that different-subject complement constructions are at least as formally complex as same-subject complements in all languages. Some languages simply lack different-subject complements of 'want' (cf. the discussion of Acehnese in §3 above).

4.8. Possessedness and definiteness

In Haspelmath (1999a), I pointed out that possessed noun phrases have a greater chance to be definite than unpossessed noun phrases, so that definite articles are relatively redundant in possessed NPs, while indefinite articles are relatively redundant in unpossessed NPs.

		possessedness	
		possessed	unpossessed
definiteness	definite	English <i>my</i> \emptyset <i>book</i> (no definite article)	English <i>the</i> <i>book</i>
	indefinite	Hebrew <i>ehad mi-sfar-ay</i> one of-books-1SG 'one of my books'	Hebrew <i>sefer</i> \emptyset (no indefinite article)

Figure 8

(i) Evidence for frequency asymmetry: See Haspelmath (1999a) for text counts in three languages.

(ii) Evidence for systematic coding asymmetry: In Haspelmath (1999a), I provided cross-linguistic evidence showing that there is a tendency for the definite article to be omitted in possessed NPs. The pattern found in English is attested (in various guises) in languages of different families.

4.9. Tense and participial voice

Comrie (1981) pointed out that there is a semantic affinity between perfect aspect and passive orientation, and we can expect this to be reflected in the frequency with which active and passive perfects and non-perfects are used. In participles, the expectation is that present participles tend to have active orientation (cf. English *stealing*), while past/perfect participles tend to have passive orientation (cf. English *stolen*).

		tense	
		present	past
participial voice	active	English <i>steal-ing</i>	English <i>having stolen</i>
	passive	English <i>being stolen</i>	English <i>stol-en</i>

Figure 9

(i) Evidence for frequency asymmetry: In Haspelmath (1994), I did a text count of a language that shows no asymmetry in its system of participles (Lezgian; Nakh-Daghestanian; cf. Haspelmath 1993a) and showed that perfect and past participles more often show object-orientation than imperfective and habitual participles.

(ii) Evidence for systematic coding asymmetry: In Haspelmath (1994), I looked at systems of participles in a range of languages and found that when there are asymmetries in the systems, passive participles tend to be associated with perfect or past-tense interpretation, and active participles tend to be associated with present-tense or imperfective interpretation. Past active

participles and present passive participles tend to be nonexistent, or (as in English) have to be expressed in more complex ways.

4.10. Sex and typical occupations

It is well-known that occupational terms tend to show coding asymmetries with respect to sex, in such a way that male terms have no particular marking, while female terms exhibit a special affix.

		sex	
		male	female
typical occupations	typically male	English <i>poet-Ø</i>	English <i>poet-ess</i>
	typically female	English <i>male nurse</i>	English <i>Ø nurse</i>

Figure 10

(i) Evidence for frequency asymmetry: I know of no text counts, but a cursory glance at a frequency dictionary of English confirms the expectation that male occupational terms are generally more frequent than female terms (e.g. *king* 176, *queen* 80; *actor* 36, *actress* 13; *policeman* 34; *policewoman* <10; Leech et al. 2001). Of course, this does not apply to typically female occupations, but at least in pre-20th century societies, when most of the relevant linguistic patterns were formed, most occupations were far more often held by men than by women. And some typically female occupations do show the opposite coding pattern.

(ii) Evidence for systematic coding asymmetry: I know of no systematic cross-linguistic study of male and female occupational terms, but there seems to be little doubt that the English (and more generally, European) pattern is systematic across languages.

4.11. Transitivity and type of causation

Among situations that can alternatively be construed as occurring spontaneously and as occurring through external causation, two broad types can be distinguished (Haspelmath 1993b): **costly events**, for which it is more likely that the event is seen as caused by an agent (e.g. 'split', 'break', 'open', 'close', 'gather'), and **automatic events**, for which it is more likely that the event is seen as occurring spontaneously (e.g. 'boil', 'freeze', 'dry', 'sink', 'melt').

		transitivity	
		transitive	intransitive
type of causation	costly events	Japanese <i>sak-Ø-u</i> 'split (tr.)'	Japanese <i>sak-e-ru</i> 'split (intr.)'
	automatic events	Japanese <i>kawak-as-u</i> 'dry (tr.)'	Japanese <i>kawak-Ø-u</i> 'dry (intr.)'

Figure 11

(i) Evidence for frequency asymmetry: Wright (2001:127-128) lists frequency figures for 28 change-of-state verbs in English and finds significant differences between different English verbs. Verbs denoting events that I would describe as "automatic" (such as *corrode*, *ferment*, *dry*, *freeze*, *melt*) show a relatively low proportion of transitive occurrences, while verbs that I would describe as "costly" (such as *break*, *open*, *bake*) show a relatively high proportion of transitive occurrences.

(ii) Evidence for systematic coding asymmetry: Evidence from 21 languages was presented in Haspelmath (1993b), showing that there is a tendency for costly events to be coded by transitive/anticausative alternations, and for automatic events to be coded by causative/intransitive alternations. There are also tendencies within particular languages to be either uniformly causativizing or uniformly anticausativizing, but the correlation between the coding types and the types of verbs can be seen despite these language-particular differences.

