Corpus data in a usage-based cognitive grammar

Joybrato Mukherjee

University of Giessen

Abstract

The present paper is intended to bridge the long-established gap between corpus-based research into actual language use on the one hand and cognitive models of the abstract language system (in terms of speaker's 'competence') on the other. For this purpose, a very useful, non-generative framework is provided by Langacker's usage-based cognitive grammar. In general, the consideration of corpus data in cognitive grammar leads to an innovative and realistic model of speakers' linguistic knowledge, i.e. a model which is data-oriented and frequency-based, functionalist and lexicogrammatical in nature. This theoretical from-corpus-to-cognition approach will be illustrated by discussing corpus data on the use of the ditransitive verb GIVE and by sketching out how the data may be included in a truly usage-based model of the lexicogrammar of GIVE.

1. Introduction: cognitive grammar and corpus data

In principle, generative models of language cognition have always been based on what Langacker (1987, 1999, 2000) has repeatedly called the 'rule/list fallacy', that is a clear distinction between a set of syntactic rules on the one hand and a list of lexical entries on the other. This is particularly true of the recent version of generative grammar, the Minimalist Program, which is guided by strict economy conditions (cf. Chomsky 1995). Langacker (2000), on the other hand, suggests a fundamentally different approach to language cognition:

> There is a viable alternative: to include in the grammar both the rules and instantiating expressions. This option allows any valid generalizations to be captured (by means of rules), and while the descriptions it affords may not be maximally economical, they have to be preferred on grounds of psychological accuracy to the extent that specific expressions do in fact become established as well-rehearsed units. Such units are cognitive entities in their own right whose existence is not reducible to that of the general patterns they instantiate.
>
> (Langacker 2000: 2)

Such 'well-rehearsed units', comprising routinised patterns of specific instantiating expressions, cut across the lexicon-syntax boundary. What is more, they are established due to the recurrent use of specific lexical items in a given

construction or, from a complementary perspective, the frequent use of specific constructions with a given lexical item. In Figure 1, one of the examples given by Langacker (1999) is shown. It visualises how combinations of specific constructions, e.g. the basic ditransitive pattern [[V][NP][NP]], and specific ditransitive verbs such as GIVE and SEND are entrenched as cognitive entities in their own right. The left-hand circle refers to the 'constructional network' of the constructional schema [[V][NP][NP]], while the right-hand circle depicts the 'lexical network' of the verb SEND. At the intersection of the two circles, the resulting pattern can be found, i.e. [[send][NP][NP]].

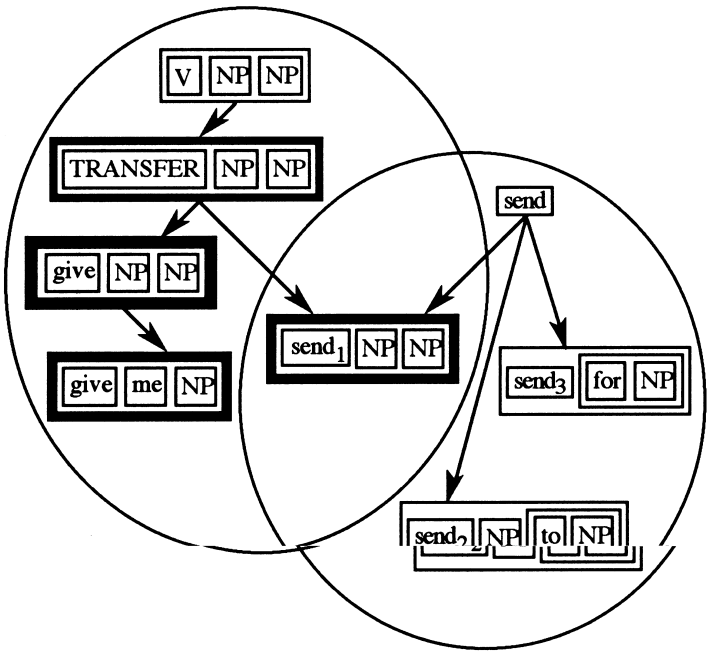


Figure 1. Lexical and constructional networks in cognitive grammar (Langacker 1999: 123)

In Figure 1, the conceptual similarities between Langacker's cognitive grammar and corpus-linguistic approaches are obvious, even though the objects of inquiry, namely language cognition and language use respectively, are no doubt different. Specifically, the concept of lexical and constructional networks (representing lexicogrammatical entities) could be easily mapped onto the notion of 'lexicogrammatical pattern' as it is described by Hunston and Francis (2000):

