Frequency vs. iconicity in explaining grammatical asymmetries

MARTIN HASPELMATH*

Abstract

This paper argues that three widely accepted motivating factors subsumed under the broad heading of iconicity, namely iconicity of quantity, iconicity of complexity and iconicity of cohesion, in fact have no role in explaining grammatical asymmetries and should be discarded. The iconicity accounts of the relevant phenomena have been proposed by authorities like Jakobson, Haiman and Givón, but I argue that these linguists did not sufficiently consider alternative usage-based explanations in terms of frequency of use. A closer look shows that the well-known Zipfian effects of frequency of use (leading to shortness and fusion) can be made responsible for all of the alleged iconicity effects, and initial corpus data for a range of phenomena confirm the correctness of the approach.

Keywords: frequency; iconicity; markedness; economic motivation

1. Introduction

The notion of *iconicity* has become very popular in the last 25 years among functional and cognitive linguists. In Croft's (2003: 102) words, "the intuition behind iconicity is that the structure of language reflects in some way the structure of experience". Iconicity is thus a very broad notion, and it has been understood and applied in a great variety of ways (see Newmeyer 1992: §§2–3 for an attempt at a survey). In this paper, I will examine just the three sub-types of (diagrammatic)[1] iconicity in (1)–(3), which have played an important role in discussions of grammatical asymmetries. I will argue that in fact none of these is relevant for explaining grammatical asymmetries, and that the phenomena in question should instead be explained by asymmetries of frequency of occurrence.

(1) Iconicity of quantity
     "Greater quantities in meaning are expressed by greater quantities
     of form."
     Example: In Latin adjective inflection, the comparative and super-
     lative denote increasingly higher degrees and are coded by increas-
     ingly longer suffixes (e.g., *long(-us)* 'long', *long-ior* 'longer', *long-
     issim(-us)* 'longest').

(2) Iconicity of complexity
     "More complex meanings are expressed by more complex forms."
     Example: Causatives are more complex semantically than the corre-
     sponding non-causatives, so they are coded by more complex forms,
     e.g., Turkish *düş(-mek)* 'fall', causative *düş-ür(-mek)* 'make fall,
     drop'.

(3) Iconicity of cohesion
     "Meanings that belong together more closely semantically are ex-
     pressed by more cohesive forms."
     Example: In possessive noun phrases with body-part terms, the
     possessum and the possessor are conceptually inseparable. This is
     mirrored in greater cohesion of coding in many languages, e.g., Mal-
     tese *id* 'hand', *id-i* 'my hand', contrasting with *siġġu* 'chair', *is-siġġu
     tiegħ-i* [the-chair of-me] 'my chair' (*siġġ(u)-i*).

While iconicity of quantity is mentioned rarely, iconicity of complexity
and iconicity of cohesion are often invoked in the functional and cogni-
tive literature (and recently to some extent also in the generative litera-
ture; see §4.5). Both have been applied to a wide range of grammatical
phenomena by many different authors.

   I argue in this paper that these three types of iconicity play no role
in explaining grammatical asymmetries of the type *long(-us)/long-ior*,
düş(-mek)/düş-ür(-mek), *id-i/siġġu tiegħ-i*. Instead, such formal asym-
metries can and should be explained by frequency asymmetries: In all
these cases, the shorter and more cohesive expression types occur signifi-
cantly more frequently than the longer and less cohesive expression types,
and this suffices to explain their formal properties. No appeal to iconicity
is necessary. Worse, iconicity often makes wrong predictions, whereas fre-
quency consistently makes the correct predictions.

   I want to emphasize that I make no claims about other types of iconic-
ity, such as

   – iconicity of paradigmatic isomorphism (one form, one meaning in the
system, i.e., synonymy and homonymy are avoided; Haiman 1980; Croft
1990a: 165, 2003: 105);

– iconicity of syntagmatic isomorphism (one form, one meaning in the string, i.e., empty, zero and portmanteau morphs are avoided; Croft 1990a: 165, 2003: 103);[2]

– iconicity of sequence (sequence of forms matches sequence of experiences; e.g., Greenberg 1963 [1966: 103]);

– iconicity of contiguity (forms that belong together semantically occur next to each other; this is similar to iconicity of cohesion, but different in crucial ways, cf. §5);

– iconicity of repetition (repeated forms signal repetition in experience, as when reduplication expresses plurality or distribution).

For most of these iconicity types, frequency is clearly not a relevant factor, and I have no reason to doubt the conventional view that the relevant phenomena are motivated by functional factors that can be conveniently subsumed under the label *iconicity*. Whether these functional factors can be reduced to a general preference for iconic over noniconic patterns is a separate question that I will not pursue here.

I also need to emphasize that I am interested in explanation of grammatical structures, perhaps more so than many other authors that have discussed iconicity. That is, I want to know WHY language structure is the way it is, whereas some authors seem to be content with observing that language structure is sometimes iconic:

The traditional view of language is that most relationships between linguistic units and the corresponding meanings are arbitrary ... But the cognitive claim is that the degree of iconicity in language is much higher than has traditionally been thought to be the case. (Lee 2001: 77)

As long as one merely observes that cases like *long(-us)/long-ior* and *düş(-mek)/düş-ür(-mek)* can be regarded as iconic in some way, I have no problem. What I am denying is that iconicity plays a motivating role and should be invoked in explaining why the patterns are the way they are.

What I observed while reading the literature on iconicity is that a number of authors (e.g., Hockett 1958: 577–578; Givón 1985, 1991) seem to use the term "iconicity" as a kind of antonym of "arbitrariness", so that almost anything about language structure that is not arbitrary falls under iconicity. I am in broad sympathy with Givón's general account of the relation between arbitrariness and non-arbitrariness in language, but I would insist on the need to identify the relevant factors as precisely as possible and to make testable predictions. It is quite possible that the disagreements about the role of frequency vs. iconicity will eventually turn out to be less severe than it may seem at the beginning, but in any event this paper should help to clarify the issues.

Iconicity and the frequency asymmetries discussed here are universal explanatory factors, so their effects should be universal. This means that in principle confirming data could come from any language, and ideally the data should come from a large representative sample of languages. Such data are still not very widely available, so this paper will continue the practice of Haiman (1983) (and much other work) of making claims about universal asymmetries that are not fully backed up by confirming data, but that nevertheless seem very plausible because of the apparent absence of counterevidence. Likewise, disconfirming data could come from any language, but of course isolated counterexamples are not sufficient to show that no systematic coding asymmetry exists. Many of the generalizations cited here are known to be merely strong tendencies, not absolute universals.

The remainder of this paper is organized as follows: §2 discusses iconicity of quantity, §§3–4 discuss iconicity of complexity, and §§5–6 discuss iconicity of cohesion. For each subtype of iconicity, I will first cite authors who have advocated it and mention examples of phenomena that are allegedly motivated by iconicity, before presenting my arguments for a frequency-based explanation of the phenomena. The final §7 presents the conclusions.

2.   Iconicity of quantity

2.1.   *Advocates and examples*

Iconicity of quantity was defined in §1 as follows:

(4)   Greater quantities in meaning are expressed by greater quantities of form.

It seems that the first author to mention this motivating principle was Jakobson (1965[1971: 352]) and (1971). Jakobson cited three examples:

(i) In many languages, "the positive, comparative and superlative degrees of adjectives show a gradual increase in the number of phonemes, e.g., *high-higher-highest*, [Latin] *altus, altior, altissimus*. In this way, the signantia reflect the gradation gamut of the signata" (1965[1971: 352]). The higher the degree, the longer the adjective.

(ii) "The signans of the plural tends to echo the meaning of a numeral increment by an increased length of the form" (1965[1971: 352]). The more referents, the more phonemes (e.g., singular *book*, plural *books*, French singular *je finis* 'I finish', plural *nous finissons* 'we finish').

(iii) In Russian, the perfective aspect expresses "a limitation in the extent of the narrated event", and it is expressed by a more limited (i.e.,

a smaller) number of phonemes (e.g., perfective *zamoroz-it*, imperfective *zamoraž-ivat* 'freeze') (Jakobson 1971).

Iconicity of quantity is mentioned approvingly in Plank (1979: 123), Haiman (1980: 528–529, 1985: 5), Anttila (1989: 17), in Taylor's (2002: 46) Cognitive Grammar textbook, and in Itkonen (2004: 28); see also Lakoff and Johnson (1980: 127).

2.2.   *Frequency-based explanation*

Any efficient sign system in which costs correlate with signal length will follow the following economy principle:[3]

(5)   The more predictable a sign is, the shorter it is.

Since frequency implies predictability, we also get the following prediction for efficient sign systems:

(6)   The more frequent a sign is, the shorter it is.

These principles have been well known at least since Horn's (1921) and Zipf's (1935) work, but somehow under the influence of the structuralist movements many linguists lost sight of them for a few decades. However, more recently cognitively oriented linguists have begun to appreciate the importance of frequency again (e.g., Bybee and Hopper 2001, among many others). I do not claim to have original insights about the way in which frequency influences grammatical structures, but I want to argue that iconicity turns out to be less important as an explanatory concept if one gives frequency the explanatory role that it deserves.

