Chapter 10. Construction grammar and lexicography

William Croft and Logan Sutton, University of New Mexico

10.1. Introduction

Construction grammar is an approach to grammatical analysis and representation that has attracted considerable attention since the seminal works were published around twenty-five years ago (Fillmore, Kay & O’Connor, 1988; Lakoff, 1987). Since that time, a number of variants of construction grammar have been developed by different scholars. Nevertheless, there is a common core of principles that characterize virtually all types of construction grammars (Croft & Cruse, 2004, chapters 9-10). Radical Construction Grammar (Croft, 2001) was developed to situate construction grammar in a crosslinguistic perspective. Its primary hypothesis on the nature of grammatical categories has significant implications for lexicography, but also holds out the prospect of making lexicographic practice in different languages more comparable.

10.2. Description

10.2.1. Construction grammar

Construction grammar contrasts with the componential approach to grammatical analysis, characteristic of structuralist and generative theories of grammar. In the componential model, different types of linguistic information—phonological, syntactic, semantic, pragmatic—are represented in self-contained components constrained by rules that apply only to that type of information. The major exception to this componential division is the lexicon, where the sound structure, syntactic behavior, and meaning and use of a word are linked together. The lexicon is of course the “component” of greatest interest to lexicographers. Even in the componential model, however, rules (interface rules, linking rules, realization rules) are required to link larger linguistic structures than individual words.

Fillmore, Kay & O’Connor (1988) use idioms to demonstrate that there is a continuum between lexical items such as building—single units with phonetic substance, syntactic behavior, and meaning—and grammatical rules or constructions such as the transitive construction [SUBJECT, VERB, OBJECT]—complex structures with (multiple) units that are fully schematic (containing no specific words), with their associated semantic linking rules. Idioms are complex structures but some of the units that form the structure are substantive (specific words), as in blow X’s nose, I can’t walk, let alone dance, or Nth cousin M times removed. Moreover, the meanings and pragmatic functions of these idioms are not predictable from more general rules of semantic interpretation, and the syntactic categories of the schematic roles in the idioms are less general than the usual parts of speech (X in blow X’s nose must be a possessive pronoun, Nth in Nth cousin M times removed must be an ordinal numeral, etc.).

Fillmore et al. argue that the structure of idioms of the type given in the preceding paragraph are not predictable from more general grammatical rules, but are themselves rule-governed, productive structures that generate special types of phrases, clauses and
sentences that have semantic interpretations and pragmatic conditions of use. In other words, they are less specific, and often contain concrete substantive units (specific words), but they are made up of their own specialized “syntactic rules” and “semantic interpretation rules”. Fillmore et al. then draw the conclusion that all grammatical units, from single words to fully schematic constructions, are of the same type: a pairing of form (both phonological and syntactic) with meaning (including pragmatic conditions of use). In between the two ends of this continuum is an entire range of grammatical types, namely different kinds of idioms, some of which are fully substantive (the more the merrier, the bigger they come the harder they fall) and others of which are partly substantive, partly schematic (the Xer, the Yer, and the idioms described in the previous paragraph). In addition, morphologically complex words—compounds, roots with derivational morphemes, and stems with inflectional morphemes—are also constructions: they are complex but the parts are morphologically bound. In other words, all grammatical structures are constructions.

10.2.2. The contribution of generic construction grammar principles to lexicography

The chief contribution of generic construction grammar to lexicography is to make the representation of all conventional grammatical units or structures more like the traditional representation of lexical units: a pairing of form (phonological and syntactic) and meaning. In fact, some construction grammarians refer to the full inventory of constructions broadly conceived as the constructicon, including all lexical items (monomorphemic and polymorphemic), and all grammatical constructions, from the fully or partially substantive (“idioms”) to the fully schematic (“syntactic structures”).

In theory, then, at least, from a construction grammarian’s perspective, lexicographers could become constructicographers. Dictionaries would then swallow grammars and become the compendium of all linguistic knowledge. We suspect that most lexicographers would prefer to maintain the culturally and commercially defined distinction between a dictionary and a grammar, at least for the time being. The perspective of generic construction grammar does not necessarily change the practical division between the dictionary and the grammar. However, it allows us to formulate the division in a more theoretically motivated manner that would be consistently applicable to dictionary-making across the full range of typological diversity of languages.

