The AAAI 2017 Spring Symposium on
Computational Construction Grammar and Natural Language Understanding
Technical Report SS-17-02

A Computational Construction Grammar for English

Remi van Trijp
Sony Computer Science Laboratory Paris
6 Rue Amyot
75005 Paris, France
remi@csl.sony.fr

Abstract

Human language users are capable of proficiently learning new constructions and using a language for everyday communication even if they have only acquired a basic linguistic inventory. This paper argues that such robustness can best be achieved through a *constructional processing model* in which grammatical structures may emerge spontaneously as a side effect of how constructions are combined with each other. This claim is substantiated by a fully operational precision model for Basic English in Fluid Construction Grammar, which is available for online testing. The precision model is the first ever to incorporate key properties from construction grammar in a large-scale setting, such as argument structure constructions and the surface generalization hypothesis, and is therefore a milestone achievement in the field of construction grammar.

One of the most fascinating features about human language processing is that human learners are already capable of successfully engaging in challenging communicative tasks even if they have not yet fully mastered a language. For example, due to the Zipfian frequency distribution of words, language learners can already get very far with a small vocabulary of high-frequency words (Nation 2001). Understanding how humans are capable of this remarkable feat is therefore important for advancing our knowledge about cognition and for developing more intelligent language technologies.

This paper argues that a key part of the solution involves *constructional language processing*, because such models are able to handle structures that are not always explicitly specified in the constraints of individual constructions. Instead, a construction grammar is able to cover structures that emerge spontaneously as a side-effect of different ways of combining the same constructions. This claim is substantiBENGated through a fully operational and bidirectional grammar of Basic English in Fluid Construction Grammar (Steels 2011, henceforth abbreviated as FCG), which can be tested online at `http://www.fcg-net.org/fcg-interactive/`. The precision model is the first ever to incorporate key properties from construction grammar in a large-scale setting, such as argument structure constructions and the surface generalization hypothesis (Goldberg 2006).

Figure 1: This Figure shows the interface of FCG-interactive. Interested readers can select the English grammar and then comprehend and formulate any English sentence if they want to test how much the grammar for Basic English is able to handle. The interface is accessible at `http://www.fcg-net.org/fcg-interactive/`.

A Grammar for Basic English

The Basic English Grammar (BENG)[1] is part of a larger research program to (a) understand the computational properties of construction grammar (Fillmore 1988; Goldberg 2006), and (b) to scale constructional processing models to broad-coverage precision grammars. Current efforts are focused on developing basic linguistic inventories that are nevertheless capable of achieving a precise (albeit sometimes partial) grammatical and semantic analysis of any English sentence, and to try and formulate those sentences again based on a meaning representation. Figure 1 shows the web interface that readers can interact with for trying the BENG.

The design of the grammar assumes that, in comprehension, nouns can be robustly distinguished from verbs in the input sentence. In order to achieve this, an input sentence is first processed by a statistical model called Parsey McParseface, which is a state-of-the-art syntactic dependency parser developed with Google SyntaxNet (Andor et al. 2016). Most of this parser's analysis is discarded, however, as the BENG only keeps the tokenized strings of the input sentence and the part-of-speech (POS) analysis

[1] The BENG is inspired by the system of Basic English (Ogden 1968), a language project and policy supported by a.o. Winston Churchill and Franklin D. Roosevelt for promoting English as a second language througout the world.

that uses a limited number of POS-tags from the universal dependency project (Petrov, Das, and McDonald 2012; de Marneffe et al. 2014). Example (1) shows an input sentence, a detailed POS analysis for each word on the first gloss (which is ignored by the BENG), and a universal POS-tag analysis on the second gloss (of which a small subset is used by the BENG).

(1)

Do	*you*	*want*	*to*	*try*
VPB	PRP	VB	TO	VB
VERB	PRON	VERB	PRT	VERB

our	*grammar?*
PRP$	NN
PRON	NOUN

As already mentioned, the universal POS-tags are mainly used by the grammar for disambiguating nouns from verbs, and to a lesser degree also adjectives. When the BENG needs to decide between analyzing a word as a noun or a verb, it will assign a higher preference score to the analysis that is congruent with the SyntaxNet analysis, without however discarding the alternative route altogether. For all other words, the BENG does not need help from the statistical model and, as explained further below, it even solves several cases where the statistical model often returns inconsistent results. Language production, in which a meaning is verbalized into an utterance, is performed entirely by the BENG without interfacing with an external model.

