

HANS-JÖRG SCHMID

NON-COMPOSITIONALITY AND EMERGENT MEANING OF
 LEXICO-GRAMMATICAL CHUNKS:
 A CORPUS STUDY OF NOUN PHRASES WITH SENTENTIAL
 COMPLEMENTS AS CONSTRUCTIONS

Abstract: The following article discusses nominal constructions with syntactic dependents. The focus lies on so-called “shell nouns”, such as *fact*, *idea* or *problem*, i.e. nouns that can be described as “containers” reducing complex pieces of information expressed by clauses. After describing the relationship between the two linguistic approaches of idiom principle and construction grammar, the article goes on to present a corpus analysis of shell nouns and their respective syntactic complementations, coming to the conclusion that these complementations can be described as constructions, in the sense that the syntactic patterns can be said to convey a meaning that goes beyond the semantic sum of the constituents.

1. Introduction: From idiom principle to construction grammar¹

It was John Sinclair who encapsulated a central insight of 20th century British corpus linguistics in his now well-known *idiom principle*: “The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments” (1991, 110). Rather than being constructed anew from scratch every time on the basis of abstract syntactic rules, as suggested by the complementary *open-choice principle*, sentences appear to be made up of lexico-grammatical units larger than single words, which are considered to be stored holistically in the mental lexicon and can be retrieved, as it were, wholesale in on-going language production and comprehension.

The idiom principle and the open-choice principle differ not only with regard to their predictions concerning the storage, retrieval and size of material in the mental lexicon and the operations involved in sentence production and comprehension, but also with regard to their implications for the lexicon-grammar

¹ I would like to thank Friedrich Ungerer for discussions of earlier versions of this paper, and Sandra Handl, Susanne Handl and Maura Enders for comments on the final one.

interface and the compositionality of meaning. The open-choice principle prevalent in generative and other “traditional” syntactic models considers lexicon and grammar as two distinct modules, with the rules of grammar setting up slots to be filled by lexical elements retrieved from the lexicon. This separation remains valid in principle, even if interaction between the modules is accepted, as manifested, for example, in the theta-component of generative grammars. The idiom principle, on the other hand, requires by definition a close cooperation of grammar and lexicon, as the chunks making up the components of sentences consist of lexical elements pre-arranged in certain grammatical relations.

As far as the meaning of complex linguistic units is concerned, the open-choice principle tacitly assumes the validity of the venerable principle of compositionality. Often attributed to Frege and apparently born from a desire to explain sentence meanings by means of logical rules, this principle states that the meaning of a complex expression “is determined by the lexical meanings of its components, their grammatical meanings and the syntactic structure of the whole” (Löbner 2002, 15). A strong interpretation of this principle entails that the source of all semantic aspects of a sentence can invariably be identified and that there must not be meaning in the sentence over and above the meanings of its parts and the relations between them.²

The importance of more or less variable, but still prefabricated chunks for sentence processing has acquired increasing recognition in recent years and gained new impetus through the emergence of various strands of Construction Grammars.³ In the more cognitively-minded version of Construction Grammar proposed by Goldberg and others, constructions are defined as “learned pairings of form with semantic or discourse function” (Goldberg 2006, 5). More specifically, “any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist” (ibid.). According to this definition, constructions are quasi non-compositional per definition.

Starting out with Fillmore’s and Kay’s work (cf. e.g. Fillmore/Kay/O’Connor 1988; Fillmore 1999) and Goldberg’s earlier book on argument-structure constructions (1995) the literature in the framework of Construction

² Interestingly, Frege does not seem to have spelled out this principle explicitly in any of his writings; quite on the contrary, as the following passage shows, he was acutely aware of the gestalt-like tendency of composite expressions to create extra, emergent meaning: “Sieht man so die Gedanken an als zusammengesetzt aus einfachen Teilen und läßt man diesen wieder einfache Satzteile entsprechen, so wird es begreiflich, daß aus wenigen Satzteilen eine große Mannigfaltigkeit von Sätzen gebildet werden kann, denen wieder eine große Mannigfaltigkeit von Gedanken entspricht. Hier liegt es nun nahe zu fragen, wie der Aufbau des Gedankens geschieht und wodurch dabei die Teile zusammengefügt werden, so daß das Ganze mehr wird als die vereinzelt Teile. (Frege 1923/1976, 72; my emphasis).

³ For recent surveys of and collections of papers on Construction Grammar see Croft and Cruse (2004, 225-290), Fried and Östmann (2004), Östmann and Fried (2005), Fried and Boas (2005), Fischer and Stefanowitsch (2006), Ungerer and Schmid (2006, 244-256).

Grammar has focussed on two types of constructions: on the one hand, rather marginal and specific constructions on the periphery of grammar, e.g. the *let alone* construction (Fillmore/Kay/O'Connor 1988), the *what's X doing Y* construction (Kay/Fillmore 1999) as well as subject-auxiliary-inversion constructions (Fillmore 1999), and basic verbal clause patterns such as Goldberg's argument-structure constructions, on the other. The reason for the latter preponderance is that there is an obvious and grammatically indispensable link between the verb of a clause and its complements, the linguistic description of which has a long, pre-construction-grammar history manifested in Dependency Grammar (Tesnière 1959) and a whole range of valency grammars (cf. Abraham 1978; Allerton 1982; Herbst et al. 2004).

In this paper, the focus shifts to nominal constructions that show dependency phenomena similar to those found with verbs. A set of four interrelated nominal patterns are investigated, all of which consist of an abstract noun and a *that*-clause or *to*-infinitive linked to the noun:

- [a] N + *that*-clause: *the fact that abstract nouns are difficult to pin down ...*
- [b] N + *to*-infinitive: *the idea to illustrate the patterns investigated ...*
- [c] N + BE + *that*-clause: *the problem is that there is a lot to study.*
- [d] N + BE + *to*-infinitive: *the solution is to focus on a bunch of examples.*

Since in patterns [a] and [b] the clauses are immediately attached to the nouns and function as complements governed by the nouns (cf. e.g. Quirk et al. 1985, 1231; Herbst 1988; Biber et al. 1999, 575), patterns of this type are included in Herbst et al.'s (2004) *Valency Dictionary of English* (see e.g. the entries for *failure*, *idea*, *pressure*, *problem*, *talent* and many others). Patterns [c] and [d] are not included, because nouns and clauses are linked by means of the copula, but the semantic relations between nouns and clauses are still very much the same as in patterns [a] and [b]: in both pairs the clauses fill the fairly abstract and unspecific shells provided by the nouns with more specific information (cf. section 2 below).

The aim of this paper is to investigate the contribution of lexical and grammatical meanings to the overall meanings of these patterns. Applying a strictly corpus-based methodology it will be demonstrated that the principle of compositionality is difficult to uphold for these patterns, because there are emergent meaning components that cannot be traced back to the input elements. The argument will be developed in four steps: in the section following this introduction a number of functions shared by the four patterns illustrated in [a] to [d] will be explained; these justify their joint treatment as one construction. The methodological section in 3 will discuss how samples of the patterns in question can be retrieved from corpora and how the reciprocal attraction of nouns and patterns can be measured using the collected data and two simple mathematical tools. Section 4 will provide a cross-tabulation of meanings of nouns and meanings of complements and whole constructions, which demonstrates that it is not always possible to derive the overall meanings of the patterns directly from their parts as suggested by the principle of compositionality. The final section 5 will

present a case study of one type of pattern, viz. N + BE + *to*-infinitive (e.g. *the aim is to ...*), in order to demonstrate how the range of possible meanings and the extent of emergent meaning of these constructions can be determined on the basis of the frequency counts of the material from the corpus. Avenues for the emergence of extra meaning are also discussed in this section.