Principle (6) straightforwardly explains Jakobson's observations about adjectival degree marking and singular/plural asymmetries, because universally comparative and superlative forms are significantly rarer than positive forms of adjectives, and plural forms are significantly rarer than singular forms (see Greenberg 1966: 34–37, 40–41). It is not possible to make such a universal statement about perfective and imperfective aspect, and the frequency of these aspectual categories depends much more on the lexical meaning of the individual verb. But for Russian, Fenk-Oczlon (1990) has shown that there is a strong correlation between length and frequency of a verb form: in general, the more frequent member of a Russian aspectual pair is also shorter.

This frequency-based explanation is not only sufficient to account for the phenomena cited by Jakobson, but also necessary, because the principle of iconicity of quantity makes many wrong predictions (as was also observed by Haiman 2000: 287). For example, it predicts that plurals should generally be longer than duals, that augmentatives should

generally be longer than diminutives, that words for 'ten' should be longer than words for 'seven', or even that words for 'long' should be longer than words for 'short', or that words for 'elephant' should be longer than words for 'mouse'. None of these predictions are generally correct (except perhaps for the last prediction, but note that *mouse* is about twice as frequent as *elephant* in English).[4]

Iconicity of quantity has never been considered particularly important, and its refutation here is only a prelude to the refutation of the other two kinds of iconicity in §§3–6.

3.   Iconicity of complexity: Advocates and examples

Iconicity of complexity was defined in §1 as follows:

(7)   More complex meanings are expressed by more complex forms.

Here are some quotations from the literature that describe this principle and refer to it as ''isomorphic'' or ''iconic''.

– Lehmann (1974: 111): ''Je komplexer die semantische Repräsentation eines Zeichens, desto komplexer seine phonologische Repräsentation.'' ('The more complex the semantic representation of a sign is, the more complex is its phonological representation.')
– Mayerthaler (1981: 25): ''Was semantisch ''mehr'' ist, sollte auch konstruktionell ''mehr'' sein.'' ('What is ''more'' semantically should also be ''more'' constructionally.')
– Givón (1991: §2.2): ''A larger chunk of information will be given a larger chunk of code.''
– Haiman (2000: 283): ''The more abstract the concept, the more reduced its morphological expression will tend to be. Morphological bulk corresponds directly and iconically to conceptual intension.''
– Langacker (2000: 77): ''[I]t is worth noting an iconicity between *of*'s phonological value and the meaning ascribed to it (cf. Haiman 1983). Of all the English prepositions, *of* is phonologically the weakest by any reasonable criterion.... Now as one facet of its iconicity, *of* is arguably the most tenuous of the English prepositions from the semantic standpoint as well ...''

In Lehmann's (1974) approach, semantic complexity is measured by counting the number of features needed to describe the meaning of an expression. A contrast between presence and absence of a semantic feature is often called ''semantic markedness'', and very often iconicity of complexity is described as a kind of ''iconicity of markedness matching'':

(8)   Marked meanings are expressed by marked forms.

This principle was already formulated by Jakobson (1963[1966: 270]), and repeated many times in the later literature, e.g.,

–   Plank (1979: 139): "Die formale Markiertheitsopposition bildet die konzeptuell-semantische Markiertheitsopposition d[iagrammatisch]-ikonisch ab." ('The formal markedness opposition mirrors the conceptual-semantic markedness opposition in a diagrammatically iconic way.')
–   Haiman (1980: 528): "Categories that are marked morphologically and syntactically are also marked semantically."
–   Mayerthaler (1987: 48–9): "If (and only if) a semantically more marked category C_j is encoded as more featured [= formally complex] than a less marked category C_i, the encoding of C_j is said to be iconic."
–   Givón (1991: 106, 1995: 58): "The meta-iconic markedness principle: Categories that are *cognitively* marked—i.e., complex—tend also to be *structurally* marked."
–   Aissen (2003: 449): "Iconicity favors the morphological marking of syntactically marked configurations."

For similar statements, see also Zwicky (1978: 137), Matthews (1991: 236), Newmeyer (1992: 763), and Levinson (2000: 136–137).

By "formally marked", these authors generally mean "expressed overtly". Typical examples of such markedness matching are given in (9).

(9)	less marked/unmarked	(more) marked
number	SINGULAR (*tree-Ø*)	PLURAL (*tree-s*)
case	SUBJECT (Latin *homo-Ø*)	OBJECT (*homin-em*)
tense	PRESENT (*play-Ø*)	PAST (*play-ed*)
person	THIRD (Spanish *canta-Ø*)[5]	SECOND (*canta-s*)
gender	MASCULINE (*petit-Ø*)	FEMININE (*petit-e*)
causation	NON-CAUSATIVE	CAUSATIVE
	(Turkish *düş-Ø-mek* 'fall')	(*düş-ür-mek* 'fell, drop')
object	INANIMATE	ANIMATE
	(Spanish *Veo la casa*	*Veo a la niña.*
	'I see the house'	'I see the girl.')

That there are universal formal asymmetries in these (and many other) categories has been known since Greenberg (1966), and Jakobson (1963[1966]) and (1965[1971]) explicitly refers to Greenberg's cross-linguistic work. However, Greenberg did not invoke iconicity to explain the formal asymmetries of the kind illustrated in (9). He had good reasons, as we will see in the next section.

4.   Iconicity of complexity: frequency-based explanation

4.1.   *Complex/"marked" expressions are rarer*

Greenberg's (1966) explanation was in terms of the frequency asymmetries in the use of the grammatical forms. He noted that "less marked" forms are more frequent, and "more marked" forms are less frequent across languages. Thus, the economy principles in (5)–(6) are sufficient to explain the asymmetries in (9) (see also Croft 2003: 110–117). The English preposition *of* is not only the most "semantically tenuous" (Langacker 2000: 77), but also the most frequent of all the English prepositions. Singulars are more frequent than plurals, nominatives are more frequent than accusatives, the present tense is more frequent than the past tense, the third person is more frequent than other persons, and the masculine is more frequent than the feminine. All of this was documented by Greenberg (1966) for a few selected languages, and the hypothesis that it holds universally has not been challenged. That causatives are generally less frequent than the corresponding non-causatives is also clear; I discuss this case in more detail below (§4.4). And among objects, inanimate referents are much more frequent than animate referents (§4.5).

This frequency-based explanation is not only sufficient to account for the relevant phenomena, but also necessary, because iconicity of complexity makes some wrong predictions. In (10), I list cases that go in the opposite direction of the patterns in (9).

(10)	less marked/unmarked	(more) marked
number	PLURAL	SINGULAR
	Welsh *plu* 'feathers'	*plu-en* 'feather'
case	OBJECT CASE	SUBJECT CASE
	Godoberi *mak'i* 'child'	*mak'i-di* (ergative)
person	SECOND P. IMPERATIVE	THIRD P. IMPERATIVE
	Latin *canta-Ø* 'sing!'	*canta-to* 'let her sing'
gender	FEMALE	MALE
	English *widow-Ø*	*widow-er*
causation	CAUSATIVE	NONCAUSATIVE
	German *öffnen*	*sich öffnen*

In all these cases, frequency makes the right predictions. Plurals like Welsh *plu* 'feathers' are more frequent than singulars (Tiersma 1982), in the imperative mood the second person is more frequent than the third person, the word *widow* is more frequent than the word *widower*, and with verbs like 'open', the causative is more frequent than the noncausative (see §4.4).

These exceptions have long been known in the literature, but linguists have often described them in terms of *markedness reversal*. The idea is that markedness values can be different in different contexts, so that, for example, third person is not absolutely unmarked with respect to second person, but in certain contexts second person can be unmarked and first person can be marked (e.g., Waugh 1982; Tiersma 1982; Witkowski and Brown 1983; Haiman 1985: 148–149; Croft 1990a: 66). But in order to reconcile the cases in (10) with iconicity of complexity, one would have to show that not only the formal coding, but also the semantic/functional markedness value has changed. This is much more difficult, and it has not been shown that it is generally true that in cases of markedness reversal, the formally unmarked term of the opposition is also semantically or functionally unmarked. For example, Tiersma's (1982) main additional evidence that "locally unmarked plurals" like Welsh *plu* 'feathers' are generally unmarked (i.e., do not merely show reversed formal coding) is that in analogical leveling, the plural survives. But analogical leveling is of course just another symptom of frequency of occurrence (cf. Bybee 1985: Ch. 3).

To make matters even more complex, some authors seem to mean frequency when they say *(functional) unmarkedness*: *Marked* means 'rare', and *unmarked* means 'frequent'. For example, in a discussion of unmarked plurals, Haiman writes:

... what is fundamentally at issue is markedness. Where plurality is the norm, it is the plural which is unmarked, and a derived marked singulative is employed to signal oneness: thus, essentially, *wheat* vs. *grain of wheat*. (Haiman 2000: 287)

The "norm" is of course the same as the more frequent situation, so what is fundamentally at issue is frequency. Linguists are of course free to define their terms in whatever way they wish, but claiming not only that formally marked elements tend to be "functionally marked" (in the sense of being less frequent), but also that this a surprising instance of "markedness matching" (or iconicity), is not helpful. The much simpler observation is that formally marked elements tend to be less frequent, and this observation is straightforwardly explained by the economy principles in (5)–(6). Neither "iconicity" nor "markedness" are relevant concepts in stating and explaining these facts (see Haspelmath 2006 for detailed argumentation that a notion of markedness is superfluous in linguistics).