> The patterns of a word can be defined as all the words and structures which are regularly associated with the word and which contribute to its meaning. [...] as a word can have several different patterns, so a

> pattern can be seen to be associated with a variety of different words. This is the opposite side of the coin.
>
> (Hunston and Francis 2000: 37, 43)

In effect, such lexicogrammatical patterns are at the basis of cognitive grammar.[1]

Another cross-correspondence between cognitive grammar and corpus-based pattern grammar is related to the fact that Langacker (1987) considers his model to be 'usage-based', which is defined as follows:

> Substantial importance is given to the actual use of the linguistic system and a speaker's knowledge of this use; the grammar is held responsible for a speaker's knowledge of the full range of linguistic conventions, regardless of whether these conventions can be subsumed under more general statements. [It is a] nonreductive approach to linguistic structure that employs fully articulated schematic networks and emphasizes the importance of low-level schemas.
>
> (Langacker 1987: 494)

Special emphasis is placed here on 'the actual use of the linguistic system'. In general, this clearly mirrors the Hallidayan assumption that system and use are inseparable because language use instantiates the system (cf. Halliday 1991: 31). More specifically, a model of language cognition should be able to account for actual usage, so that the model has to be based on actual use in the first place. It is exactly here that corpus data may play a major role in refining cognitive grammar and increasing its usage-basedness: corpora are samples of 'actual use of the linguistic system'; the 'schematic networks', 'low-level schemas' and 'linguistic conventions' correspond largely to the lexicogrammatical patterns and routines that can be identified by drawing on corpus data.

Table 1. Corpus-based insights into actual language use and their implications for a usage-based cognitive grammar

some typical features of language use as attested in corpora	implications for a usage-based cognitive grammar
• linguistic forms differ with regard to frequency and distribution	⇨ knowledge about these frequencies and distributions should be part of the model
• language use is to a large extent based on recurrent patterns of different kinds	⇨ the model should account not only for linguistic creativity but also for linguistic routine
• quantitative findings can often be explained by considering functional and context-dependent principles/factors	⇨ these principles/factors are part of speakers' linguistic knowledge and should be included in the model
• lexical and grammatical choices are interdependent	⇨ lexicogrammatical patterns should be at the basis of the model

Table 1 summarises four typical and general features of actual language use as attested in corpora. In the right-hand column of Table 1, the implications of these corpus-based findings for a truly usage-based cognitive grammar are indicated. While, in a sense, lexicogrammatical patterns have always been at the basis of cognitive grammar (cf. Figure 1), it seems to me that the first three aspects in Table 1 have so far been neglected by proponents of a usage-based cognitive grammar. In particular, existing models based on cognitive grammar include neither actual frequencies of linguistic forms nor the principles and factors that may lead language users to choose from a variety of options a specific form in a given context. This kind of information, however, can be easily obtained from corpus data. I would contend that the incorporation of this corpus-based information in cognitive grammar would certainly increase the usage-based quality of cognitive models. This theoretical approach will be exemplified in the following section by delving more closely into the patterns of the ditransitive verb GIVE in the British component of the International Corpus of English (ICE-GB, cf. Nelson et al. 2002) and by deriving from the data a genuinely usage-based cognitive model of the lexicogrammar of GIVE.

2. The relevance of corpus data to a usage-based cognitive grammar: the case of GIVE

Table 2 provides an overview of the frequency of all GIVE-patterns in ICE-GB.[2] In the following, I will be concerned with the eight most frequent patterns only; they are given in boldface in Table 2. These eight patterns alone account for more than 91% of all occurrences of GIVE in ICE-GB. In a sense, then, it is these eight patterns in particular that should be taken into consideration in a model of routinised patterns in language use, because all the other patterns are only sporadically used. Picking up on Aarts's (1991) distinction between 'performance' and 'language use', this section is thus intended to abstract away from the entirety of performance data a model of language use that accounts for frequent lexicogrammatical routines in using GIVE.

Generally speaking, type I represents the basic ditransitive pattern with both objects realised as noun phrases. I have little to say about this pattern since it can be regarded as the default case both quantitatively and structurally. Thus, the focus here should be on the reasons why language users opt for other patterns than this default pattern in specific contexts, i.e. on significant 'principles of pattern selection' (cf. Mukherjee 2001).

For type I b, one specific factor can be easily identified. This type tends to be used whenever the direct object has already been activated in the preceding text because it is part of a previous pattern. As shown in (1), this explanation accounts for some 83% of all cases of type I b. The examples in (2) to (4) nicely illustrate the fact that, generally speaking, a preceding pattern in the text (e.g. *the... the...*, *grateful for sth.*, *thank sb. for sth.*) predetermines to a large extent the following GIVE-pattern by providing the initial slot (and element) for the next

pattern.[4] In the examples, the preceding pattern is given in italics, and the overlapping GIVE-pattern is underlined.