2. Common functions of the patterns: The shell-content relation

In spite of their variability on the syntactic surface, the four patterns illustrated in the previous section share a number of underlying functional characteristics. These are illustrated in figure 1:

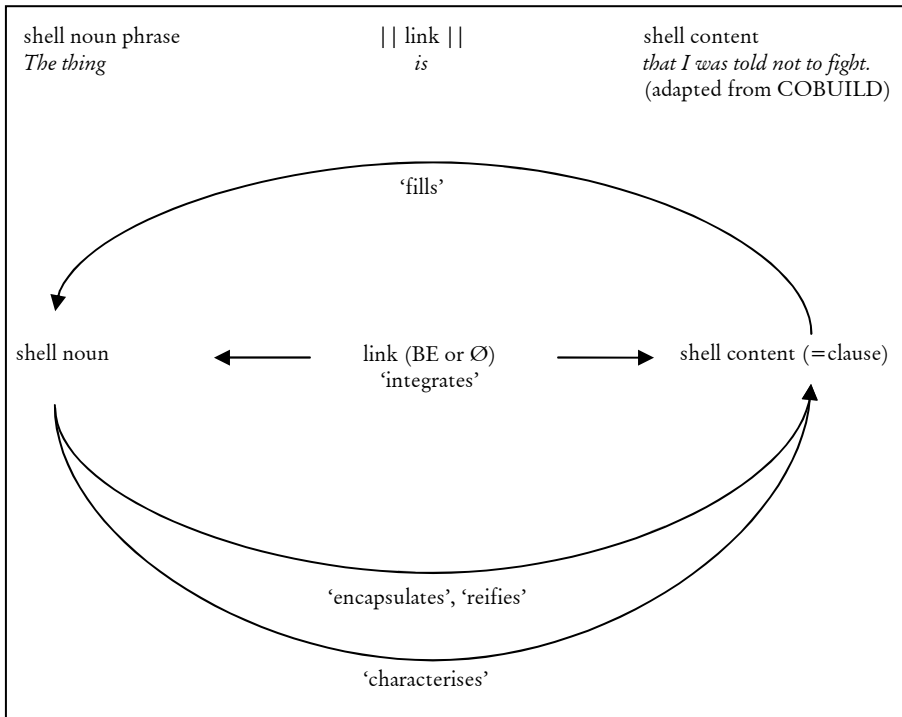


Figure 1. Functions of the shell-content relation illustrated

As already noted by Francis (1986, 36-38) and Conte (1996, 2-4), the crucial cognitive function of the abstract nouns I am concerned with here is to “encapsulate” the complex pieces of information expressed in the sentential complements as nominal concepts. To capture the notion that the nouns serve as containers for the propositions expressed by the clauses, the nouns were dubbed *shell nouns* in Schmid (2000). The result of the encapsulation is that the fact, event or state-of-affairs designated by the clause is conceptually turned into a

“thing” (in Langacker’s Cognitive-Grammar terminology; 1987, 189) or, more technically, *reified* as an apparently neatly-bounded object-like entity. In addition to reification the nouns invariably allow speakers to characterise the content of the clause in a particular way, depending on the choice of noun. Thus the noun *thing* in the example in figure 1, which arguably carries the most neutral characterisation possible, could easily be replaced by other nouns like *fact* (stressing the epistemic status of the shell content), *problem* (attitudinal characterisation), *idea* (characterisation as mental state), or *answer* (characterisation as speech act), to name just a few possibilities. On the other hand, the shell contents, i.e. the complementing clauses, fill the semantically unspecific shell nouns with concrete conceptual content. As already mentioned, the nouns themselves are not sufficiently informative to function on their own, except of course when they are used in anaphoric function, especially as topics in sentence-initial position (e.g. *I was told not to fight. This reminder was absolutely unnecessary*; COBUILD). Shell nouns and shell contents are linked to each other by the copula BE in the patterns N + BE + *that*-clause/*to*-infinitive or by the immediate attachment as NP-postmodifier in the patterns N + *that*-clause/*to*-infinitive. The links guarantee that the nominal and the clausal meanings are integrated and trigger the co-interpretation of the two elements of the relation as one piece of information (for more details see Schmid 2000, 21-31).

In addition to the cognitive and semantic functions of encapsulating/reifying and characterising, shell nouns have qualities that render them extremely helpful tools in the construction of coherent texts. In fact, their textual potential as semantically specific cohesive ties and textual signposts was the main concern of the publications by Francis (1986, see also 1994) and Conte (1996). Since I am concerned with intra-sentential aspects of shell-content constructions here, I will not go further into these aspects (cf. Schmid 2000, 329-359).

3. A glimpse of the corpus evidence

3.1 Data source

The data investigated here from a new theoretical angle were originally collected in 1996 and analysed and described in detail in Schmid (2000). The material was taken from the British section of COBUILD’s *Bank of English*, amounting at that time to 225 million running words from the subcorpora spoken conversation, transcribed BBC recordings, ephemeral texts such as brochures and leaflets, books, magazines, quality newspapers, *English Today*, *The Economist* and *The New Scientist*. It should be borne in mind that with a proportion of about two-thirds of the whole material, texts from media sources make up the lion’s share of this corpus.

3.2 Data retrieval

Intuition, unsystematic observations and dictionary evidence suggest that such nouns as *fact*, *idea*, *aim*, *problem*, *solution* or *answer* occur frequently in shell-content patterns. However, as it was one of the aims of the study reported in Schmid (2000) to inductively identify a maximum number of shell nouns, it was impossible to use the nouns themselves as inputs for corpus queries. Instead the patterns served as query inputs, with the nominal slots being defined by a part-of-speech dummy. Table 1 gives a summary of the query statements and the number of matching lines in the 225m corpus:

Query statement	Number of matching lines in the 225m corpus
Pattern N-cl with various types of adjacent clauses:	
NN+that/CS (NN = noun, CS = conjunction)	280,217
NN+to+VB (VB = base form of verbs)	560,148
Pattern N-be-to	
NN+is+to	28,463
NN+was+to	12,728
NN+has+been+to	962
NN+will+be+to	960
NN+would+be+to	1,421
NN+would+have+been+to	133
Pattern N-be-that	
NN+is+that	37,155
NN+was+that	9,104
NN+has+been+that	433
NN+will+be+that	178
NN+would+be+that	264
NN+would+have+been+that	19

Table 1. Corpus queries and numbers of matches

In addition, analogous patterns with *wh*-clauses were retrieved, yielding such instances as *the question why he didn't come* or *the problem how to define the task*, as well as anaphoric instances of shell nouns, but neither of these realizations of shell nouns is under consideration here. It should be emphasized that the scores given in table 1 reflect the raw yield in response to the corpus queries; this dataset had to be sifted half-automatically to sort out unwanted hits caused, for example, by insertions between shell noun and complement that were superficially licensed by the target pattern, (e.g. *the energy policy of the government is to have no energy policy at all* [COBUILD] where *government* would wrongly be attributed the status of the shell noun because it precedes the copula).

3.3 Assessing the reciprocal association between nouns and patterns

There is a close symbiotic relationship between shell nouns and shell contents. On the one hand, the shell contents require a particular type of semantically unspecific noun to fill the nominal slot in the patterns. Concrete nouns (cf. **the boy is that/to ...* etc.) are as impossible in the pattern as are semantically specific abstract nouns like *democracy*, *madness* or *inflation*. Shell nouns, on the other hand, typically tend to have little semantic content of their own and include a specific semantic gap to be filled. This gap is particularly noticeable in the case of many deverbal speech-reporting or thought-reporting nouns such as *answer*, *statement*, *promise* or *belief*, *feeling* and *knowledge*, where the subsequent *that*-clause plays the same role as an obligatory complement-clause saturating the valency needs of the corresponding verbs (e.g. *she promised that she would come* and *her promise that she would come*).