The contrasts in (9) show zero expression vs. overt expression, but some authors such as Lehmann (1974) and Haiman (2000) also talk

about length differences between different types of morphemes. In particular, both authors note that grammatical morphemes are universally shorter than lexical morphemes, and they claim that this iconically mirrors their more abstract or less complex meaning. But again frequency and economy account for the same facts. Iconicity makes the wrong prediction that lexical items with highly abstract or simple meanings should be consistently shorter than items with more concrete or complex meanings (as noted by Ronneberger-Sibold 1980: 239). It predicts, for example, that *entity* should be shorter than *thing* or *action*, that *animal* should be shorter than *cat*, that *perceive* should be shorter than *see*, and so on.[6]

4.2.   *Relative frequency and absolute frequency*

It is important to recognize that the relevant type of frequency for the purposes of this paper is *relative frequency*, not *absolute frequency* (cf. Corbett et al. 2001 for some discussion of this contrast). That is, what I am looking at here is the relation between the frequency of one category and the frequency of another category (within a class of lexemes or a construction): e.g., the relation between the frequency of singulars and the frequency of plurals (in nouns), the relation between the frequency of positive forms and the frequency of comparative forms (in adjectives), the relation between the frequency of inanimate objects and the frequency of inanimate objects (in transitive verb phrases), and so on.

I am not looking at the absolute frequencies of individual lexemes with a particular category. The absolute frequency of English *books*, the plural of *book*, is 131 (occurrences per million words, Leech et al. 2001), while the singular of *notebook* occurs only 8 times. But the singular and the plural should not be compared across different lexemes. The relative frequencies are as expected: *book* 243, *books* 131, *notebook* 8, *notebooks* 3. Likewise for positives and comparatives: the comparative *lower* occurs 111 times, and the positive *bright* occurs only 54 times. But the proportions (i.e., relative frequencies) are as expected: *low* 158, *lower* 111, *bright* 54, *brighter* 5.

What is crucial is that the items whose frequency and formal expression is compared are paradigmatic alternatives, i.e., that in some sense they must occur in the same slot. It is in such slots that expectations arise, so that more frequent items can make do with shorter coding because of their greater predictability. If two items are not paradigmatically related, it does not make so much sense to compare their frequency.

Another question is how big the frequency difference should be to be reflected in grammar. The answer is: significant. Perhaps one would see

bigger differences in form where the frequency differences are bigger, but this is an issue that I do not pursue in this paper.

4.3. *Adjectives and abstract nouns: Resolving an iconicity paradox*

Croft and Cruse (2004: 175) observe "a curious iconicity paradox" in connection with adjectives such as those in (11) and the corresponding abstract nouns:

(11) *long   leng-th*
     deep   dep-th
     high   heigh-t
     thick   thick-ness

They note that definitions of such adjectives presuppose a scale of length, depth, height, or thickness that is expressed by an abstract noun. Thus, *long* means something like 'noteworthy in terms of length' (cf. also Mel'čuk 1967). This abstract noun is thus conceptually simpler than the adjective, and yet it tends to be morphologically more complex across languages. The situation in (11) thus "appears to run counter to the principle that morphological complexity mirrors cognitive complexity" (Croft and Cruse 2004: 175).

Croft and Cruse try to solve the paradox, but do not seem to be very confident in their solution:

One possible explanation is that, in applying the iconic principle, we should distinguish between structural complexity (in terms of the number of elementary components and their interconnections) and processing complexity (in terms of the cognitive effort involved). Perhaps they are acquired first of all in an unanalyzed, primitive, 'Gestalt' sense, which is basically relative. Maybe in order to develop the full adult system, analysis and restructuring are necessary. Some of the results of the analysis may well be conceptually simpler in some sense than the analysand, but the extra effort that has gone into them is mirrored by the morphological complexity. (Croft and Cruse 2004: 175)

But in fact, no solution to the paradox is required, because it is a pseudo-paradox: There is no "principle that morphological complexity mirrors cognitive complexity". As we saw, morphological complexity (in the sense of length) mirrors rarity of use. It is easy to determine that adjectives are significantly more frequent than the corresponding abstract nouns. In (12), frequency figures from Leech et al. 2001 are given (the figures again indicate occurrences per million words). The example of *beautiful/beauty* shows that isolated exceptions to the coding regularity are possible.[7]

(12) *long*   392   *leng-th*   85
      deep   97   *dep-th*   41
      high   547   *heigh-t*   47
      thick   51   *thick-ness*   <10
      beautiful 87   *beauty*   44

4.4.   *The inchoative-causative alternation: Economy instead of iconicity*

In §3 and §4.1, we saw that pairs of noncausative (inchoative) and causative verbs are not uniformly coded: Sometimes the causative is coded overtly, based on the inchoative (e.g., Turkish *düş-Ø-mek* 'fall', *düş-ür-mek* 'fell, drop'), and sometimes the inchoative is coded overtly, based on the semantically causative verb. Such cases are called *anticausatives* (e.g., German *öffnen* 'open (tr.)', *sich öffnen* 'open (intr.)'; Russian *otkry-vat′-sja* 'open (tr.)', *otkryvat′-sja* 'open (intr.)').

   On the natural assumption that causatives have an additional meaning element (i.e. Russian *otkryvat′sja* means 'become open', and *otkryvat′* means 'cause to become open'), anticausative coding would be counter-iconic (as was observed by Mel'čuk 1967). This was seen as a problem by Haspelmath (1993), who assumed the iconicity-of-complexity principle (as well as markedness matching). However, Haspelmath found in a cross-linguistic study that different verb pairs tend to behave differently with respect to which member of the pair (the inchoative or the causative) tends to be coded overtly (cf. also Croft 1990b). Some verb meanings (which for convenience will be called *automatic*) tend to be coded as causatives (e.g., 'freeze', 'dry', 'sink', 'go out', 'melt'), whereas others (which for convenience will be called *costly*) are preferably coded as anticausatives (e.g., 'split', 'break', 'close', 'open', 'gather'). The idea behind the terms *automatic* and *costly* is that the automatic events do not often require input from an agent to occur, whereas the costly events tend not to occur spontaneously but must be instigated by an agent. While the automatic events conform to iconicity, it is especially the costly events that do not. Haspelmath tried to save the iconicity hypothesis by suggesting that in some way the frequency of occurrence of a particular event description is reflected in the way its meaning is treated by speakers:

Iconicity in language is based [not on objective meaning but] on conceptual meaning ... Events that are more likely to occur spontaneously will be associated with a conceptual stereotype (or prototype) of a spontaneous event, and this will be expressed in a structurally unmarked way. (Haspelmath 1993: 106–107)

This move is reminiscent of Lehmann's suggestion that rarity results in a high informational value and therefore somehow in high semantic

complexity (cf. note 6), and of the desperate attempt by Croft and Cruse to solve their iconicity paradox.

Fortunately, a much simpler explanation is available in which iconicity of complexity plays no role, and the coding preferences are explained in terms of economy: Automatic verb meanings tend to occur more frequently as inchoatives than costly verb meanings, which tend to occur more frequently as causatives. Due to economic motivation, the rarer elements tend to be overtly coded. Wright (2001: 127–128) presents some preliminary corpus evidence from English, as shown in Table 1:

Table 1. *Percentage of transitive (= causative) occurrences of some English inchoative-causative verb pairs*

verb pair	% transitive	
freeze	62%	↑ more causatives
dry	61%	
melt	72%	
burn	76%	
open	80%	↓
break	90%	more anticausatives

Thus, inchoatives and causatives behave in much the same way as singulars and plurals: Whichever member of the pair occurs more frequently tends to be zero-coded, while the rarer (and hence less expected) member tends to be overtly coded. Language-particular differences often obscure this picture (e.g., languages that never have overtly coded singulars, or languages lacking overtly coded causatives), which emerges fully only once a typological perspective is adopted.

4.5. *Differential object marking: Economy instead of iconicity*

It has long been observed (e.g., Blansitt 1973; Comrie 1989; Bossong 1985, 1998) that the overt coding of a direct object often depends on its animacy, and that such variation in object-marking can be subsumed under a general rule:

(13)   The higher a (direct) object is on the animacy scale, the more likely it is to be overtly coded (i.e., accusative-marked).

According to Comrie, this is because animate objects are not as "natural" as inanimate objects:

…the most natural kind of transitive construction is one where the A[gent] is high in animacy and definiteness and the P[atient] is lower in animacy and

definiteness; and any deviation from this pattern leads to a more marked construction. (Comrie 1989: 128)

In an interesting paper that tries to integrate insights from the functional-typological literature into an Optimality Theory (OT) framework, Aissen (2003: §3) proposes an account that appeals to a fixed constraint subhierarchy involving local conjunction of a *markedness hierarchy* of relation/animacy constraints (cf. 14) with a constraint against non-coding ($*\emptyset_{\text{CASE}}$):

(14)   markedness subhierarchy:
       *OBJ/HUM ≫ *OBJ/ANIM ≫ *OBJ/INAN

The resulting fixed constraint subhierarchy is shown in (15). Roughly this can be read as follows: Structures with zero-coded human objects are worse than structures with zero-coded animate objects, and these in turn are worse than structures with zero-coded inanimate objects.