5. Non-complementary expected patterns

Not all frequency asymmetries occur in complementary expected associations. Four such cases are presented in this section.

5.1. Agent-patient direct/inverse

Quite a few languages require special coding effort for situations in which the person scale (1/2 > 3) is not aligned optimally with the role scale (Agent > Patient). These patterns are often called "inverse patterns".

direct (1 → 3, etc.)	inverse (3 → 1, etc.)
Nocte <i>nga-ma ate hetho-Ø-ang</i> I-ERG he teach-1SG 'I will teach him.'	Nocte <i>ate-ma nga-nang hetho-h-ang</i> he-ERG I-ACC teach-INV-1SG 'He will teach me.'

Figure 12

(Nocte: Tibeto-Burman; India; DeLancey 1981:641)

(i) Evidence for frequency asymmetry: I know of no text counts, so there is no concrete evidence yet for a frequency asymmetry.

(ii) Evidence for systematic coding asymmetry: See Zúñiga (2006) for a recent overview of inverse systems in languages of the Americas.

5.2. Recipient-theme direct/inverse

Although this is less well-known, inverse-like patterns are also found in ditransitive constructions. In languages with bound object pronouns for both Recipient and Theme, it is often the case that bound object pronoun combinations are not possible when the situation is "inverse", i.e. there is no alignment of the person scale (1/2 > 3) with the role scale (Recipient > Theme). A better-known example of this is French.

direct (1 → 3, etc.)	inverse (3 → 1, etc.)
French <i>Ma mère me le donne.</i> 'My mother gives it/him to me.'	French <i>Ma mère me donne à lui.</i> 'My mother gives me to him.' (* <i>Ma mère me lui donne.</i>)

Figure 13

(i) Evidence for frequency asymmetry: Haspelmath (2004a) presents some evidence from a German corpus that the inverse combinations are significantly rarer than the direct combinations.

(ii) Evidence for systematic coding asymmetry: Haspelmath (2004a) formulates the Ditransitive Person-Role Constraint and provides evidence for it from a wide variety of languages. No counterexamples have been found.

5.3. Theme and instrument relativization

The famous Keenan-Comrie accessibility hierarchy for relativization is also an instance of economical coding. The grammatical relations that are "inaccessible to relativization" in a given language are not inaccessible in an absolute sense, because there is usually a way to get around the constraints by employing a more complex construction. Most famously, a grammatical-relation-changing affix on the verb can promote a less accessible relation to a more accessible one. For example, in (15) relativization on the instrument is not possible, but apart from the ordinary construction as in (16a), Dyirbal also allows the applicative as in (16b), in which the instrument occurs as a core argument in the absolutive case (Dixon 1994:169-71).

(15) **yugu* [*yabu numa-ngu balga-ŋu*] *jaja-ngu bura-n*.
 stick.ABS mother.ABS father.ERG hit-REL child-ERG see-NONFUT
 'The child saw the stick with which father hit mother.'

(16) a. *yabu numa-ngu balga-n yugu-ngu*.
 mother.ABS father-ERG hit-NONFUT stick-INSTR
 'Father hit mother with a stick.'
 b. *yugu numa-ngu balgal-ma-n yabu-gu*.
 stick.ABS father-ERG hit-APPL-NONFUT mother-DAT
 'Father hit mother with a stick.'

The instrument can be relativized after applicativization, as shown in Figure 14.

semantic role	
theme	instrument
Dyirbal <i>ɲuma [banaga-ɲu] yabu-ɲgu</i> father.ABS return-REL mother-ERG <i>buran.</i> see-NONFUT 'Mother saw father who was returning.'	Dyirbal <i>yugu [ɲuma-ɲgu balgal-ma-ɲu yabu-gu]</i> stick.ABS father.ERG hit-APPL-REL mother-DAT <i>jaja-ɲgu bura-n.</i> child-ERG see-NONFUT 'The child saw the stick that father used to hit mother.'

Figure 14

- (i) Evidence for frequency asymmetry: The KeeFatnan-Comrie hierarchy has not in general been presented as a hierarchy of increasing rarity of relativization, but I strongly suspect that this is what underlies it.
- (ii) Evidence for systematic coding asymmetry: Keenan and Comrie (1977) have provided ample evidence for the coding asymmetry. They do not in general say which strategies languages employ when they want to relativize on a grammatical relation that cannot be relativized on directly, but clearly all languages employ strategies that are in some way more elaborate or complex.

5.4. 'Think' vs. 'wonder' extraction

Linguists have not usually seen extraction restrictions in the context of frequency and economy, but I would like to suggest that the systematic grammatical asymmetries observed in this domain ultimately have to do with frequency, too.

subordinate-clause type	
'think that'	'wonder why'
English <i>Who do you think that I met?</i>	English <i>Who do you wonder why I met them?</i> (*Who do you wonder why I met?)

Figure 15

- (i) Evidence for frequency asymmetry: Since extraction phenomena occur very rarely in texts, it will be extremely difficult to do text counts, and indirect estimations of text frequency will have to suffice.
- (ii) Evidence for systematic coding asymmetry: See Hawkins (1999, 2004: ch. 7) for some discussion of universals of extraction.