The challenge for a corpus-based approach is of course to capture this reciprocal relation in quantitative terms. In Schmid (2000, 54-55) two simple mathematical measures, dubbed *attraction* and *reliance*, were proposed for this purpose. Their calculation is given in figure 2:

$$\text{Attraction} = \frac{\text{frequency of a noun in a pattern}}{\text{total frequency of the pattern} \times 100}$$

$$\text{Reliance} = \frac{\text{frequency of a noun in a pattern}}{\text{total frequency of the noun in the corpus} \times 100}$$

Figure 2. Calculating the measures of *attraction* and *reliance*

As the figure shows, *attraction* is calculated by dividing the frequency of occurrence of a noun in a pattern by the frequency of the pattern in the corpus. The result of this division measures the degree to which a pattern attracts a particular noun. Since the denominator of the fraction is the same for all nouns which occur in a pattern, the scores for this value are directly proportional to the raw frequencies of nouns. The measure facilitates the comparison of the relative importance of individual nouns for a pattern. For example, the attraction score of 18.45% for the noun *fact* in the pattern N + *that*-clause means that this noun accounts for almost one fifth of all instances of the pattern and is thus much more important for the pattern than, for example, *possibility* or *warning*, with attraction scores of 1.58% and 1.03% respectively (cf. table 5 in the appendix).

Reliance is calculated by dividing the frequency of occurrence of a noun in a pattern by its frequency of occurrence in the whole corpus. This measure expresses the proportion of uses of nouns in the patterns vis-à-vis other usages of the same noun. High scores for reliance such as 76% for the noun *inability* in the pattern N + *to-infinitive* indicate that the noun depends to a large extent on the pattern for a co-textual environment. The relevance of these two measures will

become clearer with a glimpse of some of the findings provided in the next section.

Compared to more sophisticated recent approaches to measuring the associations of lexical items and constructions proposed by Stefanowitsch and Gries (2003; Gries/Stefanowitsch 2004), the arithmetic applied here may look exceptionally crude and simple-minded. Unlike my approach, Stefanowitsch and Gries's "collostructional analysis" relies on fairly sophisticated distributional statistics known as the Fisher exact test. Essentially, this and other computationally less complex inferential statistical measures like the chi-square distribution are applied to test whether the frequency of co-occurrence of an item and a construction is significantly higher than predicted by chance on the basis of their isolated frequencies (see Kilgarriff 2005 for a critique of this assumption, and Gries 2006a for a follow-up on the critique). The advantage of this obviously lies in the mathematical reliability of these tests (at least as long as the scores, e.g. for the chi-square test, are of the right sizes). However, for a number of reasons the collostructional method is not only less feasible but also less dependable for the aims and data of this study than the measures of attraction and reliance suggested here. For one thing, a considerable number of shell nouns have such high relative frequencies in certain patterns that the Fisher exact test invariably yields a p-value of 0 (as was the case for the verb *give* in Stefanowitsch and Gries's analysis of the ditransitive pattern; 2003, 229). This means that the probability of an accidental association between lexical item and patterns is infinitely low, but it also means that the test does not contribute to a differentiation of the associational strengths of these nouns. Secondly, in order for Fisher exact to be calculable, the total number of constructions in the corpus other than the ones investigated has to be determined in order to serve as a total reference-population. Statistically reasonable as this clearly is, it is not however without its theoretical pitfalls. For example, investigating the construction [N *waiting to happen*] as in *there is an accident waiting to happen*, Stefanowitsch and Gries arrive at the total number of potential constructions in the corpus "by counting the total number of verb tags in the BNC, as [they] are dealing with a clause-level construction centering around the verb *wait*" (2003, 218). This choice of the total reference-population may be somewhat difficult to justify, since it tacitly assumes a potential paradigmatic substitutability of the predicate *waiting to happen* with all verbs in the corpus. Applied to the present study of noun phrases with sentential complements, the same problems arise. The total number of nouns in the corpus would seem just as unsuitable a candidate as, say, the number of all noun phrases or all complex noun phrases. Note that this is not a mathematical problem but a linguistic one. And thirdly, while it has the potential for bi-directional measures of association between words and constructions, collostructional analysis is mainly designed to test the attraction of constructions for certain words (an exception can be found in Gries 2006b, where occurrences of the verb *run* in different patterns are investigated). In the present study, the attraction perspective is quite naturally complemented by the notion of *reliance* which, as ex-

plained above, measures the degree to which nouns depend on the patterns for their occurrence.

3.4 Results

A quantitative summary of the nouns found most frequently in the two patterns is given in the appendix to this paper. Each of the four tables gives two lists of the twenty nouns with the highest scores for attraction and reliance respectively. The headers of the four lists contain information about the number of nouns (*types*) that were found in each of the patterns and the overall number of valid matches of the pattern in the corpus (*tokens*).

Quite predictably the attraction lists are dominated by high-frequency nouns, “feeding”, as it were, the patterns. This is partly due to the fact that frequent nouns are of course more likely to occur in any pattern than infrequent ones, but it can also be traced to strong associative links at least for the top scorers heading the lists. Interestingly, the constructions N + *that*-clause (table 5 in the appendix) and N + BE + *to*-infinitive (table 8 in the appendix) show a similar distribution of nouns in that there are two nouns, *fact* and *aim* respectively, which stand out as particularly strong associates of the patterns with attraction scores of 18.45% and 12.10% respectively.

The lists for reliance, on the other hand, tend to include less frequently used nouns which, in turn, exhibit particularly strong symbiotic relations with the respective patterns. The highest degrees of reliance were found for the nouns *inability*, *t Temerity*, *disinclination* and *willingness*, all boasting a reliance score above 70% (cf. table 6). This means that roughly three quarters of all occurrences of these nouns in the corpus are found in the pattern N + *to*-infinitive. Such high reliance scores are not reached in the two patterns including the copula (cf. tables 7 and 8), presumably because the semantic association between nouns and complements is less close for the complements making up a clause constituent in their own right than for the postmodifying complements directly attached to the head nouns.

In general, the four lists in the appendix undoubtedly demonstrate the semantic associations between certain types of nouns and certain patterns. The following discussion of the emergence of the composite meaning of shell-content relations will begin with a closer look at these semantic associations.

4. How does the composite meaning emerge?

Essentially, shell-content constructions consist of the shell head-nouns (and their determiners and premodifiers), on the one hand, and the complementing clauses on the other. If the principle of compositionality holds true, then the meanings of shell-content constructions would have to be a function of the lexical meanings of the nouns, the grammatical meanings of the complement types

and the syntactic links between the two components. The links presumably contribute no more than the equation, identification or reciprocal integration of the nominal meaning and the meaning of the complement. So to test the principle of compositionality it seems reasonable to cross-tabulate the meanings of nouns with the meanings of complements and check the results of this comparison against the constructional meanings.

4.1 The meanings of shell nouns

On a fairly high level of abstraction six semantic classes of shell noun uses can be distinguished. These are summarized in table 2:

Class	Examples
Factual	<i>fact, thing, point, problem, reason, difference, upshot</i>
Linguistic	<i>news, message, rumour, report, order, proposal, question</i>
Mental	<i>idea, notion, belief, assumption, aim, plan, decision</i>
Modal	<i>possibility, truth, permission, obligation, need, ability</i>
Eventive	<i>act, move, measure, reaction, attempt, tradition, trick</i>
Circumstantial	<i>situation, context, place, area, time, way, approach</i>

Table 2. Semantic classes of shell noun uses

Speakers use *factual* shell nouns to create conceptual shells for “abstract” states of affairs and facts. The types subsumed in this class include semantically neutral nouns like *fact, thing* or *phenomenon*, causal nouns (*reason, result, upshot*), evidential nouns (*evidence, proof, sign*) and, among others, attitudinal nouns like *problem, advantage* or *irony*. It should be emphasized that some of these nouns have the potential to be used in several classes: *problem*, for example, does not always serve as a factual shell noun, as in example (1), but can also be used as an eventive noun, as in example (2):

- (1) A second problem is that water prices do not simply reflect costs. (COBUILD, originally from *The Times*)
- (2) The problem was to safeguard the many civil radar sites [...] from encroachment by property development. (COBUILD: *The New Scientist*)

While in (1) it is possible to insert *the fact* yielding the paraphrase *a second problem is the fact that water prices ...*, example (2) does not accept this but seems to work better with the insertion of *how*: *the problem was how to safeguard ...* The semantic classes listed in table 2 are thus classes of noun uses rather than classes of nouns as such.