(15)   *OBJ/HUM & $*\emptyset_{\text{CASE}}$ ≫ *OBJ/ANIM & $*\emptyset_{\text{CASE}}$ ≫ *OBJ/INAN & $*\emptyset_{\text{CASE}}$

Aissen motivates these constraints by appealing to markedness matching and iconicity:

The effect of local conjunction here is to link markedness of content (expressed by the markedness subhierarchy) to markedness of expression (expressed by $*\emptyset$). That content and expression are linked in this way is a fundamental idea of markedness theory (Jakobson 1939; Greenberg 1966). In the domain of Differential Object Marking, this is expressed formally through the constraints [in (15)]. Thus they are ICONICITY CONSTRAINTS: they favor morphological marks for marked configurations. (Aissen 2003: 449)

Combined with economy constraints (*STRUC), these constraints allow Aissen to describe all and only the attested language types in her framework.

However, a much more straightforward explanation of the Differential Object Marking universal is available: Inanimate NPs occur more frequently as objects, whereas animate NPs occur more frequently as subjects. Due to economic motivation, the rarer elements tend to be overtly coded. This explanation has in fact long been known (Filimonova 2005 cites antecedents in the 19th century), though actual frequency evidence has been cited only more recently (see Jäger 2004).[8]

Thus, no appeal to markedness matching or iconicity is needed, nor is Aissen's elaborate machinery of OT constraints needed to explain Differential Object Marking.

5. Iconicity of cohesion: Advocates and examples

Iconicity of cohesion was defined in §1 as follows:

(16) Meanings that belong together more closely are expressed by more cohesive forms.

Iconicity of cohesion is discussed in detail by Haiman (1983) under the label ''iconic expression of conceptual distance'' (''The linguistic distance between expressions corresponds to the conceptual distance between them'', Haiman 1983: 782).[9] What he means by *linguistic distance* is made clear by the scale in (17), where (a)–(d) show diminishing linguistic distance (in my terms, increasing *cohesion*).

(17) Haiman's (1983: 782) cohesion scale
    a. X *word* Y   (function-word expression)
    b. X Y   (juxtaposition)
    c. X–Y   (bound expression)
    d. Z   (portmanteau expression)

I prefer the term *cohesion* to *distance* for this scale, because (b) and (c) do not literally differ in distance, and distance is not really applicable to (d).

Moreover, I want to distinguish strictly between cohesion and contiguity. That there is a functionally motivated preference for contiguity, i.e., for elements that belong together semantically to occur next to each other in speech, is beyond question (see also Hawkins 2004: Ch. 5). Newmeyer's (1992: 761–762) discussion of ''iconicity of distance'' (and similarly Givón's (1985: 202, 1991: 89) ''proximity principle'') conflate cohesion and contiguity. I only argue against an iconicity-based explanation of phenomena related to cohesion.

The following four examples of iconicity of cohesion are the most important cases cited in the literature:

(i) <u>Possessive constructions:</u> Inalienable possession shows at least the same degree of cohesion as alienable possession, because in inalienable possession (i.e., possession of kinship and body part terms) the possessor and the possessum belong together more closely semantically (Haiman 1983: 793–795, 1985: 130–136; see also Koptjevskaja-Tamm 1996). An example:

(18) Abun (West Papuan; Berry and Berry 1999: 77–82)
    a. *ji bi nggwe*
        I of garden
        'my garden'
    b. *ji syim*
        I arm
        'my arm'

(ii) <u>Causative constructions:</u> Causative constructions showing a greater degree of cohesion tend to express direct causation (where cause and result belong together more closely), whereas causative constructions showing less cohesion tend to express indirect causation (Haiman 1983: 783–787; cf. also Comrie 1989: 172–173; Dixon 2000: 74–78). The following example is cited by Dixon (2000: 69):

(19)   Buru (Austronesian; Indonesia; Grimes 1991: 211)
    a.   *Da   puna   ringe   gosa.*
        3SG.A cause 3SG.O be.good
        'He (did something which, indirectly,) made her well.'
    b.   *Da   pe-gosa   ringe.*
        3SG.A CAUS-be.good 3SG.O
        'He healed her (directly, with spiritual power).'

A similar Japanese example is provided by Horie (1993: 26):

(20)   a.   *John-wa   Mary-ni   huku-o   ki-se-ta.*
        John-TOP Mary-DAT clothes-ACC wear-CAUS-PAST
        'John put clothes on Mary.'
    b.   *John-wa   Mary-ni   huku-o   ki   sase-ta.*
        John-TOP Mary-DAT clothes-ACC wear cause-PAST
        'John made Mary wear clothes.'

The much-discussed English distinction between *kill* and *cause to die* is of course also an instance of this contrast (e.g., Lakoff and Johnson 1980: 131).

(iii) <u>Coordinating constructions:</u> Many languages distinguish between *loose coordination* and *tight coordination* (i.e., less vs. more cohesive patterns), where the first expresses greater conceptual distance and the latter expresses less conceptual distance (Haiman 1983: 788–790, 1985: 111–124). Haiman discusses coordination of clauses and cites the two examples in (21) and (22), where the greater cohesion is manifested by the absence of a coordinator. In (21a), the greater conceptual distance lies in the temporal non-connectedness, while in (22a), the greater conceptual distance lies in the lack of subject identity.

(21)   Fe'fe' (Bantoid; Cameroon; Hyman 1971: 43)
    a.   *à   kà   gén   ntēe   <u>nī</u>   njwēn   lwà'*
        he PAST go market and buy yams
        'He went to the market and also (at some later date) bought yams.'
    b.   *à   kà   gén   ntēe   njwēn   lwà'*
        he PAST go market buy yams
        'He went to the market and bought yams (there).'

(22)   Aghem (Bantoid; Cameroon; Anderson 1979: 114)
    a.   *Ò   nám   kɨbɛ́   ghá?   ?ɣía   zɨ*
        she cook fufu we.excl and   eat
        'She cooked fufu and we ate it.'
    b.   *Ò   mɔ̀   zɔ́m mám kɨbɛ́*
        she PAST sing cook fufu
        'She sang and cooked fufu.'

Wälchli (2005: Ch. 3) also discusses noun phrase coordination and cites contrasts such as (23). He calls the semantic distinction between them "accidental coordination" vs. "natural coordination", and claims that the formal contrast between loose coordination in (23a) and tight coordination in (23b) iconically reflects this semantic contrast (2005: 13).

(23)   Georgian
    a.   *gveli <u>da</u> k'ac'i*
        snake and man
        'the snake and the man'
    b.   *da-dzma*
        sister-brother
        'brother and sister'

(iv) <u>Complement clause constructions</u>: Haiman (1985: 124–130) also discusses complement-clause constructions in terms of iconicity of cohesion mirroring conceptual closeness. He observes that in the contrast in (24), the "reduced or contracted version signals conceptual closeness (same subject), while a non-reduced version signals conceptual distance (different subject)" (1985: 126).

(24)   a.   *Who do you wanna succeed?* (who=patient; same subject)
      b.   *Who do you want to succeed?* (who=agent possible; diff.
                                     subject possible)

But much better known is Givón's work on iconic form-function correspondences in complement clauses (1980, 1990: Ch. 13, 2001: Ch. 12; see also 1985: 199–202, 1991: 95–96), which posits a scale of "event integration" (called "binding hierarchy" in earlier versions) that corresponds to a scale of formal integration. In the most recent version of this, Givón posits an iconic principle of "event integration and clause union":

The stronger is the semantic bond between the two events, the more extensive will be the syntactic integration of the two clauses into a single though complex clause (Givón 2001: 40)

Among his examples are contrasts such as the following, where in each case the first example exhibits greater event integration and greater syntactic integration (non-finiteness and/or absence of a complementizer):

(25) a. *John made Mary quit her job*.   (2001: 45)
     b. *John caused Mary <u>to</u> quit her job*.

(26) a. *She wanted him to leave*.   (2001: 47)
     b. *She wished <u>that</u> he would leave*.

(27) a. *She told him to leave*.   (2001: 48)
     b. *She insisted <u>that</u> he must leave*.

(28) a. *She saw him coming out of the theatre*.   (2001: 50)
     b. *She saw <u>that</u> he came out of the theatre*.