6. The diachronic origins of economical/well-coded patterns

Functionalists have sometimes been content with pointing out usage-grammar correspondences, because they confirm the expectation that "grammars code best what speakers do most" (Du Bois 1985:363). But why are grammars well-designed for the purpose of speaking and understanding?

Why do they code best what speakers do most? Human beings are used to working with instruments that are well-designed for their purposes, and in the case of human-made artifacts, such good design is not surprising because the creators' plan provides the link between the purpose and the structure of the instrument. For language, there is no such plan, so we need a theory that explains how language use and language structure are connected.

Following Bybee (1988), Keller (1994), Kirby (1999), and related work (cf. also Haspelmath 1999b), I claim that diachronic change is the necessary link between patterns of language use and grammatical structures. Speakers do not intend to create well-designed grammars, but they behave purposefully and rationally in selecting from available variants and in creating new variants – they mostly opt for the most useful variants for their particular purposes. Through an invisible-hand process in language change, the cumulative effect of many individuals' behavior leads to useful language structures (cf. Keller 1994).

So how do economical patterns arise in language change? There are two rather different routes by which this can happen: differential phonological reduction (§6.1) and differential expansion of a new construction (§6.2). Moreover, a minor route, morphological analogy, must also be recognized (§6.3).

6.1. Differential phonological reduction

According to George Zipf (e.g. 1935, 1949), who was one of the first authors to emphasize the role of frequency for understanding linguistic structure, there is one main diachronic pathway by which frequency leads to shortness: **differential phonological reduction**. Frequent expressions tend to get shortened by phonological reduction more strongly than rare items. And of course there is no doubt that this is an important mechanism of change leading to economical patterns:⁵ Frequently occurring items are generally more predictable than rarely occurring items, so that hearers can decode the message even if it is not very carefully articulated. As a result, speakers tend to articulate them less carefully, so that they undergo faster phonological change than less frequent items. High frequency of use also leads to automatization, often implying affixation of formerly free function words (cf. Bybee 2003).⁶

Differential phonological reduction can be seen as responsible for a few of the economical patterns that we saw earlier. Some examples of phonological changes resulting in asymmetric coding are shown in (17)-(21).

(17) Person (§4.1)⁷

⁵ Historical linguists working in the Neogrammarian tradition have often ignored this insight, but there is massive evidence for it (see, e.g., the work of Witold Mańczak, such as Mańczak 1987).

⁶ Bybee (2003) states that frequency-based automatization or routinization is not only responsible for increasing cohesion (=affixation etc.), but also for phonological reduction. This is implausible, because frequently uttered and automatized expressions do not get reduced when they happen to be hard to predict. This is the case, for example, in the speech of individuals with regard to their names. I say my surname *Haspelmath* very often (at least much more often than people with other surnames), but I do not reduce it more than other people, because it has very low predictability. (I do reduce my signature, not because it is more predictable, but because there is no need for it to be recognizable at all.)

⁷ Johanna Nichols (p.c.) doubts that the cases in (17) and (19) are due to differential phonological reduction, echoing some traditional views on these changes in the

Polish 3rd person singular indicative forms became zero by special phonological reduction (cf. Proto-Slavic and the cognate language Russian):

	gloss	Proto-Slavic	Russian	Polish
1SG	'I write'	* <i>pišŏ</i>	<i>pišu</i>	<i>piszę</i>
2SG	'you write'	* <i>pišešŭ</i>	<i>pišeš'</i>	<i>piszesz</i>
3SG	's/he writes'	* <i>pišetŭ</i>	<i>pišet</i>	<i>pisze-Ø</i>

(18) Number (§4.2)

English singular of nouns became zero by special phonological reduction: Old English *dæg / dagas* (> Modern English *day/days*)
< Proto-Germanic **dag-z/*dag-ōs* (cf. Gothic *dags* 'day', *dagos* 'days')

(19) Reflexives (§4.3)

Russian reduced reflexive pronoun: *-sja*, apparently derived by special phonological reduction from full reflexive pronoun *sebjā* (at the Proto-Slavic stage or even earlier)

(20) Alienability (§4.5)

a. Old Italian Latin

<i>mogliā-ma</i>	< <i>mulier mea</i>	'my wife' (inalienable)
<i>fratēl-to</i>	< <i>fratellus tuus</i>	'your brother' (inalienable)
* <i>terra-ma</i>	(cf. <i>terra mea</i>)	'my land' (alienable)

b. Nyulnyul (Nyulnyulan; northern Australia; McGregor 1996)

<i>jan yil</i>	vs.	<i>nga-lirr</i>	(< <i>ngay lirr</i>)
I.OBL dog		1SG-mouth	I mouth
'my dog' (alienable)		'my mouth' (inalienable)	

(21) Complement clauses of 'want' (§4.7)

English same-subject *wanna*, contrasting with different-subject *want to* (*The reason I wanna come is Anna* vs. *The guest I want to come is Anna.*)

There may also be cases of differential phonological reduction of nominatives (§4.4), but it is probably very difficult to find examples of phonological reduction leading to the most of the other asymmetries. Zipf's diachronic mechanism of phonological reduction is thus less important in explaining grammatical asymmetries than one might have thought.