Uses of *linguistic* shell nouns allow speakers to portray linguistic activities and their contents and products in a number of ways: they can focus on the propositional content of the reported utterances (*news, message, rumour*) or on their illocutionary force (*report, order, proposal, question*). Transferring Leech’s (1983)

typology of speech act verbs, illocutionary uses can be divided into *assertives* (*statement, report*), *rogatives* (*question, query*), *directives* (*command, suggestion*), *commissives* (*promise, offer*) and *expressives* (*complaint, compliment*).

Mental, i.e. thought-reporting, uses have an analogous distribution, with one sub-class highlighting the conceptual content of a mental state (*idea, notion, theory*) and a second one focussing on the psychological state of the experiencer (e.g. *belief, assumption, aim, plan, decision*). Even the subclasses of illocutionary uses have counterparts in the mental domain, with *creditive* uses (*belief, opinion*) matching aspects of assertives, *dubitative* uses (*doubt, question*) rogatives, *volitional* uses (*aim, plan, wish, dream*) directives as well as commissives, and *emotive* uses (*surprise, fear*) expressives.

Modal shell noun uses are nominal means of expressing modal stances. Following Palmer's (1990) classification of modal verbs, modal shell nouns can be divided into *epistemic* uses with different degrees of certainty (*possibility, probability, certainty*), *deontic* uses with different degrees of obligation (*permission, task, necessity*) and *dynamic* uses (*ability, opportunity, tendency*).

Eventive shell noun uses encapsulate actions and processes, i.e. physically observable dynamic events. In addition to semantically unspecific *general* eventive uses of the nouns *event, change, action*, there are *specific* eventive uses (*move, habit, option*) and *attitudinal* eventive uses analogous to attitudinal factual uses (*trouble, difficulty, success*).

Finally, *circumstantial* uses subsume nouns referring to situations, times, locations, manners of doing things and conditions for doing things. Typical examples are *situation, place, time* and *way*.

4.2 "Meanings" of complements – Evidence from verb complementation

Attributing meanings to dependent grammatical constructions such as *that*-complements or *to*-infinitives is quite problematical because such semantic analyses inevitably tend to rely on the nature of the items governing the complements. The typical procedure, followed in recent semantically-oriented studies on verb-complementation like Wierzbicka (1988, 23-168), Givón (1990, 517-561), Frajzyngier and Jasperson (1991) and Langacker (1991, 438-449), is to deduce the meanings of complement types from their compatibility and co-occurrence restrictions with the governing verbs. This strategy relies on the assumption that a "systematic isomorphism [...] exists between the semantics of the complement-taking verbs, and the syntax of verb-plus-complement constructions" (Givón 1990, 515). While there is undoubtedly a smack of circularity in this assumption, the convergence of the results of the studies on verbs just mentioned and my own work on nominal complementation is so strong that the procedure may be acceptable.

A summary of the accounts of the meanings of complements/complementizers from various sources is given in table 3.

Source	<i>that</i> -clause	infinitive
Wierzbicka (1988)	SAY, KNOW	volition, future orientation (‘this will happen’)
Givón (1990)	cognition-utterance verbs (perception, cognition, mental attitude or verbal utterance)	manipulative verbs
Frajzyngier and Jasperson (1991)	“de dicto” domain	“de re” domain

Table 3. Meanings of complements according to several sources

In spite of conspicuous terminological differences table 3 indicates a convergent view that *that*-clauses convey meanings having to do with fact-like entities that are perceived, thought about and talked about, while *to*-infinitives operate in the domain of events and actions, with a focus on volition, manipulation and future orientation. The data on nouns clearly support this pattern and yield very similar compatibilities and co-occurrence restrictions of nouns and complements. This is summarized in the next section.

4.3 Compatibility of nouns and complements

Table 4 gives an overview of co-occurrence patterns of noun types and complementation types as emerging from the corpus study of more than 420,000 tokens of shell nouns. The shaded areas indicate ungrammatical combinations, such as the complementation of factual nouns like *fact* by infinitives (**the fact to go out* vs. *the fact that he went out*) or of eventive nouns with *that*-clauses (**his attempt that he went out* vs. *his attempt to go out*). The central column gives classes of shell nouns whose sub-classes differ with regard to the type of complement they take. For example, in the linguistic domain, nouns expressing the propositional content of utterances or reporting the illocutionary act of asserting combine with *that*-clauses, while nouns reporting directive and commissive speech acts take infinitives. An analogous pattern is found in the mental domain, with conceptual, creditive and dubitative uses taking *that*-clauses and volitional nouns infinitives, and in the modal domain, where nouns expressing epistemic modality take *that*-clauses and those expressing deontic and dynamic modality infinitives.

<i>that</i> -clause	divided classes of shell nouns	infinitive
factual nouns (<i>thing, fact, phenomenon</i>)		
propositional, assertive nouns (<i>news, argument, message, report, account</i>)	linguistic nouns	directive, commissive nouns (<i>order, request, promise, pledge, threat</i>)
conceptual, creditive, dubitative nouns (<i>idea, notion, belief, knowledge, doubt</i>)	mental nouns	volitional (<i>aim, goal, plan, purpose, desire, determination</i>)
epistemic nouns (<i>possibility, probability, certainty, reality, truth</i>)	modal nouns	deontic and dynamic nouns (<i>permission, job, duty, need, necessity, ability, capacity, opportunity</i>)
		eventive (<i>event, attempt, effort, priority</i>)
		circumstantial (<i>place, time, way</i>)

Table 4. Overview of co-occurrence patterns of noun types and complementation types

This pattern supports the analyses of verbal complementation summarized in section 4.2: *that*-clauses operate in the fact- and belief-related epistemic, linguistic and mental domain, infinitives in the event-related domain expressing volition, future orientation, deontic and dynamic modality as well as circumstantial information, especially means. What the corpus data demonstrate, then, is that there seems to be a semantic match of nominal meanings, on the one hand, and meanings of the two types of complements, on the other.

While the cross-tabulation of nominal and complement meanings may not be too spectacular a finding, it is necessary for the following consideration of compositionality, because it defines three types of test cases:

- 1) Nouns which accept either of the two complement types, e.g. *idea, problem* and *answer*;
- 2) Semantically unspecific eventive nouns like *time, place* and *way*, which acquire specific meanings in shell-content constructions;
- 3) Complements that acquire different meanings depending on the nouns combined with them.

These three test cases will now be studied in turn.

4.4 Test cases

4.4.1 Test case I: Same noun, different meanings

Consider the following set of fabricated examples and their glosses and semantic explanations:

- (3) My idea was to raise money for new books in the library.
'plan', 'aim' – volition, future orientation
- (4) My idea to raise money for new books wasn't bad.
'plan' – volition, future orientation
- (5) My idea that books are important is well known.
'belief' – conceptual, epistemic
- (6) My idea is that books are important for students.
'view' – conceptual, attitudinal
- (7) The problem was to raise money for books in the library.
'difficulty' – future orientation, means
- (8) The problem (how) to raise money for new books ...
'difficulty' – future orientation, means
- (9) The problem is that we haven't got any money.
'unpleasant fact' – factual
- (10) The problem that we haven't got any money ...
'unpleasant fact' – factual
- (11) The answer was to raise money for books in the library.
'solution' – action, volition, means
- (12) The answer is that we haven't got any money.
'reply' – illocution

In all three sets of examples the nominal (and constructional) meanings seem to be determined, or at least influenced, by the meaning of the two types of complements. *Idea* has a volitive meaning when complemented by the infinitive (3 and 4) and an epistemic meaning with the *that*-clause (5 and 6); *problem* acquires a future-oriented meaning with the infinitive (7 and 8) and a factual one with *that*-clauses (9 and 10); and again depending on the complement, *answer* can either have an action-related, future-oriented meaning (as in 11) or one reporting the illocution of an utterance (cf. 12). These data suggest that the constructional meaning is largely determined by the semantics of the complement. One somewhat surprising finding is the emergence of 'means' meanings in the infinitival uses of *problem* and *answer*. In both cases there is a semantic shade of 'this is how it can be achieved', which was not mentioned as part of the meaning of infinitive clauses and is not part of the nominal meaning either. It follows that 'means' is either an emergent part of the constructional meaning or a hitherto overlooked component of *to*-infinitives. We will return to this question below.