6.   Iconicity of cohesion: frequency-based explanation

My claim here is that Haiman's cohesion scale in (17) does not reflect one single underlying cause. It should be taken apart into three different distinctions: (i) overt coding vs. lack of coding (X *word* Y vs. X Y), (ii) juxtaposition vs. bound expression (X Y vs. X-Y), and (iii) portmanteau expression (Z). All three are related to frequency, but not in the same way. This is clearest in the case of portmanteau expression (or suppletion), which only occurs when the combination of the two elements has a high *absolute frequency*. For instance, in the domain of causative constructions, English has the bound causatives *sadd-en* 'make sad', *wid-en* 'make wide', *hard-en* 'make hard', but it is only for high-frequency adjectives like *good* and *small* that it has suppletive causatives (*improve* 'make good', *reduce* 'make small'). Similarly, a few cases of suppletion in possessive constructions are attested, but these all come from high-frequency nouns such as 'mother' (e.g., Juǀ'hoan *taqè* 'mother', *áíá* 'my mother', Dickens 2005: 35). The reason why high absolute frequency favours suppletion (and irregularity more generally) has long been known: High frequency elements are easy to store and retrieve from memory, so there is little need for regularity (cf. Osthoff 1899, Ronneberger-Sibold 1988).

    However, the overt-covert contrast (X *word* Y vs. X Y) and the free-bound contrast (X Y vs. X-Y) are due to frequency-induced predictability, as seen earlier for contrasts that others have explained by iconicity of quantity (§2) and by iconicity of complexity (§3–4). Predictability leads to shortness of coding by economy, and shortness of coding itself leads to bound expression, because short (and unstressed) elements do not have enough bulk to stand on their own. The phenomena that Haiman

explains through iconicity of cohesion actually all instantiate only the overt-covert contrast and/or the free-bound contrast, so what matters for them is again relative frequency.

Let us now examine the four main construction types with alleged effects of iconicity of cohesion to see how their properties can be explained in terms of relative frequency.

6.1. *Possessive constructions*

With inalienably possessed nouns, possessive constructions are of course much more frequent than with alienably possessed nouns (cf. Nichols 1988: 579). This can be easily demonstrated with corpus figures. Table 2 shows frequencies of three (hopefully representative) sets of nouns in spoken English and spoken Spanish.

We see that alienable nouns occur as possessed nouns in a possessive construction only relatively rarely (12% and 7% of the time, respectively),

Table 2. *Frequencies of selected kinship terms, body part terms and alienable nouns*

English	kinship terms[a]		body part terms[b]		alienable nouns[c]	
total	16235	100%	11038	100%	24991	100%
possessed	7797	48%	4940	45%	2967	12%
nonpossessed	8434	52%	6098	55%	22024	88%

Source: British National Corpus, spoken part
a = *mother, father, brother(s), sister(s), wife, husband, son(s), daughter(s), mum, dad, grandfather, grandmother, aunt, uncle*
b = *head, hand(s), face, finger(s), knee(s), ear(s), leg(s), wrist, hair, nose, neck, belly, skin, elbow, chest*
c = *car, dinner, health, tree, knife, bed, community, meat, money, bike, suitcase, tools, book(s), room, bedroom, kitchen*

Spanish	kinship terms[d]		body part terms[e]		alienable nouns[f]	
total	18391	100%	8863	100%	10913	100%
possessed	7362	40%	1297	15%	776	7%
nonpossessed	11029	60%	7566	85%	10137	93%

Source: Corpus del Español, spoken part
d = *madre, padre(s), hermano(s), hermana(s), esposa, marido, hijo(s), hija(s), mamá, papá, abuelo(s), abuela, tía, tío*
e = *cabeza, mano(s), cara, dedo(s), rodilla(s), oído(s), pierna(s), muñeca, pelo, nariz, cuello, vientre, piel, codo, pecho, hombro(s)*
f = *coche, cena, salud, árbol, cuchillo, cama, comunidad, pueblo, carne, dinero, bicicleta, maleta, herramientas, libro(s), habitación, dormitorio, cocina*

while it is very common for kinship terms and body part terms to occur as possessed nouns. (The fact that the figure for Spanish body part terms is relatively low here is due to the omissibility of overt possessors in body-part constructions like *levanta la mano* 'raise your hand'; strictly speaking, all notional possessors would have to be counted, but this is impossible to do automatically.)

As we saw in §4.2, what counts is relative frequencies, not absolute frequencies. Since frequent alienable nouns like 'house' or 'show' are much more frequent than rare inalienable nouns like 'kidney' or 'great niece' in most cultural contexts, the alienable nouns may well occur in a possessive construction more often than the inalienable nouns. However, the percentage of possessed occurrences of inalienable nouns will always be significantly higher than the corresponding percentage of alienable nouns. Thus, upon encountering an inalienable noun, it will be much easier to predict that it occurs in a possessive construction, and the possessive marking is therefore relatively redundant. Since languages are efficient systems, they tend to show less overt coding with inalienable nouns. Moreover, since pronominal possessors are more predictable, they show a greater tendency to become affixed, thus accounting for the contrast between juxtaposition and bound expression.

Crucially, the economy account given here makes somewhat different predictions from Haiman's (1983) iconicity account. The facts show that the predictions of the economy account are the correct ones.

First, the iconicity account is compatible with a hypothetical situation in which the pronominal possessor in the inalienable construction is actually longer than the corresponding form in the alienable possession. However, economy additionally predicts that the form of the inalienable pronominal possessor not only tends to be bound, but also tends to be shorter than the alienable possessor. This is in general borne out, and I know of no counterexamples. Some examples are given in (29).

(29)		ALIENABLE CONSTRUCTION	INALIENABLE CONSTRUCTION
a.	Nakanai (Johnston 1981: 217)	*luma taku* house I 'my house'	*lima-gu* hand-1SG 'my hand'
b.	Hua (Haiman 1983: 793)	*dgaiʔ fu* I pig 'my pig'	*d-zaʔ* 1SG-arm 'my arm'
c.	Ndjébbana (McKay 1996: 302–6)	*budmánda ngáyabba* suitcase I 'my suitcase'	*nga-ngardabbámba* 1SG-liver 'my liver'

d. Kpelle   *ŋa   pɛrɛi*   *m-pôlu*
   (Welmers 1973: 279)   I   house   1SG-back
                           'my house'   'my back'

e. Juǀ'hoan   *mí   tjù*   *m   bá*
   (Dickens 2005: 35)   1SG   house   1SG   father
                           'my house'   'my father'

Second, Haiman's account in terms of distance matching predicts that the additional element in alienable constructions should occur in the middle between the possessor and the possessum, as seen in the canonical examples from Maltese (*is-siġġu tiegħ-i* [the-chair of-me] 'my chair', see §1) and from Abun (*ji bi nggwe* [I of garden] 'my garden', see (18)). However, the extra element may also occur to the left or right of both the possessor and the possessum, as seen in (30).

(30)		ALIENABLE CONSTRUCTION	INALIENABLE CONSTRUCTION
a.	Puluwat	<u>*nay*</u>-*iy* *hamwol*	*pay-iy*
	(Elbert 1974: 55, 61)	POSS-1SG chief	hand-1SG
		'my chief'	'my hand'
b.	ʹOʹodham	*ñ-mi:stol*-<u>*ga*</u>	*ñ-je′e*
	(Zepeda 1983: 74–81)	1SG-cat-POSSD	1SG-mother
		'my cat'	'my mother'
c.	Koyukon	*se-tel*-<u>*e′*</u>	*se-tlee′*
	(Thompson 1996: 654, 667)	1SG-socks-POSSD	1SG-head
		'my socks'	'my head'
d.	Achagua	*nu-caarru*-<u>*ni*</u>	*nu-wíta*
	(Wilson 1992)	1SG-car-POSSD	1SG-head
		'my car'	'my head'

My economy account only predicts that the coding of inalienable constructions should tend to be shorter, but it says nothing about the position of the extra coding element in alienable constructions, so cases like (30a–d) are counterevidence to Haiman's iconicity account, but compatible with my economy account. Haiman (1983: 795) himself cites the Puluwat example, recognizes that it is a problem for him, and acknowledges the need to reformulate his initial generalization. But he does not seem to recognize that the facts no longer support any role of iconicity.

Finally, some languages show overt coding of inalienable nouns as well, but only when they are not possessed. An example comes from Koyukon (Athabaskan; Thompson 1996: 654, 656, 667):

(31) Koyukon   UNPOSSESSED   POSSESSED
      alienable   *teł*   *se-tel-e'*
                  socks   1SG-socks-POSSD
                  'socks'   'my socks'
      inalienable *k'e-tlee'*   *se-tlee'*
                  UNSP-head   1SG-head
                  'head'   'my head'

Haiman's iconicity does not make any predictions about unpossessed constructions, but the economy account predicts just what we see: Alienable nouns tend to have overt coding in the possessed construction, whereas inalienable nouns tend to have overt coding in the unpossessed construction.

Thus, the iconicity account is both too weak (in that it does not predict the shortness of inalienable possessive pronouns, seen in (29)) and too strong (in that it wrongly predicts that the patterns in (30) should not be possible). Economy, by contrast, makes just the right predictions.

6.2. *Causative constructions*

Again I claim that direct causatives are significantly more frequent than indirect causatives and that that explains why they exhibit more cohesive coding than indirect causatives. No appeal to iconicity is necessary.