Table 2. Frequency of GIVE-patterns in ICE-GB[3]

Type	Pattern	Sum	Freq.
I	(S) GIVE [O_i:NP] [O_d: NP]	404	38.0%
I a	(S) GIVE [O_d: NP] [O_i:NP]	1	0.1%
I b	[O_d: NP (antecedent)] (rel. pron.) [S] GIVE [O_i:NP]	23	2.2%
I c	[O_i:NP (antecedent)] (rel. pron.) [S] GIVE [O_d: NP]	2	0.2%
I d	[O_d: NP (fronted)] [S] GIVE [O_i:NP]	1	0.1%
Miscellaneous		10	0.9%
IP	[S < O_i active] BE *given* [O_d:NP] (*by*-agent)	84	7.9%
IP b	IP with [O_d:NP (antecedent)]+ rel. clause/past participle	12	1.1%
II	(S) GIVE [O_d:NP] [O_i:PP (to...)]	123	11.6%
II a	(S) GIVE [O_d:NP] [O_i:PP (for...)]	4	0.4%
II b	[O_d:NP (antecedent)] (rel. pron.) [S] GIVE [O_i:PP (to...)]	7	0.7%
II c	(S) GIVE [O_i:PP (to...)] [O_d:NP]	2	0.2%
Miscellaneous		6	0.6%
IIP	[S < O_d active] BE *given* [O_i:PP (to...)] (*by*-agent)	23	2.2%
IIP b	IIP with [S<O_d (antecedent)]+ rel. clause/past participle	17	1.6%
Miscellaneous		2	0.2%
III	(S) GIVE [O_d:NP] ~~O_i~~	247	23.2%
III b	[O_d:NP (antecedent)] (rel. pron.) [S] GIVE	16	1.5%
Miscellaneous		3	0.3%
IIIP	[S < O_d active] BE *given* ~~O_i~~ (*by*-agent)	38	3.6%
IIIP b	IIIP with [S<O_d (antecedent)]+ rel. clause/past participle	28	2.6%
IV	(S) GIVE ~~O_i~~ ~~O_d~~	10	0.9%
Miscellaneous		1	0.1%
Total		1064	100%

(1) I b [O_d: NP (antecedent)] (rel. pron.) [S] GIVE [O_i:NP]

part of a previous pattern
(19 of 23 cases = 82.6%)

(2) But it then means that *the more things they put on the menu the tinier <u>the amount</u>* <u>they give you</u> <ICE-GB:S1A-018 #24:1:B>

(3) I would anticipate doing one or two units per year and would be *grateful for <u>any financial assistance</u>* <u>that the college could give me</u>
<ICE-GB:W1B-022 #152:13>

(4) I must *thank you, Simon and your parents 'officially' for the slow cooker and table cloth* you gave us for our wedding <ICE-GB:W1B-004 #12:1>

For the passive type IP, there are many factors that seem to play a role in the process of pattern selection. The cluster of relevant factors is summarised in (5). It is not at all surprising that in more than 96% of all instances the *by*-agent is left out. An important reason for choosing type IP thus lies in the optionality of the agent. Additionally, two further factors seem to be responsible for the fact that the recipient (corresponding to the indirect object in the default active type-I pattern) is placed in the initial slot, thus serving as the grammatical subject. First, this pattern tends to be chosen whenever the direct object is significantly heavier than the initial element and is therefore placed in final position according to the 'principle of end-weight' (cf. Quirk *et al*. 1985: 1362). The correlation between weight and pattern selection is illustrated in examples (6) and (7). This factor alone accounts for 50% of all 84 cases. Second, in some 10% of all cases it is the recipient that has already been activated before and is thus taken up as the first element in the type-IP pattern. This is in line with the 'principle of end-focus' (cf. Quirk *et al*. 1985: 1357) according to which there is a general tendency to place given information before new information. In examples (7) to (9), the previously activated element which is part of (or provides the initial element for) the GIVE-pattern at hand is italicised.