The obvious way to distinguish between a dictionary and a grammar in a construction grammar sense would be to exclude fully schematic constructions, such as the transitive construction, from the dictionary (but see below for the occurrence of lexical items within such constructions). On the other hand, phrasal idioms and collocations would fall on the dictionary side of the constructional continuum, so the more the merrier and the bigger they come the harder they fall would be included in a dictionary. The inclusion of idioms in a dictionary is already a salient issue particularly in English lexicography, and is already well covered in this book (chapters 7 [Francis], 13 [Krishnamurthy] and 20 [Wu]).

A question that remains from a constructional perspective, however, is, how “substantive” must an idiom be in order for it to be included in a dictionary? For example, should the Xer, the Yer be included? The paired morphemes the in this construction are etymologically distinct from the definite article the, yet a construction
this schematic might be considered too close to the syntax to be included in a dictionary. Likewise, many so-called function words (prepositions, articles, auxiliaries, discourse markers etc.) in relatively isolating languages like English are generally included in a dictionary because they are independent words (at least orthographically). Yet they are often a small part of many different constructions. For instance, the preposition with is part of several different argument structure constructions: instrument (She hit it with a hammer), associative (I’d like a hot chocolate with whipped cream), and manner (She sang the tune with gusto), not to mention the use of with to introduce a reduced subordinate clause (With the transport workers on strike, the subway isn’t running this week).

We believe that the dictionary entries for such words should ideally provide some information about the construction(s) they occur in. In many dictionaries, including dictionaries written by documentary linguists, this morphosyntactic information is given implicitly in the example sentence(s) provided for each sense of a function word. However, in most cases no word-by-word gloss of the example is given (a rare exception is the dictionary of Zaiwa [Lustig, 2011], nor is a schematic or prose description provided of the general construction such as the Xer the Yer (with some elaboration of what X and Y can be). While lexicographers may not be expected to provide the sort of detail of a construction’s structure and use that is normally found in a reference grammar of a language, at least a general schema for the construction and a general characterization of its meaning would be desirable in a dictionary.

In construction grammar, morphologically complex words—root plus affixes—are constructions as well. In many languages, the equivalents of the morphologically independent function words of English are affixes. Hence the practical issue described in the preceding paragraph is the same but different: should a dictionary include separate entries for affixes? Some dictionaries, such as Montler’s (2012) Klallam dictionary, do, while many others do not.

Another issue is whether to organize entries by roots (monomorphemic units), by stems (mono- or polymorphemic lexical bases to which inflectional morphology is added), or by full words (however these are defined). From a construction grammar perspective, this is the morphologically bound version of the issue of how and where to include idioms in a dictionary. To some degree, a language with extensive derivation and/or compounding is a language with many idioms that are combined into a single morphological word. The word vs. stem vs. root organization issue is particularly a problem in a language with extensive prefixes, given the convention of an alphabetical organization.

For example, in Russian there are root-based dictionaries (e.g. Wolkinsky & Poltoratzky, 1961) as well as the more common word-based dictionaries, since Russian employs prefixes that combine the functions of inflectional aspect as well as derivation (word formation). Moving to morphologically richer small indigenous languages, we observe that practice varies. Dictionaries for most morphologically complex languages appear to be organized according to the (mono- or polymorphemic) lexical stem. Compounds and stems with derivational morphology whose translations are not predictable on the basis of the constituent morphemes may be entered either under the root determined to be the lexical head or in alphabetical order, depending on the particulars of the morphology and semantics involved. A morphologically complex entry
may also be included in an otherwise root-based dictionary when the translation equivalent would most normally be expressed by a monomorphemic root.

It is rare to find dictionaries with a truly word-based organization of entries when the language primarily uses prefixation and the lexical stem cannot occur as a word form in any construction without bound morphology. (This is particularly common among indigenous languages of North America.) A representative morphologically complex word-form can sometimes be used to serve as the head of a dictionary entry, as with the infinitive construction in many modern Indo-European languages. However, the minimal morphological word for at least some word classes (namely verbs) in many languages may require more than one bound (inflectional) morpheme. Dictionary makers may therefore shy away from choosing which arbitrary full word-form should serve as the representative from which to predict other inflected forms. The representative word-form organization is also problematic for alphabetization of entries in dominantly prefixing languages: entire word-classes would be grouped together within the dictionary by virtue of the chosen word forms all beginning with the same inflectional morphemes. This may then make it more difficult for the dictionary user to parse through the entries.