In order to show that this is true, ideally one would examine a corpus of a language with a regular grammatical contrast between direct and indirect causation, as illustrated in (19) for Buru and in (20) for Japanese. I hope that this paper will inspire such research, and I expect that the direct causatives are much more frequent than the indirect causatives. In the literature on English, the contrasts between the different types of periphrastic causatives have received some attention. According to Gilquin (2006: 7), the frequency in the British National Corpus of the four causative verbs that combine with an infinitive are as in (32):

(32)

	spoken	written	total
make ('*I made him go*')	898	258	1,156
get ('*I got him to go*')	350	52	402
cause ('*I caused him to go*')	15	207	222
have ('*I had him go*')	48	29	77

Since the *make* and *get* causatives are usually regarded as expressing a more direct type of causation, while the *cause* and *have* causatives express a more indirect type of causation, this is just what we would expect.

It is also possible to compare lexical causative verbs with the corresponding periphrastic *cause* causatives (this is also what Haiman 1983

mostly does for the semantic aspects). Some figures from the British National Corpus are given in (33) (these are only the forms with a pronoun object, i.e., *kill me, cause him to die*, etc).

(33)	*stop*	3267	*cause to stop*	6
	kill	2400	*cause to die*	2
	raise	466	*cause to rise*	3
	bring down	269	*cause to come down*	0
	drown	80	*cause to drown*	0

These comparisons are more problematic than those in (32) in that the length of the two types of causatives differs sharply, so one might suspect that the lexical direct causatives are more frequent simply because they are shorter. In general, such effects do not seem to be particularly strong, if they exist at all (see Haspelmath 2008: §6.5 for further discussion), but still in the ideal case we would like to perform our corpus study on a language where all causatives are expressed grammatically (i.e., even 'kill' and 'raise' are expressed as 'die-CAUS' and 'rise-CAUS'). But since many direct causatives are highly frequent (in an absolute sense) in all languages, we normally find a lot of portmanteau expression of causatives, which limits our options for corpus counts. Nevertheless, the figures in (32) and (33) should be sufficient to make a good initial case for the claim that direct causatives are generally more frequent than indirect causatives.

If this is true, then the economy account makes a further prediction: that markers of indirect causation should not only be less cohesive, but also tend to be longer. And indeed a number of languages have two causatives differing primarily in length, not in cohesion (cf. Dixon 2000: 74–78).

(34)		INDIRECT CAUSATIVE	DIRECT CAUSATIVE
a.	Amharic	*as-bälla*	*a-bälla*
	(Haiman 1983: 786,	CAUS-eat	CAUS-eat
	Amberber 2000: 317–320)	'force to eat'	'feed'
b.	Hindi	*ban-vaa-*	*ban-aa-*
	(Dixon 2000: 67,	be.built-CAUS	be.built-CAUS
	Saksena 1982)	'have sth. built'	'build'
c.	Jinghpaw	*-shangun*	*sha-*
	(Maran and Clifton 1976)		
d.	Creek	*-ipeyc*	*-ic*
	(Martin 2000: 394–399)		

Although Haiman (1983: 786) cites the example from Amharic as an instance of an iconicity contrast, it does not actually fit his iconicity explanation. The two causatives of Amharic and the other languages in (34)

do not differ in cohesion, but only in length, so the contrast is predicted only by the economy account.[10]

6.3.   *Coordinating constructions*

While Haiman's discussion of examples like (21–22) above only mentions the semantic contrast between greater and less conceptual distance, Wälchli's terminology (accidental vs. natural coordination) already points to the real motivating factor: Natural coordination (as in 21b, 22b and 23b) is "natural", i.e., frequent and expected for the pair of expressions, while accidental coordination is infrequent and hence unexpected. Thus, it is economical to use more explicit and less cohesive coding in accidental coordination, and less explicit and more cohesive coding in natural coordination.

Doing the frequency counts for clause coordination is fairly trivial. For example, in the German version of *The wolf and the seven little kids* (one of Grimm's fairy tales), there are 47 *und*-coordinations, and 41 of them show subject identity, while only 6 have different subjects. All 47 cases exhibit temporal closeness.

For noun phrase conjunction of the type *da-dzma* 'brother-and-sister' (23b), the frequency counts are less straightforward, because the definition of accidental and natural coordination is quite vague: Wälchli (2005: 5) describes natural coordination as "coordination of items which are expected to cooccur, which are closely related in meaning, and which form conceptual units". This is not specific enough to test the claim directly, but it seems plausible that for noun phrases, too, it will be possible to show that coordinations of the type 'brother and sister' will turn out to be more frequent than coordinations of the type 'the man and the snake'.

6.4.   *Complement-clause constructions*

For many of the examples given by Haiman and Givón, the frequency explanation is completely straightforward. With 'want' verbs (cf. 24a–b), the same-subject use is of course overwhelmingly more frequent than the different-subject use, for well-understood reasons (our desires naturally concern first of all our own actions), and this is often reflected in shorter coding (cf. Haspelmath 1999). This explains the contrast between English *wanna* and *want to*, and also a similar contrast between *gotta* and *got to* (*I gotta go home now* vs. *I got to go to Hawaii last winter*) that was already pointed out and correctly explained by Bolinger (1961: 27) ("condensation is tied to familiarity", cited approvingly by Haiman 1985: 126).

There are also obvious frequency asymmetries between the pairs *make/ cause* (cf. 25), *want/wish* (cf. 26), and *tell/insist* (cf. 27) which suffice to

explain the shorter coding of the first member of each pair.[11] Givón is right that in each case there is also a semantic contrast, but in order to show that the semantic contrast is indeed responsible for the formal contrast, he should provide contrasting examples of constructions with roughly equal frequency.

In contrasts such as (28a–b) (*She saw him coming out of the theatre* vs. *She saw that he came out of the theatre*), which do not exhibit a striking frequency asymmetry, another factor is clearly highly relevant: In (28a), the complement event necessarily occurs simultaneously with the main event, in contrast to (28b), where the complement event could take place at some other time (*She saw that he would come out only two hours later / that he had come out two hours earlier*). In Cristofaro's (2003: §5.3.2) terms, (28a) shows ''predetermination'' of the tense value of the complement clause, and Cristofaro rightly explains the lack of finiteness (i.e., the lack of tense) in (28a) as due to ''syntagmatic economy'': Information that can be readily inferred from the context can be left out. (See also Horie 1993: 203–212 for related discussion.)

This factor of predetermination is of course not unrelated to the broader notion of semantic closeness. If a complement-taking verb predetermines the tense value and other semantic properties of its complement, this can be seen as one facet of ''conceptual closeness'' or ''event integration''. However, such cases do not provide evidence for iconicity of cohesion, because the higher syntagmatic cohesion of *She saw him coming out of the theatre* would be expected anyway for reasons of economy.[12]

7. Conclusion

I conclude that for most of the core phenomena for which iconicity of quantity, complexity and cohesion have been claimed to be responsible, there are very good reasons to think that they are in fact explained by frequency asymmetries and the economy principle. The final result may look iconic to the linguist in some cases, but iconicity is not the decisive causal factor.

Linguists have rarely discussed the mechanism by which iconicity could come to have a causal role in shaping grammars. However, Givón claims that iconic structures are easier to process than noniconic structures:

The iconicity meta-principle: All other things being equal, a coded experience is easier to store, retrieve, and communicate if the code is maximally isomorphic to the experience. (Givón 1985: 189)

And similarly, Dressler et al. (1987: 18) say that the more iconic a sign is, the more ''natural'' it is, i.e., the easier speakers find using it.

If these claims were correct also for iconicity of quantity, complexity and cohesion, it would indeed be predicted that such iconic structures should be preferred by speakers, and we should see a significant effect of iconicity in language structures. But in fact we do not see such an effect. We see effects of frequency and predictability, i.e. of the economy principle, which (as everyone agrees) is independently needed. What we can conclude from this is that the above claims are wrong, i.e., that iconic structures are apparently not necessarily preferred in processing.

The respective role of iconicity and economy was discussed already in the 1980s. Haiman (1983: 802) recognized that formal complexity/simplicity is very often economically motivated, and he rejected the subsumption of economic motivation under iconicity, even though one might argue that the correspondence between a linguistic dimension (full vs. reduced form) and a conceptual dimension (unpredictable vs. predictable) is itself iconic. As an example of economic motivation, he cites the tendency for predictable referents to be coded with little material (short pronouns or zero), while less predictable or unpredictable referents are coded with more material (longer pronouns or full NPs) (as documented in Givón (ed.) 1983).