6.2. Differential expansion/inhibition of a new construction

Most cases of economical coding are not due to differential phonological reduction, but to differential expansion of a new, more complex construction (often called "periphrasis"). Such novel constructions typically make an existing meaning more transparent by including a special additional morpheme, and they are introduced when speakers want to call special attention to the relevant meaning, in particular when they want to express themselves in an especially clear way (e.g. in situations of potential ambiguity). That is, initially these constructions are confined to highly

Neogrammarian-inspired literature. However, the alternative account in terms of frequency-induced reduction strikes me as much more plausible, especially for (17).

unusual circumstances. A novel construction may then expand and become more frequent in an increasing number of new contexts, but it will be prevented from spreading to the contexts in which the relevant meanings occur most often. Such "inhibition of expansion" may occur for two reasons:

(i) One reason is that the most frequently occurring combinations of meanings are the most deeply entrenched in the speakers' mental grammars and are thus unlikely to be replaced by innovations. Here we see the conserving effect of usage frequency.

(ii) But another reason, present in all cases of complementary expected associations of §4, is the redundancy coming from the hearers' expectations: the expression of the meaning in question is redundant when it is typically associated with another meaning. Speakers know that hearers can predict the meaning that they want to express, so they are likely to economize and not use the novel, more explicit pattern.

Let us look at a few examples to see how this works concretely.

6.2.1. Alienability splits (§4.5)

In Classical Arabic, all nouns can take possessive affixes:

- (22) *yad* 'hand' *kitaab* 'book'
 yad-ii 'my hand' *kitaab-ii* 'my book', etc.

In Maltese, only inalienable nouns (body part terms/kinship terms) take possessive affixes; others occur in a periphrastic construction with *tiegħ-* 'of':

- (23) *id* 'hand' *ktieb* 'book'
 id-i 'my hand' **ktieb-i* 'my book'
 il-ktieb tiegħ-i (originally: 'the book my-possession')

The novel construction involving the possessive noun did not expand to inalienable nouns: Maltese does not allow **l-id tiegħ-i* 'my hand'. There are two reasons for this: (i) the possessive suffixes are more entrenched in inalienable nouns and hence more resistant to loss; (ii) with inalienable nouns, the possessive semantics is more predictable and hence the new explicit construction is redundant.

Dahl and Koptjevskaja-Tamm (1998) make the strong claim that differential expansion is the only way in which an inalienability split can arise: "We suggest the generalization that an expanding possessive construction must encroach on the territory of pronominal possession for an alienability split to arise".⁸ However, we saw in (20a-b) above two examples in which such a split apparently came about by differential phonological reduction, so this hypothesis cannot be maintained.

6.2.2. Differential Object Marking (§4.4)

In Spanish, the preposition *a* was introduced at some point to mark patients (direct objects) in transitive clauses (this occurred by meaning change of the preposition, which formerly only expressed direction and various "dative" meanings). Initially the patient-marking *a* was limited to a few pronouns and nouns, but gradually it spread to all specific human NPs and became

⁸ This view is also found in Johanna Nichols's response paper in this volume.

obligatory (*Veo a Juan* 'I see Juan'). However, it did not spread further than that: inanimate NPs do not take this preposition when occurring as patients (**Veo a la iglesia* 'I see the church'), i.e. the expansion of *a* was inhibited in the context in which patients/non-human NPs most typically occur. This was presumably (i) because verb-nonhuman NP combinations were more frequent and hence more resistant to innovation, and (ii) because the patient role is more predictable with nonhuman NPs, so that speakers could afford not to use a specific marker in this environment.

6.2.3. Sex-specific affixes (§4.10)

Imagine a language with no sex-indicating markers, perhaps Old Hungarian. In this language, an occupational term such as *orvos* 'doctor' can refer to males and females alike. At some stage a compounding strategy is introduced to specify the person's sex, e.g. *orvos-nő* 'female doctor' (*nő* 'woman'), perhaps also *orvos-férfi* 'male doctor' (*férfi* 'man'). Again, there are two potential reasons why only the *-nő*-compounds spread in the language (and in fact *-nő*-suffixation is now the main way of deriving sex-specific occupational terms). (i) Occupational terms with male meaning are highly entrenched and hence unlikely to be replaced by a new strategy; (ii) For most occupations, the information that the person is male is redundant and only the less likely meaning combination needs to be indicated by speakers.

6.2.4. Recipient-theme inverse (§5.2)

The French pattern in *me le (donne)* '(gives) it to me' is a direct continuation of the Latin *mihi illum*, i.e. it represents the old pattern. The roundabout construction involving the preposition *à* (*me donne à lui* 'gives me to him') is the innovated construction, which is now used with all full-NP recipients, but its expansion stopped before reaching the most frequent patterns: pronoun combinations showing the "direct" alignment of the person and role scales (Haspelmath 2004a). Again, two reasons can be given for this: (i) the pronoun combinations were deeply entrenched and hence more resistant to change, and (ii) the meanings of the frequent pronoun combinations were easier to predict than those of the rare combinations.

It seems that in different cases, the relevance of these two reasons for differential inhibition is different. Thus, the conserving effect of frequency seems to be very relevant for the inalienability split (§6.2.1), but less so for differential object marking (§6.2.2), and hardly at all for sex-specific affixes (§6.2.3). At present, I know of no way of assessing the respective roles of the two factors, and I leave this as an intriguing question for future research.