(5) IP [S < O_i active] BE *given* [O_d:NP] (*by*-agent)

activated before/ taken up (8 of 84 cases = 9.5%)	heavy (42 of 84 cases = 50.0%)	left out (81 of 84 cases = 96.4%)

(6) [...] Margaret Thatcher cannot be given all the credit for our record levels of radioactivity both at sea and on land <ICE-GB:W2B-014 #11>

(7) and rather nastily *she* had been tied to a chair until she was fourteen by her blind mother and never actually given any form of uhm sound or language communication <ICE-GB:S1B-003 #102>

(8) After all Saddam Hussein uh led *his people* they although they were not given much choice in the matter in an eight year war [...] <ICE-GB:S1B-035 #66>

(9) *The Italian peoples* were bound to fight in Rome's wars at their own charge [...] Some peoples were actually given Roman citizienship [...] <ICE-GB:W2A-001 #006/8>

In type II again, it is a cluster of factors that can be shown to play a role to different extents in the process of pattern selection. As shown in Table 2, type II differs from the basic type I in that the indirect object is realised as a prepositional phrase (introduced by *to*) and placed after the direct object. Heaviness

of the final element is again a relevant factor since it is involved in 39 of 123 cases (= 31.7%). But there is another factor that seems to be even more important for language users' choice of this pattern in given contexts, namely the lexical item in direct-object position. The lexical items that are frequently used as direct objects in type II can be grouped into three major types. In nearly 25% of all cases, it is the pronoun *it*. The second group contains words which, broadly speaking, are habitually associated with the preposition *to* according to the pattern information in the corpus-based *Macmillan English Dictionary* (cf. Rundell 2002). This group thus includes nouns such as *access*, *answer* and *reaction* which have a pattern themselves that could be described in COBUILD manner (cf. Sinclair 1995) as 'N *to* n'. This group also includes nouns (such as *name*) which are part of larger verb-dependent patterns containing the sequence 'N *to'* (e.g. *give one's name to sth.* and *put a name to*). Whether it is due to small-scale patterns of the noun itself or due to large-scale patterns of a verb including the noun-*to* sequence, the overall effect is the same: the noun at hand and the preposition *to* tend to co-occur fairly frequently in actual usage. The third group includes words that are so closely associated with this pattern that the resulting word-pattern combinations may be regarded as lexically stabilised idioms, e.g. *give birth to sb./sth.* and *give rise to sb./sth.*: here the type-I pattern no longer provides a genuine alternative. These three groups of lexical items in direct-object position account for some 75% of all occurrences of this pattern. The two factors that are responsible for the preference of the type-II pattern over others – namely weight of the indirect object and lexis of the direct object – are summarised in (10). The examples given in (11) to (13) are intended to illustrate the second factor in particular. In all three examples, the lexical items in direct-object position that seem to trigger off the selection of the type-II pattern are italicised. Additionally, the relevant small-scale *to*-pattern of the noun in direct-object position is in boxes in examples (12) and (13).

(10) II (S) GIVE [O_d:NP] [O_i:PP (to...)]

heavy
(39 of 123 cases
= 31.7%)

frequent lexical items in O_d-position (91 of 123 cases = 73.9%):

1. *it* (30 of 123 cases = 24.4%)
2. words that are associated with the preposition *to* in general, e.g. *access*, *aid*, *answer*, *attention*, *comfort*, *consideration*, *credence*, (*one's*) *name*, *reaction*, *reply*, *substance* (18 of 123 cases = 14.6%)
3. words bound to type II in lexically stabilised idioms, e.g. *give birth* / *rise* / *thought* / *way to sb./sth*. (43 of 123 cases = 34.9%)

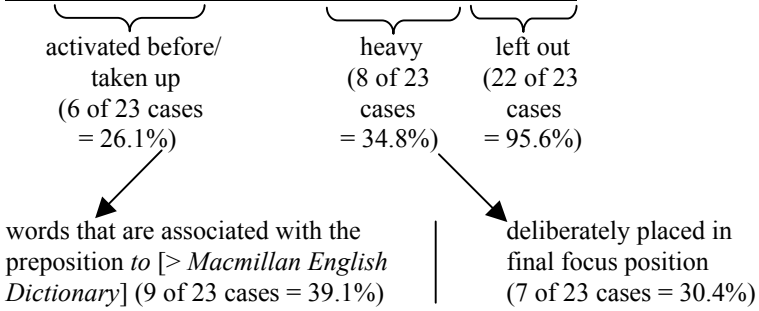
(11) so we can have an acid and alcohol and give *it* to the esterase which is a useful product <ICE-GB:S2A-034 #39>

(12) A clutch of opinion polls gave *comfort* to both sides in the simmering civil war yesterday <ICE-GB:W2C-006 #76>

(13) but when you follow that through you've got the means to give *rise* to a change in the method of accounting that's adopted in the company <ICE-GB:S2A-037 #122>