Most lexicographers for languages with extensive prefixing employ a stem-based organization, such as Onondaga (Woodbury, 2003), Skiri Pawnee (Parks & Pratt, 2008) and Lakhota (Lakota Language Consortium, 2008). For verbs in particular in these languages, the dictionary entry does not constitute a legitimate word comprehensible to a native speaker. It is only by the addition of inflectional affixes that the stems become pragmatically acceptable words. Details on the morphology to be appended to produce legitimate words must be provided by a separate grammatical description of the language (although the dictionary entry should specify particular inflectional construction information that is not predictable from the grammar).

Such a stem-based organization is not the only option for such prefixation-heavy languages. Young & Morgan produced separate word-based and root-based dictionaries of Navajo (Young & Morgan, 1987, 1992), and the aforementioned Klallam dictionary includes a word-based, stem-based, and root-based dictionary plus the affix list in a single volume. The Klallam dictionary furthermore provides a morphological analysis of the stems, also desirable for morphologically complex languages. Stems may require inflectional morphemes to constitute full words, but are often themselves lexicalized with a fair amount of morphological complexity.

Yet another solution in a physical format is to include extensive cross-references, as in Liljeblad et al.’s (2011) Northern Paiute-Bannock dictionary. This technique may be useful for organizing entries no matter whether the dictionary is arranged by root, stem, or fully formed word. All of these solutions may be rendered moot by digital hypertext dictionaries (see the entries in Section 3 of this handbook).

These issues, like the issues around organizing information about idioms in dictionaries, are ones that lexicographers—at least, lexicographers working in morphologically complex, prefix-heavy languages—are already aware of. Our point here is only to note that from a construction grammar perspective, root- vs. word-based organization of dictionaries poses the same sort of practical organizational problems as the inclusion of idiomatic phrases in less morphologically complex, more isolating languages. Indeed, it should be remembered that even morphologically complex languages will also have syntactic constructions that should be represented in a
As construction grammarians, we would suggest that the lexicographic solutions to these two sets of problems, whether in physical or digital format, would—and should—be similar.

10.2.3. Radical Construction Grammar

The most significant contribution that construction grammar can make to lexicography is the information about the syntactic behavior of words that is or could be included in a dictionary. Construction grammar provides a more sophisticated perspective on syntactic behavior, but it also shows that syntactic behavior is more complex than is generally assumed.

Syntactic structures in generative grammar are defined as phrase structure trees that terminate in a small number of lexical categories (noun, verb, adjective, etc.). Most theoretical attention in generative grammar has been focused on generalizing syntactic categories (X-bar theory). Radical Construction Grammar, on the other hand, proposes a more radical reconsideration of syntactic categories, including word classes, from both a crosslinguistic and a language-internal perspective.

The chief problem from a crosslinguistic perspective is that one cannot equate word classes from one language to the next. For example, it is frequently claimed that a particular language, for example Mandarin Chinese, lacks the category “adjective”, and that in Mandarin “adjectives are verbs”. The typical evidence that is given in this case is that predicating the sort of concept that is expressed as an adjective in English, e.g. pāng ‘fat’, uses the same construction as predicating the sort of concept that is expressed as a verb in English, e.g. kū ‘cry’ (Li & Thompson, 1981, pages 143 and 156).

The facts are actually more complicated than this, but leaving that aside, we can see that the word class(es) defined by the relevant construction(s) in Mandarin is not the same as the word class(es) defined by the relevant construction(s) in English. Having observed this fact, on what basis can we say that in Mandarin “adjectives are verbs”? The word class in Mandarin is identical neither to the English adjective class nor to the English verb class. Even if we decide to name it one or the other, there is no a priori basis to name it “verb” instead of “adjective”.

The problem is that word classes are defined by the occurrence of the words in a construction (or set of constructions) in a language; but the constructions in one language are different from the constructions in another language. Indeed, it is a principle going back to structuralist linguistic theory that a syntactic category such as a word class in a language can only be defined in terms of occurrence in the constructions in that language (that is, its distribution). But since constructions are language-specific, word classes cannot be compared across languages.