The linguistic inventory contains a vocabulary of about 35,000 lemmas of which the open-class words have been automatically extracted from online text. This vocabulary is quite large compared to human L2 speakers, but what makes the BENG a 'basic' grammar is its limited grammatical inventory. At the moment of writing this article, only 40 grammatical constructions have been implemented that are intended to cover basic phrasal patterns (noun and verb phrases; including tense-aspect-modality distinctions), argument structure constructions (intransitive, transitive and resultative constructions), voice (active vs. passive), negation, and speech acts (questions vs. topicalization vs. declaratives).

Constructions are Local Experts. Before turning to more details on how the grammar handles certain linguistic phenomena, it is important to mention two features that make a grammar in FCG different from what is common practice in the development of precision grammars. The first of these features is that FCG does not impose 'linguistic types' on its constructions to specify e.g. what the appropriate values are for a particular feature. Instead of such a top-down mechanism, each construction is an 'expert' in its own right that specifies which conditions must be satisfied before the construction can be applied. Figure 2 illustrates these conditions through the lexical construction for the verb *to break*.

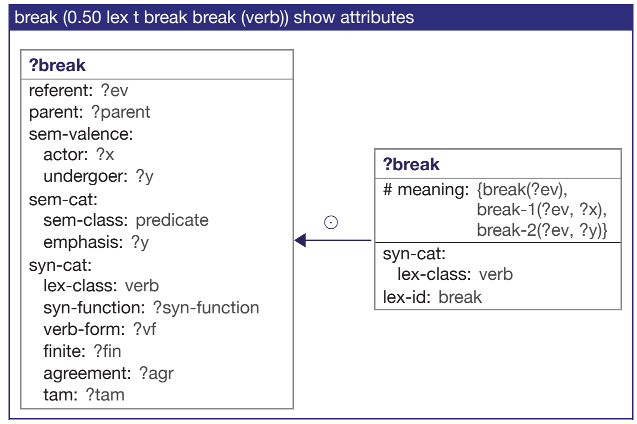


Figure 2: This Figure shows the lexical construction for the verb *to break*. A construction consists of a left- and right-hand side. Units on the right-hand side are always divided into a production lock (above the full line) and a comprehension lock (below the full line).

An FCG construction has a left- and right-hand-side. The left-hand side may specify linguistic information that a construction will always contribute to the linguistic structure that is being processed. Information on the right-hand side, however, is divided between a *production lock* (above the full line) and a *comprehension lock* (below the full line). These locks specify which information needs to be matched with the linguistic structure before the construction is applied. For instance, this lexical construction will only apply in production if a MEANING feature can be found in the input that contains the semantic frame for the verb *to break*. In comprehension, it is the comprehension lock that needs to be matched first. Readers who are interested in the details of constructional application are kindly referred to Steels (to appear).

Constructions Compete With Each Other. A second important difference, particularly with constraint-based formal grammars, is that the construction inventory does not consist of a set of constraints that are assumed to be coherent or simultaneously satisfied at all times. Instead, constructions often compete with each other or may even contain information that conflicts with specifications of other constructions. FCG processing therefore not only has to come up with an adequate solution for a given input, but must also decide which variant is the most appropriate one.

Processing is operationalized as a search problem, as is common practice in FCG (Bleys, Stadler, and De Beule 2011; Steels and Van Eecke 2016), starting with an initial search node that contains all information about the meaning that needs to be verbalized (production) or about the sentence that has been observed (comprehension). New nodes in the search tree are built through the application of constructions, until a solution node is found. For the English grammar, a node is considered as a solution node as soon as there are no more constructions left that are able to apply, so there are no specific grammaticality or other criteria to evaluate the end result.

Constructions themselves are members of one or more specialized construction sets (morphological, lexical, phrasal, functional constructions, and constructions handling argument structure, voice and discourse functions). These labels are used for optimizing linguistic processing, as the search algorithm will always only consider constructions of one label at a time. For instance, morphological constructions are considered first in comprehension because they provide information about the strings of the input sentence, but last in production because then they will provide those strings depending on the grammatical input.

In language production, a basic depth-first search algorithm is applied. This search strategy suffices because production is entirely meaning-driven (i.e. the speaker knows what to say) so there is not the same kind of uncertainty as there is for language comprehension. The uncertainty of language comprehension (e.g. word sense ambiguity) means that the application of a construction can often lead to multiple child nodes in the search tree. In this case, the depth-first search algorithm is enhanced by assigning a priority score to each of these child nodes. Priority is evaluated based on four criteria (probability, semantic coherence, locality and functional scope), which are explained in more detailed elsewhere in this paper.

Achieving Multiple Argument Realization

One salient characteristic of construction grammar is that it allows grammatical constructions to carry meaning similar to lexical constructions. In her seminal work on argument structure constructions, Goldberg (1995) has argued that the syntax-lexicon continuum of construction grammar allows the same verbal construction to occur in different argument realization patterns, as the following examples illustrate:

(2) The window broke.