4.4.2 Test case II: Unspecific nouns, specific meanings

The second type of test cases seems to support the idea that the nominal and constructional meanings are heavily influenced by the complement meaning. Consider the following three examples of the highly unspecific nouns *time*, *place* and *way* taken from COBUILD:

- (13) The best time to encourage your older child to start caring for a new baby is before the birth.
'the time when it is best possible to ...'
- (14) The place to make the right contacts – an activity long known as “networking” – will be the Net.
'the place where it is possible to ...'
- (15) Senior Republicans admit that the economy is undermining their base of voters. Some believe the only way to win is to press home personal attacks on Mr Clinton ...
'the only way in which it is possible to'

As the glosses suggest, all three examples lend themselves to interpretations in terms of dynamic modality, paraphrasable as 'it is possible to'. The only cues for these modal meanings outside the infinitive complement are the focusing premodifiers *best* in (13) and *only* in (15). The nouns themselves do not seem to have the modal meanings. As regards the infinitives, Quirk et al. (1985) seem to be right in claiming that “postmodifying *to*-infinitive clauses can either have a modal or a nonmodal sense, but the modal interpretation seems to be normal” (1985, 1269). Thus nouns whose meanings can be more or less reduced to single and very general semantic components like 'time' for *time*, 'location' for *place* and 'manner' for *way* seem to acquire modal meanings. This again points to the possibilities given in 4.3, i.e. that the modal meaning is either brought along by the infinitive or emerges from the interaction of head noun and complement (plus perhaps the focusing premodifier). However, if this was the whole story, cases of the type discussed in what follows would be impossible.

4.4.3 Test case III: Same complement, different meanings

The question addressed in the third set of test cases is whether the *to*-infinitive has the same effect in similar constructions. As the following examples and their glosses and paraphrases show, this does not seem to be the case:

- (16) The task is to raise money for new books in the library.
'obligation' – 'What we have to do is raise money for ...'
- (17) The solution is to raise money for new books in the library.
'means' – 'The way we are going to achieve it is to raise money ...'
- (18) The idea is to raise money for new books in the library.
'volition' – 'What we want to do is raise money ...'

- (19) The problem is to raise money for new books in the library.
 ‘means’ + ‘obligation’ – ‘What we have to do is raise money ... but we don’t know how to achieve it’

While all four examples contain *to*-infinitives following the copula, the meanings resulting from the interaction of this complement type with different nouns diverge: *the task is to ...* includes a modal meaning of deontic necessity (‘obligation’), *the solution is to ...* a ‘means’ meaning, *the idea is to ...* a volitive meaning and *the problem is to ...* a combination of deontic necessity and means. Since the overall structure of the construction and the complement type is identical in all four examples, and since, at the same time, the semantic impact of the constructions is not part of the nominal meanings either (e.g. *idea* does not have a volitive meaning, *problem* does not have a deontic meaning), the constructional meaning must be the result of an intricate interaction of noun and complement meanings. The data suggest that, depending on the semantic setup of the noun, different aspects of the semantic complex encodable by *to*-infinitives, i.e. volition, manipulation, and future-orientation, but also obligation and, somewhat unexpectedly, means, can be selected for the constructional meaning.

4.5 Intermediate conclusion

The juxtaposition of shell noun meanings and the meanings of the two types of complements has shown that there is a match between nominal and complement meanings. This results in distinct co-occurrence patterns that appear to be determined by the basic semantic contrast between facts and events, which is reflected in the oppositions factual vs. eventive, epistemic vs. deontic and dynamic modality and propositional content vs. illocutionary act or mental activity. The sequence of test cases presented here provides strong evidence that the constructional meanings are not simply a function of the nominal and complement meanings but include extra semantic elements that emerge from the interaction of the two components of the shell-content construction. The constructional meaning is thus not entirely predictable from the meanings of the construction’s components and the links between them; in short, the meanings of shell-content constructions seem to be non-compositional.

The precise nature of the emergent meaning does not seem to be generally predictable, however. This is due to the fact that both nouns and complement types seem to be open for a range of possible meanings and interact in unpredictable ways which can, however, be explained *post factum*. Whether a construction has emergent modal, means or volitive meanings will depend on the individual combinations of the semantic potential of nouns and complements. The final section of this paper will illustrate how this works for the N + BE + *to*-infinitive construction.

5. Case study: Dominant and emergent meanings of the N + BE + *to*-infinitive construction

While the meanings of shell-content constructions do not seem to be predictable from their components, they are not random either. In fact, if we take the view of Construction Grammar seriously that constructions are pairings of form and meanings, then it should at least in principle be possible to determine the meaning or semantic range of the constructions investigated here. I will concentrate on one variant of shell-content constructions, viz. the N + BE + *to*-infinitive, in order to demonstrate the way corpus data facilitate the semantic analysis of such lexico-grammatical chunks.

The analysis will focus on the 20 top scorers with regard to attraction and reliance as compiled in table 8 in the appendix, but also include some more types illustrating the major categories. Taken together, the 20 nouns in the attraction list account for almost 60% of the ca. 21.000 tokens of the pattern recorded in the corpus. Eleven of these 20 nouns can also be found in the list for reliance, so that the whole table includes 29 noun types which can be assumed to bring to the surface the meanings most strongly associated with the pattern.⁴

The most prominent meanings of the N + BE + *to*-infinitive pattern are represented in figure 3:

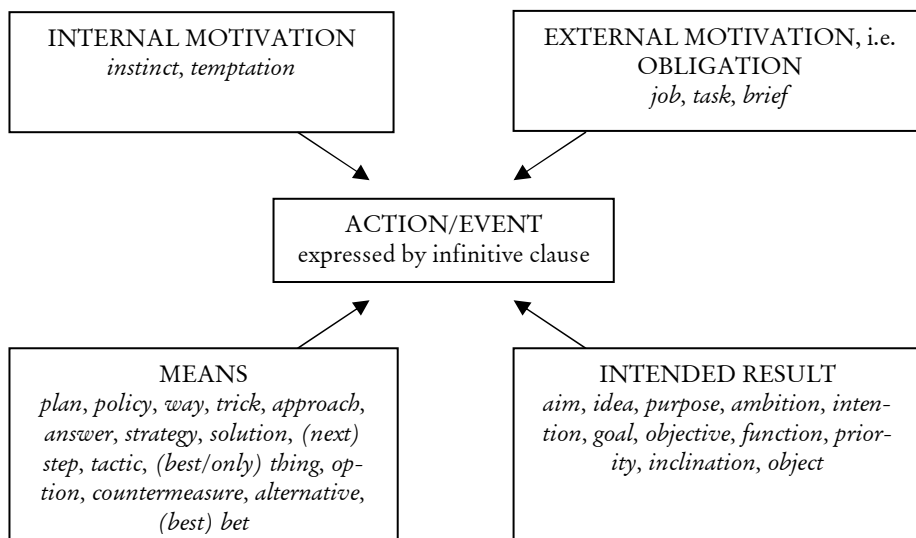


Figure 3. Semantic range of the construction N + BE + *to*-clause

⁴ This is based on the common assumption in corpus linguistics, and especially corpus-based cognitive linguistics, that the most frequently used patterns reflect their linguistically and cognitively dominant meanings and functions (for a recent survey see Tummers/Heylen/Geeraerts 2005).

As the figure suggests, the dominant meanings of the construction revolve around the action or event encoded in the infinitive clause, i.e. what I have called the “shell content”. As in all shell-content constructions (cf. section 2), the function of the noun is to reify and characterise the action in some way. These two functions, as well as the linking function of the copula BE, are shared with, or “inherited” by (cf. Goldberg 1995, 72-73; Fillmore 1999; Croft/Cruise 2004, 270-278; Goldberg 2006, 13-14 *et passim*) the general shell-content construction described in section 2.