6.3. Excluded: differential morphosyntactic reduction

A priori, one could easily imagine another way in which economical patterns are created in language change: by **differential morphosyntactic reduction**. This would mean the omission of a particular morpheme or combination of morphemes which were originally present, but came to be omitted in a more frequent construction. However, I know of no evidence that such a change has ever occurred, and one may hypothesize that it is impossible.

Consider a concrete example whose history is not known (at least to me): In Ewe (Kwa, Niger-Congo), alienable possession is indicated by simple

juxtaposition, whereas alienable possession is indicated by a postposition $\varphi\acute{e}$ (Ameka 1996: 791, 797):

(24) Ewe	<i>kofí srɔ̂</i> Kofi spouse 'Kofi's wife'	vs.	<i>kofí $\varphi\acute{e}$ awu</i> Kofi POSS garment 'Kofi's garment'
----------	--	-----	--

If differential morphosyntactic reduction is impossible, we know that the present situation does not come from an earlier pattern where $\varphi\acute{e}$ was present in both types of possession (*kofí $\varphi\acute{e}$ srɔ̂ / kofí $\varphi\acute{e}$ awu*), and it was simply dropped from the construction where it was more redundant.

However, there are two rather serious problems with this proposed universal of diachronic change: (i) In practice, it is often difficult to exclude the possibility that the change was phonologically motivated and conditioned, (ii) analogical change can create economical patterns, too (see next subsection), and we must make sure that this is taken into account in formulating the diachronic universal. Moreover, a reviewer mentions a possible counterexample: The change in many English varieties from *I've got* to *I got* (in the sense 'I have').⁹ I do not know whether this could be explained by phonological reduction, so at the very least it illustrates the difficulty of assessing the validity of the claim that differential morphosyntactic reduction does not occur.

6.4. Analogical change can create economical patterns

An economical case-marking system may arise by **selectively preserving older markers**. For example, in the Old High German *n*-declension, animate and inanimate nouns alike had a distinction between nominative and accusative (cf. 25).

(25)	Old High German	Modern German
NOM.SG	<i>affo knoto</i>	<i>Affe Knoten</i>
ACC.SG	<i>affon knoton</i>	<i>Affen Knoten</i>
	'ape' 'knot'	'ape' 'knot'

Then the nominative-accusative distinction was lost in inanimate nouns, and in Modern German only animates preserve the zero-marking in the nominative. The resulting pattern shows differential object marking and thus economical coding, but it has come into existence via a different diachronic route (cf. Haspelmath 2002:245).

Another example comes from the history of French. Old French had the case- and number-marking pattern shown in (26).

(26) Old French	SG	PL
NOM	<i>murs</i>	<i>mur</i>
ACC	<i>mur</i>	<i>murs</i>

⁹ Another potential counterexample mentioned by a reviewer is of a different nature: The frequent absence of the English complementizer *that* with the most frequent verbs that take *that*-complements (*say, know*, etc.). However, it is not clear that there was a stage at which complementizerless complement clauses with such verbs were impossible.

In later French, the nominative-accusative distinction was given up, and those forms from both numbers were selectively preserved that lead to an economical pattern (SG *mur*, PL *murs*) (see Mayerthaler 1981: ch. 4 for some relevant discussion). Further related cases of analogical change are discussed in Bybee (1985:54-56). (The approaches to analogy in Albright (this volume) and Garrett (this volume) do not seem to allow for the possibility that analogy may be motivated by the desire for an economical output; one wonders how they would deal with examples like those mentioned here.)

6.5. Is frequency really a causal factor?

A reviewer objected to the claim made here that frequency asymmetries can explain morphosyntactic asymmetries. Since this is an objection that I hear often and that still seems to be shared by many linguists, I quote the skeptical reviewer here:

"It seems to me more plausible that both the frequency and asymmetries are results, caused by something else. We say *male nurse* and (less often now) *lady doctor* because of cultural expectations—expectations that lead to *nurse* and *doctor* being used unmodified for females and males respectively. I can't see how frequency is in any sense the cause of the linguistic asymmetry."

The only problem with the idea that both frequency of use and the morphosyntactic asymmetries result from something else is that this something else cannot be identified. The only candidate that has been proposed, as far as I am aware, is "markedness", which is sometimes said to be responsible for frequency distributions (e.g. Mayerthaler 1981:136-140). However, how "markedness" should influence frequency of use remains obscure, and it is relatively easy to show that on the contrary, "markedness" effects of various sorts must be derived from frequency of use, so that the entire "markedness" concept can be abandoned (Haspelmath 2006).