Type IIP is the passive form that can be derived from the type-II pattern. Note that the systematic correspondence between the two patterns stems from the fact that in both cases the indirect object is realised as a *to*-phrase. As shown in (14), all the kinds of factors that are involved in the choice of the passive pattern IP are also involved in type IIP: previous activation of the initial element (6 of 23 cases = 26.1%), heaviness of the post-verbal element (8 of 23 cases = 34.8%), and the frequent omission of the *by*-agent (22 of 23 cases = 95.6%). In the light of the 23 cases at hand, we may also assume that two further factors may at times tip the balance in favour of type IIP: (i) the need to put the indirect object in focus according to the principle of end-focus; (ii) the use of a lexical item (e.g. *thought*) in the passive subject which may be habitually associated with the preposition *to*. The cluster of all five factors and their explanatory power in quantitative terms are summarised in (14). Example (15) illustrates the relevance of the principle of end-focus (here in order to contrast the two italicised elements at the end of the two dependent clauses). Example (16) refers to the influence of the lexical item in subject position on the selection of the type-IIP pattern.

(14) IIP [S < O_d active] BE *given* [O_i:PP (to...)] (*by*-agent)



(15) At the start of the conflict you said more time should have been given *to sanctions* but now you're saying that more time should have been given *to pursue those diplomatic initiatives* <ICE-GB:S2B-018 #92-94:2:D>

(16) It is not clear that enough *thought* has been given *to* the consequences of these proposals for the movement of traffic outside the areas immediately affected <ICE-GB:W1B-027 #41:4>

What all type-III patterns have in common is the fact that the indirect object is omitted. Note that in many of these cases, the verb GIVE is not parsed as 'ditransitive' but as 'monotransitive' in ICE-GB. For various reasons, how-

ever, I regard all instances of GIVE as examples of ditransitivity. Without going into details about this theoretical issue, it is necessary to point out that my approach to ditransitivity is inherently lexico-semantic (rather than, say, merely syntactic) in nature. In other words, the underlying assumption is that the verb GIVE always triggers what Goldberg (1995) calls the 'ditransitive construction' at a cognitive level. However, as pointed out by Goldberg (1995) herself, not all argument roles of the process of giving (i.e. the 'agent', the 'recipient' and the 'patient') need to be explicitised at the level of syntactic surface structure. Among many others, Matthews (1981), Jackson (1990), Newman (1996) and Biber *et al*. in the *Longman Grammar of Spoken and Written English* (1999) show that specific elements may be left out because, for example, they can be recovered from the context or can be inferred from world knowledge. In a sense, then, GIVE should be regarded as a ditransitive verb in all its occurrences from a cognitive-semantic point of view because it is bound to evoke an event type which includes three argument roles, even though some 'implicit' argument roles may not be explicitised.

Type III is the second most frequent pattern of GIVE in ICE-GB. It does not come as a surprise that corpus data reveal that this pattern is used whenever the recipient is indeed recoverable from the context or when its specification is irrelevant in a given context. In fact, this pertains to all 247 cases at hand. Furthermore, the pattern tends to be chosen whenever specific lexical items are used in direct-object position. That is to say, the omission of the indirect object seems to be linked to lexical items which may imply no need for any specification of the recipient because it is only the mere existence of a recipient that is relevant but not the particular kind of recipient.[5] In (17), those 21 words are listed that are used at least three times as direct objects in the type-III pattern of GIVE. Note that these 21 words alone account for roughly 50% of all cases of this pattern.[6] As in the type-II patterns, it thus seems as though specific lexical items may serve as pointers to the type-III pattern. Some examples are given in (18) to (20).

(17) III (S) GIVE [O_d:NP] ~~O_i~~

contextually recoverable / specification irrelevant
(all 247 cases = 100.0%)

frequent lexical items (≥3):
account (9), *birth* (3), *command* (3), *detail* (10), *effect* (3), *evidence* (20), *example* (9), *hint* (3), *impression* (10), *indication* (7), *information* (5), *instruction* (5), *it* (4), *lecture* (8), *message* (3), (*sb.'s*) *name* (4), *notice* (3), *reason* (3), *signal* (3), *talk* (3), *way* (6) (124 of 247 cases = 50.2%)

(18) So for instance we can give a very nice *account* of coarticulation [...] <ICE-GB:S2A-030 #12>

(19) It helps to clarify the poet's ambiguous comments beforehand by giving an actual *example* of what he means <ICE-GB:W1A-018 #33>

(20) And it's that sort of thing that gave the *impression* which I'm sure he was trying to do <ICE-GB:S1B-038 #103>