The corpus data reveal that four types of characterisations predominate in the case of the N + BE + *to*-infinitive construction: the INTENDED RESULT of the action (highlighted by the nouns *aim*, *idea*, *purpose*, *ambition*, *intention*, *objective*, *function* and *priority*); the MOTIVATION for the action, which can either be INTERNAL (*instinct*, *temptation*, the least prominent meaning) or EXTERNAL, in which case it is expressed as an OBLIGATION (*job*, *task*, *mission*, *duty*, *brief*); and the MEANS with which the goal of the action is meant to be achieved (*plan*, *policy*, *way*, *trick*, *approach*, *strategy*, *solution*, *the best/only thing*, *option*, *alternative*, *the best bet*). The boundaries between these classes are by no means clear and rigid, since nouns like *plan* or *policy* share with the group of INTENDED RESULT a distinct mental meaning, but also focus on the way in which a certain goal is to be reached. Viewed together, these four perspectives represent three key elements of a very general cognitive frame of actions: the motivation for the action, the goal as well as the means, instrument or manner of the action. The other conceptual core components of actions are the action itself and, of course, the agent. While the action is encoded in the infinitive clause, the agent can either be backgrounded altogether since it is syntactically unnecessary to express it (cf. example 20), or be encoded by a possessive determiner or genitive noun preceding the shell noun as is illustrated in example (21):

- (20) In Canada a plan to slaughter the largest buffalo herd in the world has triggered off an unprecedented revolt ... (COBUILD: BBC).
- (21) Trackless surgery still has its limitations. It can be used only in ablative procedures, in which the surgeon's job is to destroy a body's rotten fabric. (COBUILD: *The Economist*)

The frequency distribution of the nouns in the corpus data suggests that there are particularly strong associations between the construction and the nouns of the types INTENDED RESULT and OBLIGATION. Thus the noun *aim* is by far the most frequently used noun in this pattern (with an attraction score of 12.10%) and also heads the list for reliance. This noun can be considered an anchor or “leit-noun” of the construction. The association is so strong that it is a fairly safe bet that informants would name this noun first or at least very early if they were presented with the pattern *the N is to ...* and asked to fill in the first nouns that come to their minds. Other members of the *aim* group, e.g. *ambition*, *intention*, *objective* and *purpose* are also among the top scorers, as are the prominent representatives of the OBLIGATION group *job*, *task* and *brief*.

Example (21) does not only illustrate the encoding of agents but also shows how the meanings of INTENDED RESULT and OBLIGATION are related to each other: the infinitive in this example expresses the goal (*to destroy a body's rotten fabric*) in a way analogous to *the surgeon's aim is to destroy ...*, but the noun *job* characterises this goal as something the surgeon is obliged to achieve. Apparently, the semantic complex of INTENDED RESULT and OBLIGATION is so strongly associated with the construction that it has the power to “rub off” on nouns that do not carry these meanings in isolation. Example (22), featuring the general mental noun *idea*, is a case in point:

- (22) It's now been over a month since Soviet president Mikhail Gorbachov cut off all oil and most natural gas supplies to Lithuania. The Kremlin's idea is to bring the country to its knees for having declared independence from the Soviet Union. (COBUILD: BBC)

Despite the unspecific meaning of *idea*, the construction undoubtedly carries the meaning of INTENDED RESULT; *idea* could be replaced by *aim* without any changes in meaning. The idea of “rubbing off” is justified insofar as the *aim* meaning of *idea* is indeed lexicalised, but mainly activated in this pattern (cf. OALD4, s.v. *idea* “AIM/INTENTION [C, U] the aim, intention, or purpose of doing something; *The idea is to teach children to save money*”). Significantly, the *aim*-meaning of the noun *idea* emerged fairly late in the diachronic development of this noun; the first attested occurrence of this meaning in all quotations in the OED, i.e. not just those given for *idea*, dates from the 1830s; substantial numbers of examples do not occur before the 1880s (cf. Schmid 1996, 99).

From the perspective of emergent meaning, the most interesting group is of course the MEANS group, since here the construction meaning is less easy to trace back to any of the construction components. The meaning of MEANS was not attributed to the *to*-infinitive by any of the sources referred to in Section 4.2 above. Looked at in isolation the nouns collected in this group make up a rather mixed bag including nouns actually expressing ‘manner’ or ‘means’, such as *way*, *countermeasure* and perhaps *alternative*. The most frequently used types of nouns giving rise to the MEANS meaning, however, are mental nouns such as *plan*, *policy*, *strategy*, *tactic* and the eventive nouns *approach*, *trick* and *solution*, as well as the illocutionary noun *answer* and the focusing noun phrases *the best/only thing*. In the construction, these diverse nouns all highlight the MEANS of an action and are at the same time influenced by the meanings of INTENDED RESULT and OBLIGATION dominating the construction meaning. Examples (23) to (26) illustrate how this blend of meanings emerges and reveal how the component of MEANS enters the picture:

- (23) To get sturdy seedlings, the trick is to keep the compost slightly on the dry side, never too wet, but make sure that it doesn't dry out fully. (COBUILD: Magazines)
- (24) We are winning against drugs but the only foolproof way is to bring in blood tests. (COBUILD: *The Guardian*)

- (25) Eric's approach is to shoot it in the shade on a bright sunny day, at f/11 and 1/30 sec on Fuji RDP 100. (COBUILD: Magazines)
- (26) The Hankeys were employed until the contracts were finished but were left owing the bank Pounds 88,000. As their overdraft had been secured on their home in South Ferriby, the only answer was to sell. (COBUILD: *The Times*)

Example (23) is fairly representative of uses of the noun *trick* in this construction because this sentence begins with another infinitive in adverbial function encoding the goal of the action, while the shell-content infinitive functioning as subject complement encodes an action representing how the goal can be achieved: in order "to get sturdy seedlings", you "keep the compost slightly on the dry side ...". Semantically similar, but couched in a different syntactic format, is example (24), where the goal is encoded as the first main clause "we are winning against drugs". In (25) and (26) the goals are not explicitly expressed but given earlier in the co-text (not included in example 25). Cases of this type seem to provide the missing link between the INTENDED RESULT meaning of the construction, on the one hand, and the MEANS meaning, on the other, because they show that the shell-content clauses can encode the means or manner of reaching a goal that is not explicitly encoded but understood or presupposed. In (26), for example, the shell content *to sell* encodes an action carried out in order to reach the goal of being able to pay back pounds 88,000 to the bank, and it is the noun *answer* that serves as a linguistic cue for this fairly complex chain of inferences. The semantic impact of the construction thus appears to be influenced by pragmatic aspects linking the event-components of INTENDED RESULT and MEANS (of reaching the intended result).

Example (27), a particularly interesting case already quoted as (2) above, contains the noun *problem* which is not included among the 20 top scorers but appears very soon further down the list (attraction score 1.31%).

- (27) Several years ago I was involved in a project aimed at measuring the reflectivity of metallised glass. The problem was to safeguard the many civil radar sites round Britain from encroachment by property development. Increasing numbers of buildings were being designed then with metallised glass. (COBUILD: *New Scientist*)
- (27') ... the (ultimate) aim was to safeguard ...
- (27'') ... the problem was how to safeguard ...

Here one goal is explicitly expressed in the preceding sentence (... *aimed at measuring the reflectivity of metallised glass*). The *problem*-sentence talks about another goal, *to safeguard the many civil radar sites round Britain from encroachment by property development*, which comes across as being superordinate to the first one: in order *to safeguard the radar sites* it was necessary *to measure the reflectivity of metallised glass* because, and this reason is given in the final sentence quoted here, *increasing numbers of buildings were being designed then with metallised glass*. Even though the infinitive *to safeguard the radar sites* thus encodes a goal, and despite the fact that *problem* does not have a MEANS component, the emergent meaning oscillates between the INTENDED RESULT interpretation

given in (27') and the MEANS one paraphrased in (27''). As before, the semantic complex encoded by the construction seems to emerge from an inferential chain. In fact, as far as textual argumentation is concerned, the *problem*-sentence serves as a justification of the information in the first sentence, yielding a paraphrase like “we wanted to measure X because we had to safeguard Z” – a paraphrase that unveils the OBLIGATION component (*we had to ...*) also hidden beneath the surface of this example.