Of course, the *male nurse/lady doctor* asymmetry is related to cultural expectations, but these imply linguistic expectations: Since nurses tend to be women, words for female nurses occur more often than words for male nurses. As a result, saying *lady nurse* would be redundant and is avoided by speakers. One might be tempted to derive the redundancy directly from the cultural expectations, and in this particular case this would probably lead to the same results (though cultural expectations, or the real-world frequencies they presumably derive from, are much harder to measure than linguistic frequencies). But crucially, linguistic frequencies do not necessarily match real-world frequencies or non-linguistic expectations. Singulars are more frequent than plurals, though the world is hardly populated by more individuals than groups, and it would be hard to argue that there is a non-linguistic "expectation" that countable entities should occur singly. For whatever reason, people talk more about single entities than about pluralities, and this suffices to explain the facts of language. The important general point is that while linguistic frequency can usually be derived from other factors, these factors are quite heterogeneous. What matters to us grammarians is that the results of frequency are homogeneous, so that we can focus on linking frequency of use to economical patterns, leaving the extra-linguistic causes of frequency aside.

The skeptical reviewer also brings up the possibility of the direction of causation being the reverse: The availability of short coding could make speakers more likely to say a particular thing. In general this does not seem to be the case: Languages that have symmetrical coding patterns usually show the same kinds of frequency asymmetries as languages with asymmetrical coding. Thus,

- 'want' complements tend to be same-subject not only in asymmetric Italian, but also in symmetric Greek (see §4.7);
- possessed NPs are preferentially definite not only in asymmetric English, but also in symmetric Italian (§4.8, Haspelmath 1999a);
- active/present and passive/past participles are preferred not only in asymmetric English, but also in symmetric Lezgian (§4.9, Haspelmath 1994);
- automatic events like 'freeze' tend to be more rarely used transitively not only in asymmetric Japanese, but also in symmetric English (§4.11);
- inverse recipient/theme constructions ('show me to him') are rarer than direct constructions ('show him to me') not only in asymmetric French, but also in symmetric German (§5.2, Haspelmath 2004a).

I conclude that frequency of use really is the relevant causal factor, and the reviewer's skepticism is unjustified.

7. Conclusion: The relation between diachrony and language universals

In this paper, I have made five main points:

(i) A very large number of morphosyntactic implicational universals can be explained by invoking economic motivation (Haiman 1983)¹⁰: more frequently used expressions are shorter than semantically similar, but more rarely used expressions, because they are more predictable.

(ii) Apparently all universal morphosyntactic asymmetries (in the sense of §2) are economically motivated (see (1) above). This is a meta-universal, a universal about the explanation of universals.

(iii) Economical patterns are created by speakers in language use, and when innovative patterns spread through the community, they are manifested in the results of language change.

(iv) There are at least three different diachronic paths through which economical patterns arise: differential phonological reduction (§6.1), differential expansion of a new construction (§6.2), selective analogical change (§6.4).

(v) One obvious possible diachronic path does not seem to be well attested: differential morphosyntactic reduction (§6.3). However, this generalization is problematic, because morphosyntactic reduction is not easy to differentiate from phonological reduction, and counterexamples have been noted.

If for the sake of the discussion we consider (27) as the most basic question of this volume,

- (27) Do synchronic universals arise from universals of change, or do universals of change arise from synchronic constraints?

¹⁰ Economic motivation is also the relevant factor for many of the phenomena that Haiman (1983) and others have attributed to "iconic motivation", as is shown in Haspelmath (2007+c).

the answer given by the present paper is an intermediate one. While I do not deny that some synchronic universals derive from synchronic universals of the cognitive code (e.g. perhaps the generalization that grammatical rules do not include numerical specifications), the implicational universals considered here arise in language change. If we created an artificial language violating some of the universals, it would be learnable without problems, but it would be predicted that after a few centuries the language would give in to the pressure to change to more normal, economical patterns. In this regard, I am thus in agreement with Bybee (1988, this volume), Hopper (this volume), Blevins (this volume), and Garrett (this volume), who emphasize the importance of diachronic change for understanding synchronic universals.

However, I would argue that many diachronic changes cannot be understood without taking into account the result they lead to. It is not an accident that differential phonological reduction, differential expansion of a new construction, and selective analogical change may all lead to synchronic economical patterns (rather than, say, to counter-economical ones). Rather, the changes are motivated by economy, i.e. by the innovating speakers' desire to speak economically. This is not a synchronic grammatical constraint, or even a more general cognitive constraint. It is simply a constraint on any rational behaviour.

Thus, with regard to the phenomena considered in this paper, the answer to (27) is: both. Implicational universals of the sort exemplified in §4-5 arise from universal tendencies of change toward economical patterns, and these universal diachronic tendencies themselves are motivated by the (synchronic) constraint that speech behaviour should be rational and take both speakers' and hearers' needs into account.

References

- Aissen, Judith (2003). 'Differential object marking: Iconicity vs. economy.' *Natural Language and Linguistic Theory* 21: 435-483.
- Albright, Adam (This volume). 'Explaining universal tendencies and language particulars in analogical change.'
- Ameka, Felix (1996). 'Body parts in Ewe grammar', in Hilary Chappell and William McGregor (eds.) (1996). *The grammar of inalienability*. Berlin: Mouton de Gruyter, 783-840.
- Blansitt, Edward L. Jr. (1973). 'Bitransitive clauses.' *Working Papers in Language Universals (Stanford)* 13: 1-26.
- Blevins, Juliette (This volume). 'Consonant epenthesis: natural and unnatural histories.'
- Bossong, Georg (1985). *Differenzielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Narr.
- Bossong, Georg (1998). 'Le marquage différentiel de l'objet dans les langues d'Europe', in Jack Feuillet (ed.) *Actance et valence dans les langues de l'Europe*. Berlin: Mouton de Gruyter, 193-258.
- Bybee, Joan L. (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam: Benjamins.
- Bybee, Joan L. (1988). 'The diachronic dimension in explanation', in John A. Hawkins (ed.) *Explaining language universals*. Oxford: Blackwell, 350-379.