In fact, the problem is more serious than this. Linguists disagree about word classes in a language because they use different constructions from the same language to define the word classes. This is true just as much for a well-described language like English (are there two word classes, prepositions and particles, or just one?) as it is for a less well documented indigenous language like Nootka (are there distinct noun, verb and adjective word classes, or just one class?). The reason is the same in both cases: it depends on which construction is used to define the word class or classes.
Many typologists have come to accept that languages have their own word classes or syntactic categories; that is, syntactic categories are language-specific. However, as the immediately preceding paragraph indicates, even the constructions in a single language do not define the same word classes. The distribution of words in constructions varies from one construction to the next. That is, word classes are not just language-specific but construction-specific. This is the primary principle of Radical Construction Grammar (Croft, 2001).

This would seem to be a completely anti-universalist approach to syntactic theory, but in fact it is not. First, the argumentation in this section reveals something which is rarely made explicit in discussions of any syntactic theory; word classes are defined by their distribution in syntactic constructions. This fact is obscured by the use of a variety of terms instead of “construction”—criteria, tests, arguments, etc. for a word class—and a variety of terms instead of a word’s “distribution in constructions” to define its word class—a word’s behavior, properties, features, traits, etc.

The fact that word classes are defined in terms of constructions allows us to bring in the fundamental property of constructions given in section 10.2.1, namely that they are pairings of form and meaning, applying to everything from individual words to complex, fully schematic structures. In other words, we can use meaning as well as form to reconstruct a crosslinguistically valid approach that captures the intuitions behind syntactic categories such as “noun”, “verb” and “adjective”.

For example, the intuition behind the claim that “Mandarin adjectives are verbs” has to do with both the meaning of the word classes and the function of the predication construction. The word classes of English and Mandarin are comparable to the extent that they include semantic translation equivalents, in this case property concept words (“adjectives”) and action concept words (“verbs”). The construction used in this example is the construction in Chinese for predicating property concepts and also action concepts. The predication function is also prototypically associated with “verb” status in grammatical theory and hence the word class label chosen is “verb” rather than “adjective”.

The constructions typically used for defining the major lexical categories or parts of speech are the constructions used for predication, modification and reference, described in Croft (2001) as the propositional act functions. Other constructions are also used to define parts of speech, in particular inflectional constructions. However, word classes based on inflectional constructions should be treated as different classes from those defined by propositional act constructions.

There are two reasons for the crosslinguistic problems in identifying and labeling word classes. First, most languages have more than one construction for each of the above three functions. For example, English uses the copula be for predicating property concepts (He is fat) but no copula and inflection for person/number and tense for predicating action concepts (He cried). Second, and most importantly, the semantic classes of words that occur in the propositional act constructions differ from language to language and also from one propositional act construction to another in the same language. This diversity is a basic fact of grammar. In fact, typologists have uncovered a number of universals about the mapping between semantic classes of words and their distribution in constructions performing particular functions (see Comrie, 1989; Croft, 2003, for textbook surveys of language universals). But grammarians and lexicographers
have to accept that there is a high degree of variation in how words are distributed across the constructions of a language (and how their semantic equivalents in other languages are distributed across the functionally equivalent constructions in those languages).

10.2.4. Radical Construction Grammar and lexicography: the inclusion of major distributional information

Traditional lexicographic practice has followed traditional syntactic approaches to syntactic categories, in which major parts of speech provide all of the information available about the occurrence of words in syntactic constructions. The primary grammatical (syntactic distributional) information that is found in lexical entries in both “major language” dictionaries and “indigenous language” dictionaries is word class, and sometimes word subclass. In a survey of 42 bilingual dictionaries, 39 provided some kind of lexical category information. These are almost never elaborated beyond the traditional Western categories of “noun”, “adjective”, “verb”, etc. aside from specification of the valence of verbs. The latter itself tends not to be any more specific than “transitive” or “intransitive” (see discussion below).

However, a word class or even subclass label does not explicitly tell us what constructions the word can occur in. The chief contribution of construction grammar (and Radical Construction Grammar) to lexicography is to replace word class information in lexical entries with constructional distributional information. That is, the abbreviations used for syntactic information will refer to specific constructions, not word classes.