(3) The boy broke the window.

The English grammar adopts this view and implements Goldbergian-style argument structure constructions as proposed by van Trijp (2011; 2015). More specifically, the meaning of a lexical construction of a verb activates the verb's *semantic frame* and its *frame elements* (Fillmore 1975). For example, a break-event may involve a participant that is the one who breaks another participant. The following examples use a first-order logic representation (symbols with a question mark are variables) for the meaning of the verbs *to break* and *to shout* with theory-neutral labels (e.g. *break-1* or *shout-1*) for the verb's frame elements:

(4) *break(?ev), break-1(?ev, ?x), break-2(?ev, ?y)*

(5) *shout(?ev), shout-1(?ev, ?x), shout-2(?ev, ?y)*

Lexical constructions for verbs provide more information that determine a verb's combinatorial *potential*, without however making any commitment to how the construction will combine with other constructions. For argument structure constructions, two features on the left-hand side of the construction are important: *sem-valence* and *emphasis*. As can be seen in Figure 2, the construction for *to break* repeats the variable names *?x* and *?y* that appeared in the verb's meaning in the *sem-valence* feature to indicate that the frame element *break-1* can potentially be construed as an Actor, and that the frame element *break-2* can potentially be construed as an Undergoer. The feature *emphasis* (a subfeature of *sem-cat*) repeats the variable *?y*, which indicates that the frame element *break-2* (the Undergoer of the event) is emphasized by the verb. By placing the emphasis on the Undergoer, the verb will effectively behave as an ergative verb when it is combined with other constructions. The verb *to shout*, on the other hand, places its emphasis on the Actor, which makes the verb behave as an accusative verb.

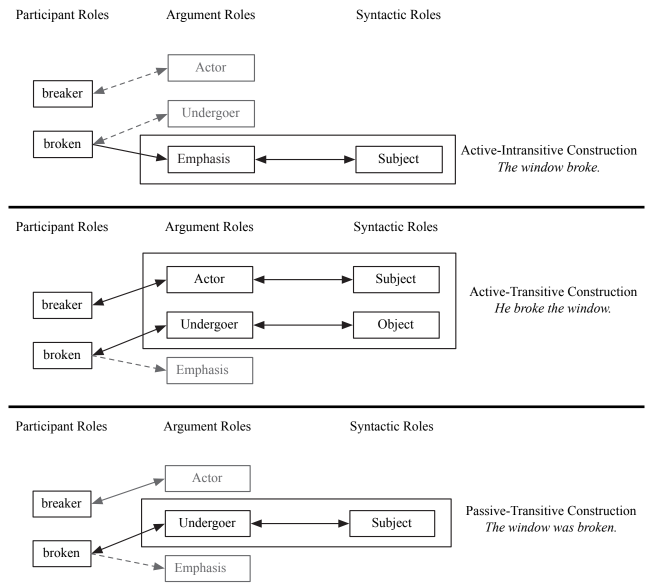


Figure 3: Argument structure constructions select the information they need from a verb's combinatorial potential. The Intransitive Construction (on top) maps the verb's emphasized role (i.e. the same as the Actor for accusative verbs and the Undergoer for ergative verbs) onto Subject. The Transitive Construction (middle) maps Actor and Undergoer onto Subject and Object. The Passive Construction (bottom) maps the Undergoer onto Subject.

Argument structure constructions may use this information for deciding the actual argument realization pattern in which the verb occurs, as illustrated in Figure 3. For instance, the Intransitive Construction will map the emphasized frame element of a verb onto the Subject role. The subject of *the window broke* is thus the Undergoer because *break* is an ergative verb, and the subject of *he shouted* is the Actor because *shout* is an accusative verb. The Transitive Construction, on the other hand, selects the Actor and Undergoer roles and maps them onto Subject and Object respectively. Other argument structure constructions, such as the Resultative Construction, may even add new semantic and syntactic roles to the valence of the verb.

A Functional Approach to Lexico-Phrasal Constructions

The free combination of constructions allows us to do more than apply the same verb in multiple argument realization patterns: it can also be used for implementing a *functional* approach to lexico-phrasal constructions in which the same word form can assume the role of a different part-of-speech. For example, verbal ing-forms or ed-participles often appear as adjectives, as in *the inspiring book*.

Mainstream approaches have mainly focused on the structural properties of hierarchical phrases and therefore treat different uses of the same word form as involving a different lexical item, similar to positing a different verb entry or derivational rule for every argument realization pattern. Functional approaches, on the other hand, hypothesize that hierarchical structures in language emerge because semantics is compositional (Spranger and Steels 2012).