- Bybee, Joan L. (2003). 'Mechanisms of change in grammaticization: The role of frequency', in Richard Janda and Brian Joseph (eds.) *Handbook of historical linguistics*. Oxford: Blackwell, 602-623.
- Bybee, Joan L. (This volume). 'Substantive and formal universals as emergent phenomena: the origins of Structure Preservation.'
- Comrie, Bernard (1981). 'Aspect and voice: some reflections on perfect and passive', in Philip J. Tedeschi and Annie Zaenen (eds.) *Tense and Aspect*, ed. by, 65-78. (Syntax and Semantics, 14) New York: Academic Press.
- Comrie, Bernard (1989). *Language Universals and Linguistic Typology*. 2nd edn. Oxford: Blackwell.
- Croft, William (1990). *Typology and universals*. Cambridge: Cambridge University Press.
- Croft, William (1991). *Syntactic categories and grammatical relations: the cognitive organization of information*. Chicago: The University of Chicago Press.
- Croft, William (2003). *Typology and universals*. 2nd edn. Cambridge: Cambridge University Press.
- Dahl, Östen and Maria Koptjevskaja-Tamm (1998). 'Alienability splits and the grammaticalization of possessive constructions', in Timo Haukioja (ed.) *Papers from the 16th Scandinavian Conference of Linguistics*. Turku: University of Turku, 38-49.
- Dayley, Jon P. (1985). *Tzutujil grammar*. (University of California Publications in Linguistics, 107.) Berkeley: University of California Press.
- DeLancey, Scott (1981). 'An interpretation of split ergativity and related patterns.' *Language* 57: 626-57.
- Dixon, R.M.W. (1994). *Ergativity*. Cambridge: Cambridge University Press.
- Du Bois, John (1985). 'Competing motivations', in John Haiman (ed.) *Iconicity in syntax*. Amsterdam: Benjamins, 343-365.
- Durie, Mark, Bukhari Daud and Mawardi Hasan (1994). 'Acehnese', in Cliff Goddard and Anna Wierzbicka (eds.) *Semantic and lexical universals: theory and empirical findings*. Amsterdam: Benjamins, 171-201.
- Faltz, Leonard M. (1985). *Reflexivization: A Study in Universal Syntax*. New York: Garland.
- Garrett, Andrew (This volume). 'Markedness and universals in the evolution of paradigm leveling.'
- Greenberg, Joseph H. (1963). 'Some universals of grammar with particular reference to the order of meaningful elements', in Joseph H. Greenberg (eds.) *Universals of grammar*, 73-113. Cambridge, Mass.: MIT Press.
- Greenberg, Joseph H. (1966). *Language universals, with special reference to feature hierarchies*. The Hague: Mouton.
- Haiman, John (1983). 'Iconic and economic motivation.' *Language* 59: 781-819.
- Haspelmath, Martin (1993a). *A grammar of Lezgian*. (Mouton Grammar Library, 9.) Berlin: Mouton de Gruyter.
- Haspelmath, Martin (1993b). 'More on the typology of inchoative/causative verb alternations', in Bernard Comrie and Maria Polinsky (eds.) *Causatives and transitivity*. (Studies in Language Companion Series, 23.) Amsterdam: Benjamins, 87-120.
- Haspelmath, Martin (1994). 'Passive participles across languages'. In: Barbara Fox and Paul J. Hopper (eds.) *Voice: Form and Function*. (Typological Studies in Language, 27.) Amsterdam: Benjamins, 151-177.
- Haspelmath, Martin (1999a). 'Explaining article-possessor complementarity: economic motivation in noun phrase syntax'. *Language* 75.2: 227-43.