However, Givón (1985: 197) sees the correlation between unpredictability and amount of coding material as primarily iconic (see also Givón 1991: 87–89), and he objects to Haiman's economy account:

. . . the principle of economy has not been working here by itself, since the end result of such a situation would have been the exclusive use of zero anaphora for all topic identification in discourse. (Givón 1991: 87–89)

But that economy (favoring the speaker's needs) is not the only relevant factor in communication should be clear from the beginning—if there were no opposite principle of distinctiveness (favoring the hearer), we would have no linguistic forms at all. Another argument that Givón makes is the following:

It may well be that Zipf-like economy considerations were indeed involved in the diachronic . . . shaping of the quantity-scale . . . But the end result is nonetheless an iconic—isomorphic—relation between code and coded. And such a relation surely carries its own meta-motivation, i.e., [the iconicity meta-principle, cited at the beginning of this section]. (Givón 1991: 87–89)

This last sentence simply does not follow. If the end result is iconic in the eyes of the analyst, this does not mean that it is iconically motivated, i.e., that iconicity is a relevant causal factor. The empirical evidence from

frequency distributions and cross-linguistic coding types that was cited in this paper shows that iconicity may well be irrelevant for an explanation of the grammatical asymmetries considered here. That is, in the debate between Haiman and Givón, Haiman was right to favor economy over iconicity in explaining the quantity scale for referent expressions. However, as I hope to have shown here, Haiman's economy explanation should be extended also to many other cases that he and others explained in terms of iconicity.

Notes

1.   In C. S. Peirce's received typology of signs, there are three types of icons: diagrams, images, and metaphors (see, e.g., Dressler 1995 for discussion). Nowadays metaphor is not generally discussed under the heading of iconicity, and imagic iconicity is relevant primarily for onomatopoeia. This paper is exclusively concerned with possible iconicity effects in grammar, so only diagrammatic iconicity will be considered here. The relevance of Peirce's semiotic concepts to the study of grammar was first brought to linguists' attention by Jakobson (1965).

2.   The idea that (syntagmatic and paradigmatic) isomorphism can be considered an instance of Peircean iconicity was apparently first proposed by Anttila (1989, originally published in 1972). A number of authors have noted that this represents a fairly extreme extension of Peirce's original concept, and Itkonen (2004) flatly rejects the subsumption of isomorphism under iconicity.

3.   Haiman (1985: 194–195) recognizes that "the motivation for the reduction is also partly economic: one gives less expression to that which is familiar or predictable", but he does not consider the possibility that the motivation may be entirely economic.

4.   Lakoff and Johnson (1980: 127) apply their principle "MORE OF FORM IS MORE OF CONTENT" (which they call a metaphor, not relating it to iconicity) to cases of iteration (*She ran and ran and ran and ran*) and lengthening (*He is bi-i-i-i-ig!*). Such extragrammatical phenomena may well be motivated by a kind of iconicity of quantity. However, their attempt to extend the principle to grammatical reduplication fails: While many cases of reduplication signal "more of content" (e.g., plurality, continuative aspect), this is by no means always the case (Moravcsik 1978 also mentions a widespread sense of diminution and attenuation, and more specific senses such as indifference and pretending). Grammatical reduplication is apparently just like affixation in that the reduplicated form is always the rarer one.

5.   The third vs. first/second person contrast has also been interpreted as a kind of "iconicity of absence" (closely related to iconicity of quantity as seen in §2): Haiman (1985:

4–5), citing Benveniste (1946), claims that the third person, as a non-speech act participant, can be seen as an "absent" person, a "non-person" that is iconically represented by a non-desinence (i.e., zero). But neither Benveniste nor Haiman mention imperatives, where the hearer is present, but a second-person desinence is typically absent (see (10) below). (See also the discussion in Helmbrecht 2004: 228–229.)

6. Lehmann (1974: 113) notes that length correlates with rarity, but instead of following Zipf in explaining length with reference to frequency/rarity, he suggests that rarity can also be seen as equivalent to improbability or informational value. He then assumes that informational value correlates with semantic complexity and infers that rare items tend to be semantically complex. But evidently informational value in the statistical sense is very different from semantic complexity. Talking about *animals* or *perceiving* is perhaps in some technical sense of high informational value (even though it is not very informative), but it is hard to argue that *animal* and *perceive* are semantically more complex than *cat* or *see*.

7. A reviewer observes that the English pair *widow*/*widower* in (10) is also an isolated exception and asks how it is different from *beautiful*/*beauty*. The answer is that the *widow*/*widower* contrast is not isolated from a cross-linguistic point of view: There is a general tendency for this pair to show overt coding on the male member (e.g., German *Witwe*/*Witw-er*, Russian *vdova*/*vdov-ec*), whereas *beautiful*/*beauty* is isolated not only within English, but also cross-linguistically.

8. Also much of the earlier functionalist literature is insufficiently explicit with regard to the causal factor. For example, Comrie (1989: 128) only invokes the "naturalness" of certain associations between role and animacy, a relatively vague notion compared to frequency.

9. Cf. also Lakoff and Johnson's (1980: 128–132) principle "CLOSENESS IS STRENGTH OF EFFECT", which is, however, not related to iconicity by them, but is regarded as a metaphor. The frequency-based perspective here suggests that Lakoff and Johnson's metaphor-based account is not necessary.

10. A further observation is that direct vs. indirect causation is not the only semantic parameter by which competing causatives differ. Dixon (2000: 76) lists the following parameters and observes that they all tend to correlate with the degree of "compactness" of the causative marker (i.e., its shortness).

LONGER MARKER	SHORTER MARKER
action	state
transitive	intransitive
causee having control	causee lacking control
causee unwilling	causee willing
causee fully affected	causee partially affected
accidental	intentional
with effort	naturally

Not all of these can be subsumed under "less conceptual distance", but they can be plausibly related to frequency asymmetries. This is a matter for future research.

11. Leech et al. 2001 give the following figures for the verbal lexemes, which can be taken as representative for the complement-clause constructions as well: *want* 945, *wish* 30; *tell* 775, *insist* 67; *make* 2165, *cause* 206.

12. Cristofaro (2003: Ch. 9), while pointing to the importance of the factor of predetermination, still wants to retain semantic integration and iconicity as explanatory factors for complement-clause constructions. But like Haiman and Givón, she does not even consider the potential explanatory value of frequency-based economy.

References

Aissen, Judith
  2003   Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21(3), 435–483.
Amberber, Mengistu
  2000   Valency-changing and valency-encoding devices in Amharic. In R. M. W. Dixon, and Alexandra Y. Aikhenvald (eds.), *Changing Valency*. Cambridge: Cambridge University Press, 312–332.
Anderson, Stephen C.
  1979   Verb structure. In Larry Hyman (ed.), *Aghem Grammatical Structure* (Southern California Occasional Papers in Linguistics 7). Los Angeles: University of Southern California, 73–136.
Anttila, Raimo
  1989   *Historical and Comparative Linguistics*. 2nd ed. Amsterdam: Benjamins.
Benveniste, Émile
  1946   Relations de personne dans le verbe. *Bulletin de la Société de Linguistique de Paris* 43, 1–12.
Berry, Keith and Christine Berry
  1999   *A description of Abun: A West Papuan language of Irian Jaya*. (Pacific Linguistics, B-115) Canberra: Australian National University.
Blansitt, Edward L.
  1973   Bitransitive clauses. *Working Papers in Language Universals (Stanford)* 13, 1–26.
Bolinger, Dwight
  1961   *Generality, Gradience, and the all-or-none*. The Hague: Mouton.
Bossong, Georg
  1985   *Differenzielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Narr.
  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. and Paul Hopper (eds.)
  2001   *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.
Comrie, Bernard
  1989   *Language Universals and Linguistic Typology*. 2nd ed. Oxford: Blackwell.
Corbett, Greville, Andrew Hippisley, Dunstan Brown, and Paul Marriott
  2001   Frequency, regularity and the paradigm: A perspective from Russian on a complex relation. In Bybee, Joan and Paul Hopper (eds), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins. 201–226.
Cristofaro, Sonia
  2003   *Subordination*. Oxford: Oxford University Press.
Croft, William
  1990a   *Typology and Universals*. Cambridge: Cambridge University Press.
  1990b   Possible verbs and the structure of events. In Tsohatzidis, S. L. (ed.), *Meanings and Prototypes: Studies in Linguistic Categorization*. London: Routledge, 48–73.

2003　　　*Typology and Universals*. 2nd ed. Cambridge: Cambridge University Press.

Croft, William and Alan Cruse
2004　　　*Cognitive Linguistics*. Cambridge: Cambridge University Press.

Dickens, Patrick J.
2005　　　*A Concise Grammar of Ju|'hoan*. Cologne: Köppe.

Dixon, R. M. W.
2000　　　A typology of causatives: Form, syntax and meaning. In Dixon, R. M. W. and Alexandra Y. Aikhenvald (eds.), *Changing Valency*. Cambridge: Cambridge University Press, 30–83.

Dressler, Wolfgang U.
1995　　　Interactions between iconicity and other semiotic parameters in language. In Raffaele Simone (ed.), *Iconicity in Language*. Amsterdam: Benjamins, 21–37.

Dressler, Wolfgang U., Willi Mayerthaler, Oswald Panagl and Wolfgang U. Wurzel
1987　　　*Leitmotifs in Natural Morphology*. (Studies in Langauge Companion Series 10). Amsterdam: Benjamins.

Elbert, Samuel
1974　　　*Puluwat Grammar*. (Pacific Linguistics, B-29.) Canberra: Australian National University.

Fenk-Oczlon, Gertraud
1990　　　Ikonismus versus Ökonomieprinzip: Am Beispiel russischer Aspekt- und Kasusbildungen. *Papiere zur Linguistik* 42(1), 49–69.