In fact, each word has its own distributional pattern of occurrence across all the constructions of a language (Gross, 1979; Croft, 2001, inter alia). This is, of course, a lexicographic nightmare. Even for languages with large electronic corpora available, it would be a daunting if not impossible task to index every construction that a word occurs in, for every word in the dictionary. However, there is also no doubt that the limited and inexplicit morphosyntactic information found in a traditional lexical entry means that valuable linguistic information falls in the cracks between a grammar and a dictionary.

The fact that constructions have functions (meanings) allows us to propose an ordered list of the functionally most important constructions in a language, for which it would be useful to have distributional information recorded in a lexical entry. This will require identifying the range of constructions in the language whose distribution is described in the dictionary. The introduction to the dictionary ideally would have a small sketch grammar of the constructions which the dictionary entries refer to (and appropriate abbreviations).

(i) The part of speech categories (n., v., adj., etc.) found in traditional dictionaries generally reflect at least one of two distinct types of constructions: morphological inflectional constructions and propositional act constructions. Morphological inflectional information has traditionally been included in dictionaries and will not be discussed further here (but see below).

We suggest that part of speech categorization should specifically refer to occurrence in the predication, modification (attributive) and referring constructions of the language. Attributive constructions are often syntactically well differentiated: numeral and possessive classifiers, alienable vs. inalienable (attributive) possession, genitive-like vs.
adjective-like numerals, different subclasses of property words (“adjectives”) according to the attributive constructions of the language, etc. Predicating constructions are typically multiple as well: “verbal” vs. one or more types of “nonverbal” predication (see Stassen, 1997). Reference, when it is not prototypical, is often accomplished by derivational morphology (action nominals such as liberation, “deadjectival” nominals such as freedom), and so is already usually covered by lexicographic practice. However, other languages, particularly North American languages, use a relative clause-like construction for many nonprototypical referring expressions (e.g. Arikara nuunawaawihú “photographer”, lit. “the one who makes images”, nuunawi’oóxit “the one who is quiet, the quiet one”, napahaátu “the red one, the one that is red”). Such relative-like constructions may stand beside other “deverbal” and “deadjectival” derivational constructions within the language, but the choice of which relative constructions to include in the dictionary will be somewhat arbitrary

(ii) Syntactic subclass categories (vi, vt, etc.) found in traditional dictionaries generally reflect major constructions in the clause, for “verb” or predicate subclasses. (Major constructions in referring phrases are already described by modifying propositional act constructions; see (i).) The most important constructions at the clause level are argument structure constructions (Goldberg, 1995), which specify the number, obligatoriness, form and semantic type of argument phrases (“subject”, “object”, and “oblique” phrases and their associated adpositions or case inflections) that combine with particular predicates. Examples of this information for lexical items are the construction information in the FrameNet project (https://framenet.icsi.berkeley.edu/fndrupal/; see also chapter 9 [Atkins]); the information in Levin (1993) and its computational implementation in VerbNet (http://verbs.colorado.edu/~mpalmer/projects/verbnet.html); and in a crosslinguistic database of more manageable dimensions, the Leipzig Valency Classes project (http://www.eva.mpg.de/lingua/valency/index.php).

Argument structure information is already sometimes indicated in dictionaries of languages where verbal inflections (indexation/agreement) are used to encode argument structure constructions. Examples include Onondaga (active, inactive, transitive; Woodbury, 2003) and Creek (various combinations of set I, II, III indexation markers; Martin & Mauldin, 2000).

English is distinctive (in fact, typologically unusual) in that the same verb can occur in multiple argument structure constructions without any morphological alteration (e.g. I sprayed water on the lawn, I sprayed the lawn with water, Water sprayed onto the lawn). In most languages, such alternations are accompanied by morphological derivation, usually described as voice or diathesis, going under names such as causative/anticausative, applicative, antipassive, passive, inverse, middle. This information can be combined with information about the argument structure constructions associated with these morphological derivations. Current lexicographic practice varies to the degree that such semi-derivational possibilities are included.

Another example of an argument structure-like constructional distribution that is significant is the sort of clausal complement types (or nonfinite or deranked clause-like complement types) that a predicate occurs with, usually described as finite complements, infinitives, gerunds, nominalizations, participles, secondary predicates, etc. English is a good example of a language with a great variety of complement types, but other
languages have a variety of clausal complement constructions associated with different classes of predicates.