In a functional grammar, parts-of-speech can be considered as the grammaticalization of semantic functions. For instance, nouns typically introduce a class of objects and adjectives typically introduce predicates that can be used for further refining those objects. Nominal phrases are groupings of these functions to establish a *referring expression* (e.g. *the inspiring book*) or an *inquiring expression* (e.g. *which book*). Words that occur frequently enough in a particular usage may be classified and stored as such in order to optimize processing efficiency (e.g. *beginning* may be stored both as a noun or the ing-form of a verb), without however losing the grammar's creativity in allowing words to spontaneously take on the role of a different part-of-speech.

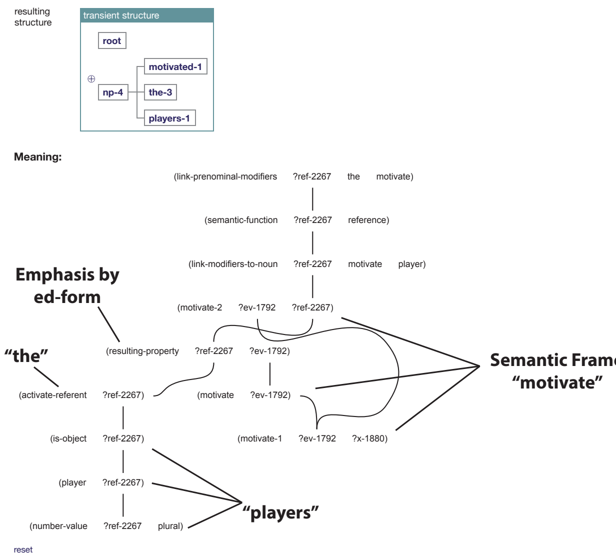


Figure 4: This figure shows the syntactic structure and semantic representation of the NP *the motivated players*. As can be seen in the semantic network, the full semantic frame for *motivate* is included using a prefix predicate-logic representation. The ed-form emphasizes the undergoer-role in this semantic frame (*motivate-2*), which is linked to the head noun *players*, thereby capturing the facts that the players are motivated (and not the motivators).

Using Verbs as Adjectives. The Basic English Grammar puts semantics in the driver's seat and allows words to appear in different functions without requiring additional lexical constructions or derivational rules. Let's consider how the grammar handles ed-participles (which are often syncretic with past tense forms) and ing-forms. The following two examples show different uses for both forms:

(6) a. *The book has* **inspired** *many people.* (ed-particple)
  b. *Michelle Obama* **inspired** *many people with her speech.* (past tense)
  c. *They thanked the goalkeeper for his* **inspired** *saves.* (adjective)

(7) a. *Her speech is* **inspiring** *many people to act.* (progressive ing-form)
  b. *Michelle Obama gave an* **inspiring** *speech.* (adjective)

This kind of part-of-speech (POS) ambiguity is still an unsolved problem in natural language processing (Manning 2011), as POS taggers (even though superficially achieving a near-perfect evaluation score of 97%) often fail to make these kinds of distinctions. However, from the functional perspective adopted by the BENG, there is no need for the POS tagger to make these fine-grained distinctions, because the grammar is free to use the same lexical construction for different functions.

More specifically, a verb's *ed-form* always seems to highlight the result or the undergoer of an event. The different uses of example (6) are therefore semantically motivated: an English Perfect construction typically emphasizes the current relevance or result of an event that occurred in the past, the Passive construction assigns Subject-status to the undergoer of an event, and so on. When used as an adjective, the ed-form indicates that the modified noun is the participant that was affected by the event expressed by the verbal lexical construction, e.g. *the motivated players* are players that have been motivated by someone or something. Figure 4 shows the syntactic structure and semantic analysis of the nominal phrase *the motivated players*. The lexical unit for the word *motivated* was built using the same verbal construction as would be used for other uses of the word. The whole semantic frame of the verb therefore appears in the semantic network, including its two frame elements *motivate-1* (the actor) and *motivate-2* (the undergoer). Notice that the semantics contains a predicate *resulting-property*, which is linked through variable equalities to the *motivate-2* role, thereby indicating that *motivated* expresses the undergoer role.

For ing-forms, a similar story applies: ing-forms all seem to express the event as unbounded or ongoing. Here, the classification of the verb as accusative or ergative is relevant again: when used as an adjective, the ing-form expresses the fact that the modified noun is the actor if the verb is accusative (e.g. *the shouting man*), and that it is the undergoer if the verb is ergative (e.g. *the breaking bridge*).