Fenk-Oczlon, Gertraud
1991　　　Frequenz und Kognition—Frequenz und Markiertheit. *Folia Linguistica* 25(3–4), 361–394.

Filimonova, Elena
2005　　　The noun phrase hierarchy and relational marking: Problems and counterevidence. *Linguistic Typology* 9(1), 77–113.

Gilquin, Gaëtanelle
2006　　　The verb slot in causative constructions: Finding the best fit. *Constructions* SV1-3. (www.constructions-online.de)

Givón, Talmy
1980　　　The Binding Hierarchy and the typology of complements. *Studies in Language* 4, 333–377.
1985　　　Iconicity, isomorphism and non-arbitrary coding in syntax. In John Haiman (ed.), *Iconicity in Syntax*. Amsterdam: Benjamins, 187–219.
1990　　　*Syntax: A Functional-Typological Introduction. Vol. II*. Amsterdam: Benjamins.
1991　　　Isomorphism in the grammatical code: Cognitive and biological considerations. *Studies in Language* 15(1): 85–114.
1995　　　Markedness as meta-iconicity: Distributional and cognitive correlates of syntactic structure. In *Functionalism and Grammar*, T. Givón, 25–69. Amsterdam: Benjamins.
2001　　　*Syntax: An Introduction. Volume II*. Amsterdam: Benjamins.

Givón, T. (ed.)
1983　　　*Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam: Benjamins.

Greenberg, Joseph H.
1963[1966]　Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of Grammar*. 2nd ed. 1966. Cambridge, Mass.: MIT Press, 73–113.

1966   *Language Universals, with Special Reference to Feature Hierarchies.* (Janua Linguarum, Series Minor 59). The Hague: Mouton.

Haiman, John
1980   The iconicity of grammar. *Language* 56, 515–540.
1983   Iconic and economic motivation. *Language* 59, 781–819.
1985   *Natural Syntax.* Cambridge: Cambridge University Press.
2000   Iconicity. In Geert Booij, Joachim Mugdan and Christian Lehmann (eds.), *Morphology: An International Handbook* Vol. I. Berlin: de Gruyter, 281–288.

Haspelmath, Martin
1993   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.
1999   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. (Handout available from author's website.)
2006   Against markedness (and what to replace it with). *Journal of Linguistics* 42(1), 1–46.
2008   Creating economical morphosyntactic patterns in language change. To appear in Good, Jeff (ed.), *Language Universals and Language Change.* Oxford: Oxford University Press.

Hawkins, John A.
2004   *Efficiency and Complexity in Grammars.* Oxford: Oxford University Press.

Helmbrecht, Johannes
2004   Ikonizität in Personalpronomina. *Zeitschrift für Sprachwissenschaft* 23, 211–244.

Horie, Kaoru
1993   A cross-linguistic study of perception and cognition verb complements: A cognitive perspective. PhD dissertation, University of Southern California.

Hockett, Charles F.
1958   *A Course in Modern Linguistics.* New York: MacMillan.

Horn, Wilhelm
1921   *Sprachkörper und Sprachfunktion.* Berlin: Mayer and Müller.

Hyman, Larry
1971   Consecutivization in Fe'fe'. *Journal of African Languages* 10(2), 29–43.

Itkonen, Esa
2004   Typological explanation and iconicity. *Logos and Language* 5(1), 21–33.

Jäger, Gerhard
2004   Learning constraint sub-hierarchies: The bidirectional gradual learning algorithm. In Blutner and H. Zeevat (eds.), *Pragmatics in OT*, R. Palgrave MacMillan, 251–287.

Jakobson, Roman
1963[1966] Implications of language universals for linguistics. In Joseph H. Greenberg (ed.), *Universals of language.* 1966. Cambridge, MA: MIT Press. (First edition 1963). 263–278. 2d ed.
1965[1971] Quest for the essence of language. In *Selected Writings*, vol. II. The Hague: Mouton, 345–359. (Originally published in *Diogenes* 51, 1965).
1971   Relationship between Russian stem suffixes and verbal aspects. In *Selected Writings*, vol. II. The Hague: Mouton, 198–202.

Johnston, Ray
    1981   Conceptualizing in Nakanai and English. In Franklin, Karl (ed.), *Syntax and Semantics in Papua New Guinea Languages*. Ukarumpa, Papua New Guinea: SIL, 212–224.

Koptjevskaja-Tamm, Maria
    1996   Possessive noun phrases in Maltese: Alienability, iconicity and grammaticalization. *Rivista di Linguistica* 8(1), 245–274.

Lakoff, George and Mark Johnson
    1980   *Metaphors We Live By*. Chicago: University of Chicago Press.

Langacker, Ronald
    2000   The meaning of of. In *Grammar and Conceptualization*. Berlin: Mouton de Gruyter, 73–90.

Lee, David
    2001   *Cognitive Linguistics: An Introduction*. Melbourne: Oxford University Press.

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.

Lehmann, Christian
    1974   Isomorphismus im sprachlichen Zeichen. In Seiler, Hansjkob (ed.), *Linguistic workshop II: Arbeiten des Kölner Universalienprojekts 1973/4*, (Structura 8). München: Fink, 98–123.

Levinson, Stephen C.
    2000   *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge/MA: MIT Press.

Maran, L. R. and J. R. Clifton
    1976   The causative mechanism in Jinghpaw. In Shibatani, Masayoshi (ed.), *The Grammar of Causative Constructions*. New York: Academic Press, 443–458.

Martin, Jack B.
    2000   Creek voice: Beyond valency. In Dixon, R. M. W. and Alexandra Y. Aikhenvald (eds.), *Changing Valency*. Cambridge: Cambridge University Press, 375–403.

Matthews, Peter
    1991   *Morphology*. 2nd ed. Cambridge: Cambridge University Press.

Mayerthaler, Willi
    1981   *Morphologische Natürlichkeit*. Wiesbaden: Athenaion.
    1987   System-independent morphological naturalness. In Dressler et al., *Leitmotifs in Natural Morphology*. (Studies in Language Companion Series 10). Amsterdam: Benjamins, 25–58.

McKay, Graham R.
    1996   Body parts, possession marking and nominal classes in Ndjébbana. In Chappell, Hilary and William McGregor (eds.), *The Grammar of Inalienability*. Berlin: Mouton de Gruyter, 293–326.

Mel'čuk, Igor' A.
    1967   K ponjatiju slovoobrazovanija. *Izvestija Akademii Nauk SSSR*, serija literatury i jazyka 26 (4), 352–362.

Moravcsik, Edith A.
    1978   Reduplicative constructions. In Greenberg, Joseph H. (ed.), *Universals of Human Language*. Vol. 3. *Word Structure*. Stanford: Stanford University Press, 297–334.

Newmeyer, Frederick
   1992   Iconicity and generative grammar. *Language* 68, 756–796.
Nichols, Johanna
   1988   On alienable and inalienable possession. In Shipley, William (ed.), *In Honor of Mary Haas*. Berlin: Mouton de Gruyter, 557–609.
Osthoff, Hermann
   1899   *Vom Suppletivwesen der indogermanischen Sprachen*. Heidelberg: Hörning.
Plank, Frans
   1979   Ikonisierung und De-Ikonisierung als Prinzipien des Sprachwandels. *Sprachwissenschaft* 4, 121–158.
Ronneberger-Sibold, Elke
   1980   *Sprachverwendung–Sprachsystem: Ökonomie und Wandel* (Linguistische Arbeiten 87). Tübingen: Niemeyer.
   1988   Entstehung von Suppletion und Natürliche Morphologie. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 41, 453–462.
Saksena, Anuradha
   1982   Contact in causation. *Language* 58, 820–831.
Taylor, John R.
   2002   *Cognitive Grammar*. Oxford: Oxford University Press.
Thompson, Chad
   1996   On the grammar of body parts in Koyukon Athabaskan. In Chappell, Hilary and William McGregor (eds.), *The Grammar of Inalienability*. Berlin: Mouton de Gruyter, 651–676.
Tiersma, Peter
   1982   Local and general markedness. *Language* 58, 832–849.
Wälchli, Bernhard
   2005   *Co-compounds and Natural Coordination*. Oxford: Oxford University Press.
Waugh, Linda R.
   1982   Marked and unmarked: A choice between unequals. *Semiotica* 38, 299–318.
Welmers, Wm. E.
   1973   *African Language Structures*. Berkeley: University of California Press.
Wilson, Peter J.
   1992   *Una descripción preliminar de la gramática del Achagua (Arawak)*. Bogotá: Asociación Instituto Lingüístico de Verano.
Witkowski, S. R. and Cecil H. Brown
   1983   Marking reversal and cultural importance. *Language* 59, 569–582.
Wright, Saundra Kimberly
   2001   Internally caused and externally caused change of state verbs. Ph.D. dissertation, Northwestern University, Evanston, IL.
Zepeda, Ofelia
   1983   *A Papago Grammar*. Tucson: University of Arizona Press.
Zipf, George K.
   1935   *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. Boston: Houghton Mifflin.
Zwicky, Arnold
   1978   On markedness in morphology. *Die Sprache* 24, 129–143.