From type III, the passive form IIIP can be derived. Again, the optionality of the *by*-agent is most important for the process of pattern selection because it is omitted in 31 out of 38 cases (81.6%). Additionally, specific lexical items in the subject position (i.e. the subjectivised direct objects of the type-III pattern) tend to be closely associated with this pattern. That is to say, not only is the type-IIIP pattern used whenever neither the agent nor the recipient needs to be explicitised but also when particular words refer to the patient of the action. In (21) those words are listed that occur at least twice in this pattern in ICE-GB, accounting for some 45% of all instances. Some of them are exemplified in (22) to (24).[7]

(21) IIIP [S < O_d active] BE *given* ~~O_t~~ (*by*-agent)

left out (31 of 38 cases = 81.6%)

recurrent lexical items (≥2):
approval (2), *limit* (2), *information* (2), *detail* (7), *time* (2), *directions* (2)
(17 of 38 cases = 44.7%)

(22) He's called Malachi in the opening verse but no biographical *information* is given about him <ICE-GB:S2A-036 #78>
(23) uh *directions* are given from Ushant uh from the Scillies uh from the South coast of Ireland down to Cape Ortegal or Finisterre <ICE-GB:S2B-043 #20>
(24) More specific implementation *details* are given at the end of the report <ICE-GB:W1A-005 #5:1>

The last pattern to be mentioned is type IIIP b. This type is similar to pattern I b in that the patient (i.e. the subjectivised direct object) serves as an antecedent to which a relative clause or a past participle construction refers back. As shown in (25), there is again a clear tendency for language users to choose this pattern with a fronted antecedent whenever this antecedent has already been part of a preceding pattern in the text at hand. Examples (26) to (28) illustrate this dependency on the previous pattern (given here in italics: *know of sth.*, *consider sth.*, *trace on to ... sth.*) the last element of which provides the starting-point for the subsequent GIVE-pattern. It should be noted in passing that the *by*-agent is not as frequently omitted as in all other passive patterns mentioned so far. In fact, in more than one third of all cases (10 of 28 cases = 35.7%), the agent is stated explicitly. Thus, the optionality of the *by*-agent as such turns out to be less forceful a factor for this particular passive form.

(25) IIIP b IIIP with [S<O_d (antecedent)] + relative clause/past participle

part of a previous pattern (16 of 28 cases = 57.1%)

with or without *by*-agent (10 vs. 18 cases = 35.7% vs. 64.3%)

(26) and he will also *know of* *the increased uh support* given uh in the uh announcement last week by my right honourable friend the Social Security Secretary <ICE-GB:S1B-056 #46:1:B>

(27) [...] it also is of relevance when *considering* *the evidence* given by Mr Holt because there is a clear conflict [...] <ICE-GB:S2A-068 #40:1:A>

(28) But what I have simply done is to *trace on to a map* *the directions* that are given which give you some indication [...] <ICE-GB:S2B-043 #19:1:A>