Grammatical Operators. So how does the English grammar recognize the semantic function of a word and build the desired phrasal structures? The answer is that phrases can be built in a different way for each sentence depending on what kind of grammatical cues are available. When trying to build phrases using the words of the input sentence, the processor will give priority to grammatical words and inflections that are implemented as operators. For instance, a determiner (such as the article *the*) will always build a nominal phrase, after which Modifier-Noun constructions proceed from left-to-right and consider any word as modifier until the head noun is found. Likewise, auxiliaries are reliable cues for detecting when verb forms are part of a verbal phrase. Here is an example of how an NP and VP get recognized through such grammatical words:

(8) *I have read an inspiring book.*

- The pronoun *I* is recognized as an NP.
- The Perfect construction matches on Auxiliary-HAVE and the past-particple *read* and builds a VP.
- The article *an* is a determiner and builds an NP. Next, a Determiner-Nominal construction makes *inspiring* a member of the NP and assigns an adjectival function to it. Finally, a Modifier-Noun construction recognizes *book* as the head of the NP.

In the absence of a determiner, noun phrases are recognized from right to left. In this case, the NP is not built by a determiner, but by the head noun itself. The same is true for verbs: in absence of auxiliaries, the VP construction will build a VP for any verb form that has not been made member of another phrase yet by a different construction:

(9) *I read inspiring books.*

- The pronoun *I* is recognized as an NP.
- Since no grammatical words are present, the NP- and VP-constructions build an NP for *books* and a VP for *read*.
- A Modifier-Noun construction now recognizes *inspiring* as a modifier of *books* and assigns an adjectival function to it.

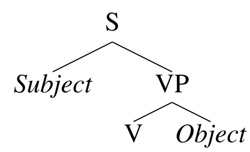
The last sentence is particularly tricky for non-informed NLP systems, because *read* could potentially be an adjective, past-tense form, past participle, or a noun; *inspiring* could be an adjective, progressive verb form, or a gerund; and *books* could be a verb or a noun. Thanks to careful management of how grammatical constructions are applied, the English grammar thus greatly reduces the task of the statistical parser by only requiring it to recognize *books* as a noun, and *read* and *inspiring* as anything else than a noun.

The grammar is also robust in the case of NPs without a head noun. For instance, in the utterance *The inspired are usually happy*, a nominal phrase will be built by the determiner, and *inpired* spontaneously assumes the role of the phrase's head because no noun can be found in the sentence.

Functional Structure and Information Structure Constructions

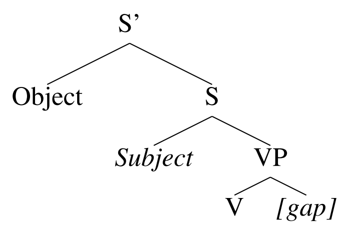
We now come to the clausal level, where we are a.o. concerned with the ordering of a clause's constituents and determining the speech act of a sentence. In mainstream approaches, issues concerning word order and information structure are secondary to syntactic tree representations: grammatical functions such as *subject* and *object* are defined in terms of *argument positions* in the syntax tree, as shown in example (10). All of the semantics of the utterance and decisions concerning word order are directly read off the tree.

(10)



In the case of marked word order, e.g. a long-distance dependency pattern in which the object is topicalized, a problem occurs because the object is no longer in its 'default' argument position (i.e. there is a 'gap'). A widespread solution in formal grammars is the *filler-gap* solution, such as the one adopted by Sag (2010). This solution requires three additional steps. First, some mechanism needs to be implemented that detects that the direct object is gapped and that alters the valence of the verb through some principle or lexical rule. Next, information about the gap must be communicated upwards in the syntax tree in a stepwise fashion. The final step is to introduce a filler-gap-rule that stops the upwards percolation and that identifies the topicalized phrase as the gap's filler. Such an analysis is an example of how word order is explicitly licensed in a hierarchical, top-down fashion through the set of grammar rules. Depending on the grammatical theory, the final structure looks like the one shown in example (11).

(11)



Even though such analyses do not involve actual transformations anymore, it is clear that the default and marked syntactic structures reflect the distinction between deep and surface structures in transformational grammars. Construction grammarians, on the other hand, assume that these so-called 'surface structures' must be studied in their own right instead of considering them to be derived from other structures. Goldberg (2006, p. 25) calls this assumption the 'Surface Generalization Hypothesis,' which again requires that constructions can be freely combined with each other. For example, she argues that the utterance *A dozen roses, Nina bought her mother* involves the same lexical and argument structure constructions as the sentence *Nina sent her mother a dozen roses*, but that the former yields a different pattern because it is combined with a Topicalization construction (ibid. at p. 21).