6. Summary and conclusion

In this paper I have studied a set of lexico-grammatical patterns that qualify as constructions since they can carry meanings that are not entirely predictable from the semantics of the component parts, thus violating the principle of compositionality. The final section has focussed on one variant of these shell-content constructions, the pattern N + BE + *to*-infinitive, illustrating some mechanisms of emergent meaning. The overall picture suggested by this corpus study is that the frequent combinations of nouns and patterns are determined by symbiotic matches of the meanings of nouns and complementizers. Thus it is certainly not an arbitrary finding that, for example, *the aim is to ...* or *the fact that ...* are by far the most frequent representatives of the respective patterns, since the semantic match between the notion of *aim* and the semantic range of infinitives and the notion of *fact* and the meaning of *that*-clauses is particularly tight. The meanings that can be attributed to the four constructions are thus motivated by the converging meanings of complementizers and frequent head nouns.

Emergent meaning, not deducible from the component parts, apparently seems to come about as a result of the activation of the construction meaning, even in cases where the nouns contribute very little. The examples discussed suggest two avenues how this evolves: on the one hand, the construction apparently leaves its mark on unspecific nouns (such as *idea*) and has so to speak “rubbed off” its meaning in the course of repeated occurrences. For semantically more specific nouns like *answer* or *problem*, on the other hand, the effect can be explained in pragmatic terms as a result of inferential processes relying on the semantic complex associated with the construction.

It is very likely that this inference-based emergent meaning of the construction is indeed the result of a diachronic diversification of its semantic range. On the one hand, the earliest attestations of the pattern N + BE + *to* which I was able to find in the quotations in the OED and the diachronic part of the Helsinki Corpus all include nouns expressing the prototypical meaning of INTENDED RESULT, viz. *desire* (1571, OED), *aim* (1625, Helsinki), *hope* (1625, Helsinki) and *intention* (1616, OED). So this is not only the most frequent and prominent meaning of the construction, but also historically the original one. The OBLIGATION sense follows fairly soon; it is first attested in the OED in a quotation from the year 1647 featuring the noun *task*. Nouns encoding the MEANS meaning, on the other hand, occur only much later in the construction:

problem is attested in the pattern for the first time in 1833, *answer* quite surprisingly not before 1975. The diachronic diversification of the constructional meaning can be accounted for in terms of the fossilization of invited inferences along the lines suggested in section 5 above.⁵

Works Cited

Dictionaries

- A Valency Dictionary of English. A Corpus-based Analysis of the Complementation Patterns of English Verbs, Nouns and Adjectives* (2004). Thomas Herbst, David Heath, Ian F. Roe and Dieter Götz. Berlin/New York: Mouton de Gruyter. [VDE]
- Collins COBUILD English Language Dictionary* (1987). John M. Sinclair, ed. London et al.: Collins [COBUILD].
- Oxford Advanced Learner's Dictionary of Current English* (⁴1989). Albert S. Hornby, edited by Anthony P. Cowie. Oxford: Oxford University Press. [OALD4]
- Oxford English Dictionary* (²1989). Simpson, John and Edmund S.C. Weiner, eds., Oxford: Clarendon. [OED]

Other Sources

- Abraham, Werner, ed. (1978). *Valence, Semantic Case and Grammatical Relations*. Amsterdam: John Benjamins.
- Allerton, David J. (1982). *Valency and the English Verb*. London et al.: Academic Press.
- Biber, Douglas, Susan Conrad, Geoffrey Leech, Stig Johansson and Edward Finegan (1999). *Longman Grammar of Spoken and Written English*. London et al.: Longman.
- Conte, Maria-Elisabeth (1996). "Anaphoric Encapsulation." Walter de Mulder and Liliane Tasmowski, eds. *Coherence and Anaphora. Belgian Journal of Linguistics* 10, 1-10.
- Croft, William and David A. Cruse (2004). *Cognitive Linguistics*. Cambridge/New York: Cambridge University Press.
- Fillmore, Charles J. (1999). "Inversion and Constructional Inheritance." Gert Webelhuth, Jean-Pierre Koenig and Andreas Kathol, eds. *Lexical and Constructional Aspects of Linguistic Explanations*. Stanford, CA: Stanford University, 113-128.
- Fillmore, Charles J., Paul Kay and Mary Catherine O'Connor (1988). "Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*." *Language* 64, 501-538.

⁵ See Traugott and König (1991) and Traugott and Dasher (2002: 34-40) for an account of the invited-inferences theory of semantic change.

- Fischer Kerstin and Anatol Stefanowitsch (2006). "Konstruktionsgrammatik: Ein Überblick." Kerstin Fischer and Anatol Stefanowitsch, eds. *Konstruktionsgrammatik. Von der Anwendung zur Theorie*. Tübingen: Stauffenburg, 3-17.
- Frajzyngier, Zygmunt and Robert Jasperson (1991). "That-clauses and Other Complements." *Lingua* 83, 133-153.
- Francis, Gill (1986). *Anaphoric Noun*. Birmingham: University of Birmingham Printing Section.
- (1994). "Labelling Discourse: An Aspect of Nominal-Group Lexical Cohesion." Malcolm Coulthard, ed. *Advances in Written Text Analysis*. London: Routledge, 83-101.
- Frege, Gottlob (1923/1976). "Das Gedankengefüge." Günther Patzig, ed. *Gottlob Frege. Logische Untersuchungen*. Göttingen: Vandenhoeck und Ruprecht.
- Fried, Mirjiam and Jan-Ola Östman (2004). "Construction Grammar: A Thumbnail Sketch." Mirjiam Fried and Jan-Ola Östman, eds. *Construction Grammar in a Cross-linguistic Perspective*. Amsterdam/Philadelphia: John Benjamins, 11-86.
- Fried, Mirjiam and Hans C. Boas, eds. (2005). *Grammatical Constructions. Back to the Roots*. Amsterdam/Philadelphia: John Benjamins.
- Givón, Talmy (1990). *Syntax: A Functional-typological Introduction, Vol. 2*. Amsterdam/Philadelphia: John Benjamins.
- Goldberg, Adele E. (1995). *A Construction Grammar Approach to Argument Structure*. Chicago/London: Chicago University Press.
- (2006). *Constructions at Work. The Nature of Generalization in Language*. Oxford et al.: Oxford University Press.
- Gries, Stefan T. and Anatol Stefanowitsch (2004). "Extending Collostructional Analysis: A Corpus-based Perspective on 'Alternations'." *International Journal of Corpus Linguistics* 9, 97-129.
- Gries, Stefan T. (2006a). "Null-hypothesis Significance Testing of Word Frequencies: A Follow-up on Kilgarriff." *Corpus Linguistics and Linguistic Theory* 1, 277-94.
- (2006b). "Corpus-based Methods and Cognitive Semantics: The Many Meanings of *to run*." Stefan Th. Gries and Anatol Stefanowitsch, eds. *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis*, Berlin/New York: Mouton de Gruyter, 57-99.
- Herbst, Thomas (1988). "A Valency Model for Nouns in English". *Journal of Linguistics* 24, 265-301.
- Kay, Paul and Charles J. Fillmore (1999). "Grammatical Constructions and Linguistic Generalizations: The *What's X doing Y?* Construction." *Language* 75, 1-33.
- Kilgarriff, Adam (2005). "Language is Never, Ever, Ever, Random." *Corpus Linguistics and Linguistic Theory* 1-2, 263-276.

- Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar, Vol. I: Theoretical Prerequisites*. Stanford, CA: Stanford University Press.
- (1991). *Foundations of Cognitive Grammar, Vol. II: Descriptive Application*, Stanford, CA: Stanford University Press.
- Leech, Geoffrey N. (1983). *Principles of Pragmatics*. London/New York: Longman.
- Löbner, Sebastian (2002). *Understanding Semantics*. London: Arnold.
- Östman, Jan-Ola and Mirjam Fried (2005). “The Cognitive Grounding of Construction Grammar.” Jan-Ola Östman and Mirjam Fried, eds. *Construction Grammars. Cognitive Grounding and Theoretical Extensions*. Amsterdam/Philadelphia: John Benjamins, 1-13.
- Palmer, Frank R. (1990). *Modality and the English Modals*. London/New York: Longman.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik (1985). *A Comprehensive Grammar of the English Language*. London et al.: Longman.
- Schmid, Hans-Jörg (1996). “The Historical Development and Present-day Use of the Noun *Idea* as Documented in the OED and other Corpora.” *Poetica* 47, 87-128.
- (2000). *English Abstract Nouns as Conceptual Shells. From Corpus to Cognition*. Berlin/New York: Mouton de Gruyter.
- Sinclair, John M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stefanowitsch, Anatol and Stefan Th. Gries (2003), “Collostructions: Investigating the Interaction of Words and Constructions.” *International Journal of Corpus Linguistics* 8-2, 209-243.
- Tesnière, Lucien (1959). *Éléments de Syntaxe Structurale*. Paris: Librairie C. Klincksieck.
- Traugott, Elizabeth C. and Ekkehard König (1991). “The Semantics-pragmatics of Grammaticalization Revisited.” Elizabeth C. Traugott and Bernd Heine, eds. *Approaches to Grammaticalization. Vol. I: Focus on Theoretical and Methodological Issues*. Amsterdam/Philadelphia: John Benjamins, 189-218.
- Traugott, Elizabeth C. and Richard B. Dasher (2002). *Regularity in Semantic Change*. Cambridge et al.: Cambridge University Press.
- Tummers, Jose, Kris Heylen and Dirk Geeraerts (2005). “Usage-based Approaches in Cognitive Linguistics: A State-of-the-art Report.” *Corpus Linguistics and Linguistic Theory* 1, 225-261.
- Ungerer, Friedrich and Hans-Jörg Schmid (2006). *An Introduction to Cognitive Linguistics*. London et al.: Pearson Longman.
- Wierzbicka, Anna (1988). *The Semantics of Grammar*. Amsterdam/Philadelphia: John Benjamins.

Appendix

Noun	Freq. in pattern	Attraction	Noun	Freq. in pattern	Freq. in corpus	Reliance
fact	26,106	18.45%	realisation	820	1,185	69.20%
evidence	5,007	3.54%	proviso	111	250	44.40%
idea	4,812	3.40%	assumption	1,391	3,151	44.14%
doubt	4,010	2.83%	assertion	596	1,492	39.95%
belief	3,696	2.61%	belief	3,696	9,344	39.55%
view	3,532	2.50%	insistence	796	2,069	38.47%
hope	2,727	1.93%	fact	26,106	68,472	38.13%
news	2,572	1.82%	premise	274	765	35.82%
feeling	2,511	1.77%	misapprehension	44	123	35.77%
impression	2,279	1.61%	suggestion	2,033	5,854	34.73%
possibility	2,232	1.58%	dictum	84	249	33.73%
claim	2,194	1.55%	stipulation	48	145	33.10%
suggestion	2,033	1.44%	misconception	91	284	32.04%
speculation	1,922	1.36%	truism	47	150	31.33%
knowledge	1,794	1.27%	reminder	812	2,688	30.21%
sign	1,738	1.23%	notion	1,655	5,713	28.97%
notion	1,655	1.17%	coincidence	627	2,196	28.55%
point	1,511	1.07%	speculation	1,922	6,778	28.36%
warning	1,460	1.03%	supposition	46	164	28.05%
fear	1,432	1.01%	impression	2,279	8,206	27.77%

Table 5. Pattern: N-*that*; total number of types: 350; total number of tokens: 141,476

Noun	Freq. in pattern	Attraction	Noun	Freq. in pattern	Freq. in corpus	Reliance
time	19,496	8.54%	inability	1,843	2,400	76.79%
way	15,194	6.66%	temerity	118	160	73.75%
attempt	13,543	5.94%	disinclination	45	62	72.58%
chance	11,722	5.14%	willingness	1,804	2,493	72.36%
decision	10,089	4.42%	unwillingness	337	470	71.70%
ability	9,830	4.31%	attempt	13,543	20,728	65.34%
right	8,522	3.74%	refusal	2,406	3,892	61.82%
need	8,150	3.57%	ability	9,830	16,957	57.97%
opportunity	7,799	3.42%	urge	918	1,626	56.46%
effort	5,221	2.29%	reluctance	1,158	2,131	54.34%
power	4,194	1.84%	eagerness	285	535	53.27%
failure	4,179	1.83%	readiness	676	1,368	49.42%
desire	4,159	1.82%	opportunity	7,799	17,805	43.80%
plan	3,866	1.69%	inclination	339	801	42.32%
place	3,791	1.66%	propensity	189	449	42.09%
reason	3,499	1.53%	desire	4,159	9,973	41.70%
bid	3,343	1.47%	tendency	1,763	4,297	41.03%
campaign	3,098	1.36%	temptation	859	2,193	39.17%
pressure	2,649	1.16%	obligation	765	2,041	37.48%
position	2,476	1.09%	incentive	969	2,690	36.02%

Table 6. Pattern: N-*to*; total number of types: 200; total number of tokens: 228,165

Noun	Freq. in pattern	Attraction	Noun	Freq. in pattern	Freq. in corpus	Reliance
problem	2,672	8.62%	upshot	105	313	33.55%
thing	1,532	4.94%	snag	250	784	31.89%
truth	1,235	3.98%	drawback	140	735	19.05%
fact	1,218	3.93%	implication	274	1,514	18.10%
trouble	1,034	3.34%	guess	266	1,620	16.42%
point	1,020	3.29%	irony	395	3,085	12.80%
result	977	3.15%	downside	51	512	9.96%
view	933	3.01%	inference	37	375	9.87%
reason	897	2.89%	corollary	19	198	9.60%
idea	790	2.55%	hunch	43	451	9.53%
news	749	2.42%	gripe	17	186	9.14%
difference	642	2.07%	stipulation	12	145	8.28%
answer	633	2.04%	worry	253	3,119	8.11%
theory	561	1.81%	assumption	229	3,151	7.27%
reality	509	1.64%	truth	1,235	17,421	7.09%
hope	482	1.56%	complication	34	484	7.02%
fear	437	1.41%	likelihood	125	1,857	6.73%
argument	430	1.39%	disadvantage	101	1,556	6.49%
danger	395	1.27%	regret	99	1,754	5.64%
irony	395	1.27%	paradox	63	1,149	5.48%

Table 7: Pattern: N-be-that; total number of types: 366; total number of tokens: 30,992

Noun	Freq. in pattern	Attraction	Noun	Freq. in pattern	Freq. in corpus	Reliance
aim	2,646	12.10%	aim	2,646	9,324	28.38%
idea	1,141	5.22%	brief	118	835	14.13%
job	1,041	4.76%	objective	370	2,883	12.83%
task	946	4.32%	countermeasure	1	10	10.00%
thing	758	3.46%	ambition	373	4,678	7.97%
plan	629	2.88%	intention	494	6,484	7.62%
purpose	586	2.68%	task	946	13,439	7.04%
solution	551	2.52%	priority	425	6,055	7.02%
step	515	2.35%	trick	368	6,080	6.05%
intention	494	2.26%	inclination	46	801	5.74%
answer	431	1.97%	tactic	71	1,300	5.46%
priority	425	1.94%	instinct	150	2,853	5.26%
goal	399	1.82%	purpose	586	11,524	5.09%
ambition	373	1.71%	alternative	334	7,046	4.74%
policy	370	1.69%	solution	551	12,161	4.53%
objective	370	1.69%	temptation	82	2,193	3.74%
trick	368	1.68%	option	328	9,913	3.31%
alternative	334	1.53%	bet	95	3,037	3.13%
option	328	1.50%	object	187	6,027	3.10%
way	316	1.44%	function	174	5,651	3.08%

Table 8. Pattern: N-be-to; total number of types: 162; total number of tokens: 21,876