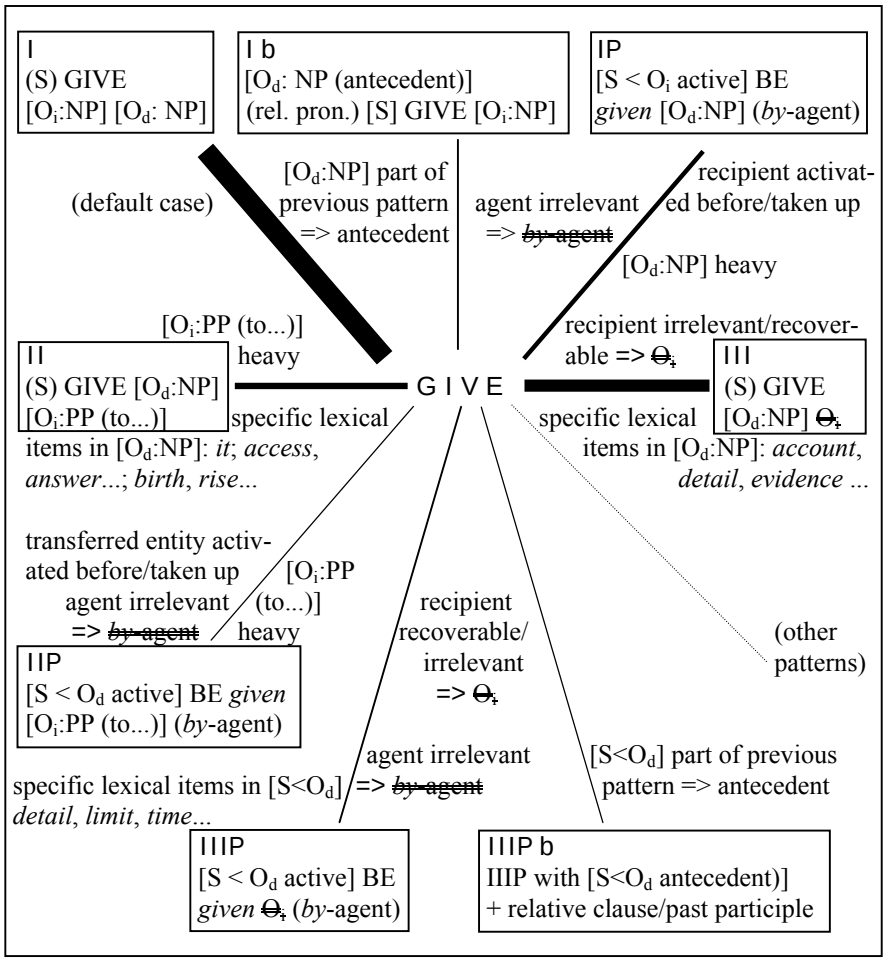


Figure 2. A usage-based cognitive model of the lexicogrammar of GIVE

The actual use of the eight most frequent GIVE-patterns and the relevant principles of pattern selection as described above provide an empirically sound basis for a truly usage-based cognitive model of the lexicogrammar of GIVE. Such a usage-based model on the basis of ICE-GB is visualised in Figure 2.

In two regards, the tentative model suggested in Figure 2 is more elaborated and more 'usage-based', as it were, than traditional lexical networks in cognitive grammar (as, for example, shown in Figure 1). Firstly, the thickness of the lines between GIVE and its patterns depends on the frequency of GIVE in each pattern. Figure 2 thus puts into operation what has been suggested, among others, by Lamb (2002: 91), namely that different "[d]egrees of entrenchment [can be] accounted for by variability in the strengths of connections." Secondly, at all lines connecting GIVE and its patterns there is information on why a particular pattern is used in a given context. Such principles of pattern selection can be identified only by looking at large amounts of natural data in context and have so far not been taken into consideration in cognitive grammar. More specifically, traditional network models in cognitive grammar have focused on what is structurally possible. Corpus data, however, provide information on what is likely to occur and why. As I have argued elsewhere (cf. Mukherjee 2002), both aspects are part of speakers' linguistic knowledge and should therefore be covered by a truly usage-based cognitive grammar.

3. Conclusions and prospects for future research

The present paper is informed by the belief that corpus linguistics and cognitive linguistics are not at all mutually exclusive but can fruitfully complement each other in developing a genuinely usage-based model of language cognition, i.e. of speakers' knowledge of the underlying language system. A genuinely usage-based model defies the rigid Chomskyan dichotomy between 'competence' and 'performance'.[8] In fact, such a model is intended to bridge the gap between system and use and to mirror speakers' linguistic knowledge along the lines of Hymes's (1972, 1992) concept of 'communicative competence', in which the ability to use linguistic forms and structures idiomatically (e.g. in terms of frequently co-occurring forms) and appropriately (e.g. in terms of pragmatic principles) is integral to speakers' knowledge of the language. This view is closely related to the Hallidayan idea that language use and language system are intricately interwoven, which makes it possible and reasonable to derive from a corpus-based analysis of actual language use a usage-based model of the cognitive entrenchment of the language system. In effect, this approach capitalises on Schmid's (2000: 39) "From-Corpus-to-Cognition Principle: Frequency in text instantiates entrenchment in the cognitive system." In particular, I hope to have shown that lexical network models in cognitive grammar can be refined in two regards by taking into account corpus data: not only is it possible to introduce frequency-based information on different strengths of linkage between lexical items and constructions but also to introduce in the

model context-dependent principles of pattern selection (such as lexico-grammatical co-selections, pragmatic principles and activation statuses of discourse entities). Thus, corpus-linguistic methodology obviously opens up new and promising perspectives in cognitive linguistics.

By including quantitative trends and context-dependent principles of pattern selection in usage-based models of language cognition, future research in this field should try to quantify the influence that each of the relevant factors exerts on the process of pattern selection and to empirically describe the prototypicality of a specific pattern in a given context (cf. Gries's 2001 model of a 'multifactorial analysis'). In order to establish more reliable quantitative trends (in terms of, say, lexical co-selections of a given pattern), it will certainly be useful to analyse larger corpora such as the British National Corpus. From a more theoretical perspective, future research into the refinement of the usage-based model as sketched out in the present paper will have to address the question as to whether the principles of pattern selection should be integrated with each individual lexicogrammatical pattern of a given verb or, alternatively, whether they should best be regarded as a separate subcomponent of a usage-based model. As shown in Figure 1, constructional networks provide, in a sense, mirror images of lexical networks, which begs the question as to whether it is necessary and reasonable to posit separate constructional networks in a usage-based model. While proponents of construction grammar (e.g. Goldberg 1995) place special emphasis on the constructional nature of language cognition, other researchers (e.g. Nemoto 1998) call into question the plausibility of the concept of abstract and entirely delexicalised constructions.