Through a computational implementation of English argument structure constructions and long-distance dependencies in Fluid Construction Grammar, van Trijp (2014; 2015) has demonstrated that it is perfectly feasible to formalize such an analysis and that it in fact reduces all of the additional steps required in analyses that are restricted to local tree configurations. Moreover, while an explicit Topicalization construction can be redundantly stored in memory for more efficient processing, van Trijp also showed that different word orders may emerge spontaneously as a result of differences in the information structure of the utterance (i.e. which information is deemed topical or focal). The implementation therefore offered a proof-of-concept that deep grammatical analyses are possible even for challenging phenomena such as long-distance dependencies using only a limited set of constructions whose different possible combinations lead to different outcomes. This general design principle has also been used for the implementation of the Basic English Grammar.

Chopping Down the Syntax Tree

The only way to achieve the Surface Generalization hypothesis and the free combination of constructions, is to 'chop down' the syntax tree and no longer treat syntactic structure as the basic representation device from which all other information must be inferred (van Trijp 2016). Instead, constructions may take different perspectives (or a combination of perspectives) on the same linguistic structure, including functional structure and information structure.

Subject-Verb Construction. The English grammar assumes that the functional structure of a sentence (e.g. the assignment of Subject and Object) expresses the *vantage point* that the speaker takes vis-à-vis a state-of-affairs (Dik 1997). The Subject is the *primary perspective*. For example, in an active sentence such as *The boy opened the window*, the scene is presented from the perspective of *the boy*. In a passive sentence such as *The window was opened by the boy*, the same scene is presented from the viewpoint of *the window*.

Subjects often appear immediately in front of the VP, but in non-declarative clauses there is often Subject-Auxiliary Inversion, and in long-distance dependencies the Subject may be distantly instantiated. A very powerful generalization, therefore, is that Subjects should appear before the main verb of their VP. Obviously, there are exceptions to this generalization, but remember that constructions are allowed to compete with each other and that the linguistic inventory is not required to contain a coherent set of constraints.

Figure 5 shows the formal details of the Subject-Verb (SV) Construction. As can be seen, the construction expresses a conceptual meaning in production and has a very underspecified word order constraint in comprehension (i.e. the Subject must *precede* the head of its VP, but the construction does not specify where exactly the Subject may be located). The only feature that I want to draw attention to is the *parent* feature in the *?VP* unit, which establishes an immediate-dominance relation with the *?clause* unit.

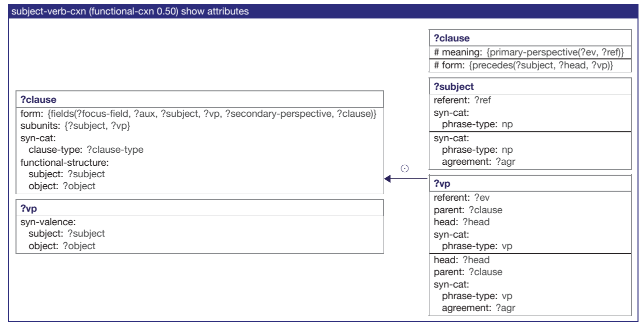


Figure 5: The Subject-Verb Construction is a highly underspecified construction that simply states that an NP that precedes the main verb of a VP may be the subject of the verb. The construction often leads to many hypotheses in comprehension, but the best solution is found through search preferences such as semantic integrity and locality.

The English grammar allows multidominance relations between units (i.e. a unit may have more than one parent) except if a unit has an explicit parent-feature with a pointer to the parent. A VP can only have one parent in the grammar, which is the clause that it governs. Nominal phrases, on the other hand, can be member of multiple clauses. For instance, the NP *a man* is both the direct object of the main clause and the subject of the subclause in *I saw a man running down the street*. By specifying only the parent of the VP, the SV Construction is able to restrict the clause to having only one Subject, but it still allows that Subject to play a different syntactic role in another clause.

Parsing Preferences. Due to its underspecified word order constraint, the construction will often lead to several hypotheses in comprehension. The English grammar will nevertheless be able to reliably retrieve a verb's subject because the best hypothesis is selected based on the parsing preferences of the processor, which include probability, semantic coherence, locality, and functional scope:

- **Probability.** Sometimes, the search algorithm is supplied with more than one construction to consider at the time. This happens when constructions are *competitors*. For example, two morphological constructions may cover the same string, such as the word *work*, which can be both a noun or a verb. In this case, the construction that offers an analysis that is compatible with the statistical parser's suggestions is preferred.
- **Semantic coherence.** If constructional application leads to multiple hypotheses, priority is given to the hypothesis that has the highest degree of semantic coherence. This search preference is inspired by typological research indicating that elements of a sentence tend to show semantic integrity rather than being independent phrases (Dik 1997).