(iii) The constructions described above will make reference to relational/contextual inflections (Havel, 2002) as part of the construction, for languages that have such morphological inflections. Relational/contextual inflections encode relations between syntactic elements, and include: indexation/agreement (heads or modifiers), case marking, linkers, and special forms (cf. Croft, 2003, chapter 2). Other lexicographically relevant information has to do with inherent inflections, such as gender, number and definiteness for referring expressions, gradability for modifiers, and tense, aspect and modality for predications. Of course many inherent semantic categories are expressed through independent words in some languages that will have their own lexical entries (such as the English articles the and a(n)); and others are sometimes expressed periphrastically (such as the English progressive aspect construction be V-ing). From a constructional perspective, what matters is distribution of referring expressions and predicates in the constructions that express these meanings, whether the constructions are inflectional or periphrastic.

Similar semantic categories tend to be expressed inflectionally across languages, which motivates the choice of entering such constructional information in dictionaries. The semantic categories that are most useful to enter constructional information on in dictionaries are the ones that Havel (1985) calls the most relevant categories: those that most strongly semantically interact with the meaning of the word or root that they combine with. These are the categories that are going to be the most lexically idiosyncratic—that is, not apply to all predicates, modifiers or referring expressions—and therefore are the most informative in a lexical entry.

The most relevant (in both senses of the word) semantic constructions for referring expressions or object words (“nouns”) are gender/class and number (countability). This is actually a relational category, since it is manifested in agreement (Corbett, 1991). However, since the noun is the controller of agreement, gender/class information is usually given in the lexical entry for nouns. It is already current lexicographic practice to include gender/class information in lexical entries, but the agreement constructions that use gender/class should be included in the inventory of constructions in the dictionary.

Number inflection reflects constructions indicating the countability, or lack thereof, of the word: mass nouns do not occur in plural constructions, pluralia tanta do not occur in singular constructions; and languages with dual constructions have dual-only words as well, and other distributional patterns occur. Some dictionaries that include subclass information may include countability construction distributions. The same tends not to be true in dictionaries of major Western languages. More comprehensive bilingual (or monolingual) dictionaries of, e.g., German, Spanish, Russian, French, and English may include specific countability constructions among the particular idioms under a given lexical entry. However, no generalization regarding distribution among countability constructions is usually given with class or subclass information in an entry: the dictionary user is left to carefully examine the example constructions to determine this him- or herself. English is quite flexible on placing a word in either a countable or noncountable (“mass”) construction, although other languages require morphological derivation, as with voice and diathesis discussed above. Semantic shifts may also occur
when a word is used in different countability constructions that are not completely regular (the semantic relationship between hair [on the head] and a hair [strand] is not the same as that between chicken [meat] and a chicken [bird]). Ideally these facts should be recorded in lexical entries.

The most relevant semantic category for modifiers/property words (“adjectives” or “modifers”) is gradability, manifested in occurrence in degree modification constructions (intensifiers, downtoners, etc.). The most relevant semantic category for predicates/action words (“verbs”) is aspect. Aspect is a semantically complex category (see Croft, 2012, chapters 2-3 for a detailed analysis), and many constructions may be used for lexicographic information. As with countability, English is extremely flexible on placing a word in different types of aspectual constructions (that construe the event as being stative or dynamic, telic or atelic, punctual or durative, semelfactive or iterative, etc.), while other languages require morphological derivation. Also like countability constructions, semantic shifts may occur when a word is used in different aspectual constructions that are not completely regular (for example, be reading describes the middle of a action, whereas be flashing describes repetition of the action).

10.2.5. Conclusion

There is no doubt that the suggestions for providing distributional information for certain constructions made in this chapter will require considerably more effort and empirical analysis (via corpora or native language consultant elicitation) than traditional lexicographic practice. The alternative is for grammars to make exhaustive lists of the words that occur in those constructions. This is not common practice among writers of reference grammars (unless the word class is very small). As a result, valuable information about constructions and the distribution of words in constructions falls between the traditional practices of lexicographers and grammarians. While we do not rule out including this constructional distributional information in grammars, we believe that a more construction-oriented notation of lexical distributions in dictionaries can substantially increase their lexicographic value.

References