Finally, brief mention should be made of the issue of genre distinctions. In the present paper, the influence that specific genres may exert on the frequency of individual GIVE-patterns has been left out of consideration. Future research into corpus-based cognitive models should certainly delve more closely into the correlations between specific genres and the frequency of linguistic forms. It remains to be seen, though, whether genre-specific factors should best be regarded as full-fledged principles of pattern selection at the centre of a usage-based model or as additional factors on the periphery of such a model.[9]

Notes

1. Note that Langacker (1999: 122) himself states that "lexicon and grammar grade into one another so that any specific line of demarcation would be arbitrary". This description is of course largely reminiscent of the Hallidayan approach to "lexicogrammar as a unified phenomenon, a single level of 'wording', of which lexis is the 'most delicate' resolution" (Halliday 1991: 31-32).
2. It should be noted that the data in Table 2 are based on a manual analysis of all occurrences of GIVE and not on the parsing information included in ICE-GB. The reason why the data were analysed manually is the fact that many

instances of GIVE are not parsed as ditransitive in ICE-GB but, for example, as monotransitive (especially in the case of type-III patterns) or as complex-transitive (especially in the case of type-II patterns). In contrast, I regard all instances of GIVE as examples of ditransitivity on cognitive and semantic grounds (cf. Goldberg 1995 and Newman 1996). It is for this reason that phrasal verbs such as GIVE AWAY, GIVE IN and GIVE UP have not been taken into account, because their semantics tends to be quite different from GIVE. Note also that not all instances of GIVE can be grouped into any of the patterns listed in Table 2. However, such 'miscellaneous' cases are rare and thus of a marginal nature.

3. The pattern formulas are based on the following notational conventions: [...] obligatory element; [...(...)] obligatory element with a specific form/function; (...) optional element; ~~O_t~~/~~O_d~~ clause element which is not part of the lexicogrammatical pattern at the level of syntactic surface structure (although the corresponding argument role is taken to be implicitly evoked by GIVE at a cognitive level).
4. In fact, this is reminiscent of what Hunston and Francis (2000: 211) refer to as 'pattern flow': "Pattern flow occurs whenever a word that occurs as part of the pattern of another word has a pattern of its own."
5. Since, from a lexico-semantic point of view, the existence of a recipient is already inherent in the event type evoked by the ditransitive verb GIVE, there is no need to explicitise the recipient as an indirect object at the level of syntactic surface structure in these cases. For example, in phrases such as *give a lecture* and *give a talk* some kind of recipient is always implied (e.g. an unspecified audience). Accordingly, Newman (1996: 54), in his cognitive study of GIVE, describes such implicit argument roles as 'unfilled elaboration sites'.
6. As a matter of fact, many of the lexical items could be complemented by other items of the same semantic field that also occur in this GIVE-pattern in ICE-GB, e.g. *give a lecture/a talk* (+ *a paper*, *a speech*, *a statement*...), *give instructions* (+ *advice*, *help*, *orientation*...) and *give a message* (+ *an answer*, *an outline*, *a response*, *a warning*...). The important point here is that the lexis in direct-object position is semantically restricted.
7. Note that the analysis of type IIIP is based on 38 instances only. One could easily hypothesise that the list of recurrent lexical items would have been much more similar to the list given for the type-III pattern if some 250 cases had been scrutinised. Here, larger corpora are needed.
8. It is for this reason that the term 'competence' is not used in the present paper. A cognitive model that is based on corpus evidence, as suggested in the present paper, has not much in common with a generative model of competence. Thus, it is not very useful to take over and extend or redefine the term competence, which would automatically lead to terminological confusion (cf. Taylor 1988). Instead, I prefer to speak of a usage-based model of speakers' linguistic knowledge.

9. Note

Mukherjee, J. (2001), 'Principles of pattern selection: a corpus-based case study', *Journal of English linguistics*, 29: 295-314.

Mukherjee, J. (2002), 'The scope of corpus evidence', in: R. Brend, W. Sullivan and A. Lommel (eds.) *LACUS forum XXVIII: what constitutes evidence in linguistics?* Houston, TX: LACUS. 103-114.

Mukherjee, J. (forthcoming), *English ditransitivTj*□.□□□□□ □ □ □.□□□□□ □□□*s*□□□□*Tm*□□□*ı*