- **Locality.** If constructional application leads to multiple hypotheses, priority is given to the hypothesis that involves elements of the sentence that are closest to each other. This search preference is based on research in functional linguistics and psychology that shows that languages tend to keep information that belongs together close to each other (Gibson 1998; 2000; Hawkins 2004; Rijkhoff 1992).
- **Functional Scope.** If constructional application leads to multiple hypotheses, priority is given to the hypothesis that involves elements with the largest functional scope. This preference is simply a way to capture 'headedness' of phrases.

The search preferences may sound a little bit abstract at first, but we'll illustrate them using the Subject-Verb Construction. Example (12) illustrates the semantic coherence and locality preference:

(12) The book I think is beautiful.

Assuming that all noun and verb phrases are already recognized here, applying the Subject-Verb construction in comprehension would lead to four hypotheses in which *the book* is considered as the subject of *think* and *is*, and in which *I* is considered as the subject of those verbs as well. Since all four solutions have the same semantic coherence effect (i.e. they establish a relation between an NP and a VP), the search algorithm chooses the solution that scores best on the locality principle, which means designating *I* as the subject of *think*.

When the Subject-Verb Construction is considered again, it still yields two hypotheses: both *the book* and *I* could be the subject of *is*. If locality were the dominant search preference, the processor would wrongly choose the hypothesis in which *I* is the subject. That choice would however lead to a solution that has less semantic coherence than the one where *the book* is the subject, because the meaning of *the book* would remain disconnected from the other parts of the sentence. Since semantic coherence outweighs the locality preference, *the book* is correctly assigned the subject role. Both VPs in the sentence now have a subject, so the construction cannot apply anymore.

Now consider example (13).

(13) Students with strong work ethics shoudn't fear a cynical professor.

The challenge of this example is that the subject of the sentence could be either the complex NP *students with strong work ethics* or its subordinate NP *strong work ethics*, because the Subject-Verb Construction fully abstracts away from the location of the Subject in the linguistic structure. In this case, the processor will prefer the solution that involves units with the largest functional scope: since the scope (or domain) of the complex NP is larger than the one of any of its subparts, it is considered as the subject of *fear*.

The highly underspecified Subject-Verb Construction is therefore applicable to any sentence in which a scene is described, whether that scene is described through a main clause, a subclause, or nominal complement; and it applies even when the subject appears in a long-distance relation with its verb. When this construction is applied together with other functional constructions and speech act constructions, the position of the subject in the sentence is robustly detected/assigned as a side-effect of each different combination. Readers who are interested in a detailed account of this interaction, are kindly referred to van Trijp (2014; 2016).

Conclusions and Further Work

This paper reported on the first computational construction grammar for English that achieves both production and comprehension in a large-scale setting. More specifically, the grammar has been designed as a precision model for Basic English, but which is nevertheless able to achieve detailed (albeit sometimes partial) semantic analyses of sentences whose structures go beyond the scope of the constraints of individual constructions in the inventory. This latter feature of the grammar is made possible through the fact that constructions are allowed to be combined freely with each other, which allows structures to emerge spontaneously as a side-effect of different combinations of the same constructions.

The grammar can be considered as a milestone in construction grammar theory, as it operationalizes and demonstrates the feasibility of key properties of constructional language processing in a large-scale setting, including a functional approach to lexico-phrasal processing and a Goldbergian-style approach to argument structure constructions, word order and long-distance dependencies.

Obviously, the efforts reported in this paper are part of an ongoing research program and many important steps need to be taken to advance this work. The most important one is to come up with evaluation criteria for assessing the adequacy of the grammar, particularly in comparison to prior work. This step is non-trivial, because the BENG introduces several design principles that are radically different from existing language technologies (which are often trained on corpora annotated using an incompatible syntactic theory), which means that existing evaluation criteria cannot be applied blindly.

The grammar is also of potential interest to both psycholinguistic research and NLP applications. In linguistic psychology, a lot of recent work has focused on 'good enough' representations for language comprehension (Ferreira, Bailey, and Ferraro 2002), which investigates how partial semantic analyses can lead to robust linguistic behavior. As for NLP applications, this paper has shown how even a basic grammar may solve issues concerning fine-grained classifications in POS-taggers by greatly reducing the complexity of the tagging task, and how it is able to extract detailed semantic analyses from any input. In other words, we are now only beginning to see the potential impact of constructional language processing for both scientific research and language technologies, and many exciting new research avenues have opened up for further exploration.