- Haspelmath, Martin (1999b). 'Optimality and diachronic adaptation.' *Zeitschrift für Sprachwissenschaft* 18.2: 180-205.
- Haspelmath, Martin (1999c). 'On the cross-linguistic distribution of same-subject and different-subject complement clauses: Economic vs. iconic motivation.' Paper presented at the ICLC, Stockholm, July 1999.
- Haspelmath, Martin (2002). *Understanding morphology*. London: Arnold.
- Haspelmath, Martin (2004a). 'Explaining the Ditransitive Person-Role Constraint: a usage-based account.' *Constructions* 2/2004, 49 pp. (free online journal, University of Düsseldorf)
- Haspelmath, Martin (2004b). 'Does linguistic explanation presuppose linguistic description?' *Studies in Language* 28.3: 554-579.
- Haspelmath, Martin (2006). 'Against markedness (and what to replace it with).' *Journal of Linguistics* 42.1: 1-46.
- Haspelmath, Martin (2007+a). 'A frequentist explanation of some universals of reflexive marking'. Max Planck Institute for Evolutionary Anthropology, <http://email.eva.mpg.de/~haspelmt/papers.html>.
- Haspelmath, Martin (2007+b). 'Parametric versus functional explanations of syntactic universals.' To appear in a volume on parameters edited by Theresa Biberauer.
- Haspelmath, Martin (2007+c). 'Frequency vs. iconicity in explaining grammatical asymmetries.' Submitted.
- Hawkins, John A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Hopper, Paul (This volume). 'Emergent serialization in English: pragmatics and typology.'
- Jäger, Gerhard (2004). 'Learning constraint sub-hierarchies: The Bidirectional Gradual Learning Algorithm.' In Reinhard Blutner and Henk Zeevat (eds.), *Pragmatics in OT*. Palgrave MacMillan, 251-287.
- Keenan, Edward L. and Comrie, Bernard (1977). 'Noun phrase accessibility and universal grammar.' *Linguistic Inquiry* 8: 63-99.
- Kemmer, Suzanne (1993). *The middle voice*. Amsterdam: Benjamins.
- Kiparsky, Paul (This volume). 'Universals constrain change; change results in typological generalizations.'
- Kirby, Simon (1999). *Function, selection, and innateness: The emergence of language universals*. Oxford: Oxford University Press.
- König, Ekkehard and Siemund, Peter (1999). 'Intensifiers and reflexives: a typological perspective', in Zygmunt Frajzyngier and Traci S. Curl (eds.) *Reflexives: forms and functions* (Typological studies in language, 40) Amsterdam: Benjamins, 41-74.
- Koptjevskaja-Tamm, Maria (1996). 'Possessive noun phrases in Maltese: Alienability, iconicity, and grammaticalization.' *Rivista di Linguistica* 8.1: 245-274.
- Lazard, Gilbert (2001). 'Le marquage différentiel de l'objet', in Martin Haspelmath and Ekkehard König and Wulf Oesterreicher and Wolfgang Raible (eds.) *Language typology and language universals: An international handbook* Vol. 2. Berlin: de Gruyter, 873-885.
- Leech, Geoffrey, Paul Rayson, and Andrew Wilson (2001). *Word frequencies in written and spoken English based on the British National Corpus*. Harlow, England: Pearson Education.
- Lefebvre, Claire and Anne-Marie Brousseau (2002). *A grammar of Fongbe*. Berlin: Mouton de Gruyter.

- Lehmann, Christian. (1982 (1995)). *Thoughts on grammaticalization*. Munich: Lincom.
- Mańczak, Witold (1987). *Frequenzbedingter unregelmäßiger Lautwandel in den germanischen Sprachen*. Wrocław: Ossolineum.
- Mayerthaler, Willi (1981). *Morphologische Natürlichkeit*. Wiesbaden: Athenaion. (translated as Mayerthaler 1988)
- Mayerthaler, Willi (1988). *Naturalness in morphology*. Ann Arbor: Karoma.
- McGregor, William (1996). 'The grammar of nominal prefixing in Nyulnyul', in Hilary Chappell and William McGregor (eds.) 1996. *The grammar of inalienability*. Berlin: Mouton de Gruyter, 251-292.
- Newmeyer, Frederick J. (2004). 'Against a parameter-setting approach to language variation.' *Linguistic Variation Yearbook* 4: 181-234.
- Newmeyer, Frederick J. (2005). *Possible and probable languages: a generative perspective on linguistic typology*. Oxford: Oxford University Press.
- Nichols, Johanna (1988). 'On alienable and inalienable possession', in William Shipley (ed.) *In honor of Mary Haas*. Berlin: Mouton de Gruyter, 475-521.
- Reh, Mechthild (1985). *Die Krongo-Sprache (niino mó-di). Beschreibung, Texte, Wörterverzeichnis*. (Kölner Beiträge zur Afrikanistik, 12.) Berlin: Reimer.
- Siewierska, Anna. 2005. 'Third person zero of verbal person marking.' In: *WALS*, 418-421.
- Silverstein, Michael (1976). 'Hierarchy of features and ergativity', in R.M.W. Dixon (ed.) *Grammatical categories in Australian languages*. Canberra: Australian Institute of Aboriginal Studies, 112-71.
- Thompson, Chad (1996). 'On the grammar of body parts in Koyukon Athabaskan', in Hilary Chappell and William McGregor (eds.) 1996. *The grammar of inalienability*. Berlin: Mouton de Gruyter, 651-676.
- Tiersma, Peter (1982). 'Local and general markedness.' *Language* 58: 832-49.
- WALS*: Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press.
- Wright, Sandra Kimberly (2001). *Internally caused and externally caused change of state verbs*. Ph.D. dissertation, Northwestern University (Evanston, IL).
- Xrakovskij, Viktor S. (ed.) (2001). *Typology of imperative constructions*. Munich: Lincom Europa.
- Zipf, George K. (1935). *The psycho-biology of language: an introduction to dynamic philology*. Houghton Mifflin (Republished 1965 by MIT Press.)
- Zipf, George K. (1949). *Human behavior and the principle of least effort: an introduction to human ecology*. Cambridge, MA: Addison-Wesley.
- Zúñiga, Fernando (2006). *Deixis and alignment: Inverse systems in indigenous languages of the Americas*. (Typological Studies in Language, 70.) Amsterdam: Benjamins.