Acknowledgements The research reported in this paper was conducted at and funded by the Sony Computer Science Laboratory Paris. I am greatly indebted to my colleagues, particularly Luc Steels, for the discussions and crucial ideas that got this project started; Paul Van Eecke, who made the online version of the grammar a reality; Katrien Beuls, for her feedback and for leading the way in scaling FCG grammars; and Miquel Cornudella, for his unwavering help. Many thanks also go to François Pachet, director of Sony CSL Paris, for his trust, and Pratik Bhoir; for assistance in our web endeavours. I would also like to thank Aline Godfroid for directing me to the relevant literature on language acquisition studies. Finally, many thanks go to Jóhanna Barðdal, Hans Boas, Bill Croft, Mirjam Fried, Adele Goldberg and Stefan Müller, with whom I have had several fruitful discussions in the past years about my work and whose work has been a major source of inspiration.

References

Andor, D.; Alberti, C.; Weiss, D.; Severyn, A.; Presta, A.; Ganchev, K.; Petrov, S.; and Collins, M. 2016. Globally normalized transition-based neural networks. *arXiv* 1603.06042. Available at https://arxiv.org/abs/1603.06042.

Bleys, J.; Stadler, K.; and De Beule, J. 2011. Search in linguistic processing. In Steels, L., ed., *Design Patterns in Fluid Construction Grammar*. Amsterdam: John Benjamins. 149–179.

de Marneffe, M.-C.; Dozat, T.; Silveira, N.; Haverinen, K.; Ginter, F.; Nivre, J.; and Manning, C. 2014. Universal Stanford dependencies: A cross-linguistic typology. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)*, 4585–4592. Reykjavik: European Language Resources Association.

Dik, S. C. 1997. *The Theory of Functional Grammar. Part 1: The Structure of the Clause*. Berlin: Mouton De Gruyter.

Ferreira, F.; Bailey, K. G.; and Ferraro, V. 2002. Good-enough representations in language comprehension. *Current Directions in Psychological Science* 11(1):11–15.

Fillmore, C. J. 1975. Against checklist theories of meaning. In Cogen, C., ed., *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*. Berkeley CA: Berkeley Linguistics Society. 123–131.

Fillmore, C. J. 1988. The mechanisms of "Construction Grammar". In *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, 35–55. Berkeley CA: Berkeley Linguistics Society.

Gibson, E. 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition* 68:1–76.

Gibson, E. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In Miyashita, Y.; Marantz, A.; and O'Neil, W., eds., *Image, Language, Brain*. Cambridge, MA: MIT Press. 95–126.

Goldberg, A. E. 1995. *A Construction Grammar Approach to Argument Structure*. Chicago: Chicago UP.

Goldberg, A. E. 2006. *Constructions At Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Hawkins, J. 2004. *Efficiency and Complexity in Grammars*. Oxford: OUP.

Manning, C. D. 2011. Part-of-speech tagging from 97% to 100%: Is it time for some linguistics? In Gelbukh, A. F., ed., *Computational Linguistics and Intelligent Text Processing*, volume 6608 of *Lecture Notes in Computer Science*. Berlin: Springer. 171–189.

Nation, I. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Ogden, C. K. 1968. *Basic English: International Second Language*. New York: Harcourt, Brace & World. A revised and expanded version of 'The System of Basic English', ed. by E.C. Graham.

Petrov, S.; Das, D.; and McDonald, R. 2012. A universal part-of-speech tagset. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)*, 2089–2096. Istanbul: European Language Resources Association.

Rijkhoff, J. 1992. *The Noun Phrase: A Typological Study of its Form and Structure*. Amsterdam: John Benjamins.

Sag, I. A. 2010. English filler-gap constructions. *Language* 86(3):486–545.

Spranger, M., and Steels, L. 2012. Emergent functional grammar for space. In Steels, L., ed., *Experiments in Cultural Language Evolution*. Amsterdam: John Benjamins. 207 – 232.

Steels, L., and Van Eecke, P. 2016. Insight grammar learning using pro- and anti-unification. *forthcoming*.

Steels, L., ed. 2011. *Design Patterns in Fluid Construction Grammar*. Amsterdam: John Benjamins.

Steels, L. to appear. Basics of Fluid Construction Grammar. *Constructions and Frames*. Manuscript available at www.fcg-net.org.

van Trijp, R. 2011. A design pattern for argument structure constructions. In Steels, L., ed., *Design Patterns in Fluid Construction Grammar*. Amsterdam: John Benjamins. 115–145.

van Trijp, R. 2014. Long-distance dependencies without filler-gaps: A cognitive-functional alternative in Fluid Construction Grammar. *Language and Cognition* 6(2):242–270.

van Trijp, R. 2015. Cognitive vs. generative construction grammar: The case of coercion and argument structure. *Cognitive Linguistics* 26:613–632.

van Trijp, R. 2016. Chopping down the syntax tree: What constructions can do instead. *Belgian Journal for Linguistics* 30.