Decomposed phrasal constructions

Petter Haugereid
NTNU Trondheim

Proceedings of the 14th International Conference on
Head-Driven Phrase Structure Grammar

Stanford Department of Linguistics and CSLI's LinGO Lab

Stefan Müller (Editor)

2007

Stanford, CA: CSLI Publications

pages 120–129

Haugereid, Petter. 2007. Decomposed phrasal constructions. In Stefan Müller (ed.), *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar, Stanford Department of Linguistics and CSLI's LinGO Lab*, 120–129. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2007.7.

Abstract

In this paper I suggest an interface level of semantic representations, that on the one hand corresponds to morpho-syntactic entities such as phrase structure rules, function words and inflections, and that on the other hand can be mapped to lexical semantic representations that one ultimately needs in order to give good predictions about argument frames of lexical items. This interface level consists of basic constructions that can be decomposed into five sub-constructions (*arg1-role*, *arg2-role* ... *arg5-role*). I argue in favour of phrasal constructions in order to account for altering argument frames and maybe also coercion without having to use lexical rules or multiple lexical entries.

1 Introduction

Every syntactic theory will have to decide on which component of the grammar shoulders the burden of subcategorization, the lexicon or the syntax. While frameworks like HPSG and LFG are mainly lexicalist, Construction Grammar and some versions of Minimalism are more in favour of letting the syntax do most of the labour.

This paper presents an HPSG-like approach which aims at making a clear distinction between morpho-syntactic elements such as phrase structure rules, function words and inflections on the one hand, and open class lexical items on the other. I believe that open class lexical items do not have grammatical content in the sense that they are assigned a particular category and that they require particular argument frames. The fact that they can be coerced is a strong indication that they do not have any fixed grammatical information in the way that function words and inflections do. I also believe that what Borer (2005, 11) refers to as an "intricate web of layers of a complex perceptual structure and emerging world knowledge" is what open class lexical items are representing. And it is in the end this intricate web of layers that the lexical item represents that makes us prefer a particular category and argument frame.

However, writing a grammar based on such a theory is a huge task, considering the enormous amount of factors involved. What I will focus on in this paper are the syntactic rules, the function words and the inflections that make up the grammatical frame that the open class lexical items appear in. I will also sketch an interface to the "web of layers" that can be employed in order to restrict the number of possible argument frames.

The main objective behind such an approach is to be able to account for altering argument frames and maybe also coercion without having to use lexical rules or multiple lexical entries.

I assume five argument roles that are different from the functional argument roles like Subject and Complement used in the HPSG literature. They are also

[†]Thanks to Lars Hellan, Stefan Müller and four anonymous reviewers for helpful comments.

not necessarily linked to functions like Subject, Direct Object and Indirect Object. They are maybe more inspired by the initial stratum in Relational Grammar (see Blake (1990)). The five roles are not directly linked to a particular syntactic realization. That is, a role can be realized either as a phrase structure rule, as an inflection or as a function word. The argument roles are ultimately assumed to be determined by the semantics of the verb, and correspond vaguely to thematic roles:

- **Arg1-role**: The agent or source.
- **Arg2-role**: The patient.
- **Arg3-role**: The benefactive or recipient.
- **Arg4-role**: The goal.
- **Arg5-role**: The antecedent.[1]

The argument roles function as a meeting point between semantics and syntax. I have intentionally been vague in the semantic definitions above, and the role names *arg1-role*, *arg2-role* etc. are chosen not only because similar names are used in Relational Grammar, but also because they are neutral. One role can correspond to several semantic roles in lexical semantics.

This approach can be seen as an attempt to extract the semantics of syntax. So given a syntactic construction, one can infer certain semantic roles even though one does not get the full lexical semantics. I believe that the full semantic representation comes from the semantics of syntax *plus* the meaning that the open class lexical item represents.

2 Construction Grammar

Goldberg (1995) gives a number of phrasal constructions that independent of the lexical meaning of the words can be said to have a meaning. Examples of such constructions are:

i) *The English Ditransitive Construction* (see (1)), which has the following syntactic active structure: [SUBJ [V OBJ OBJ2]],

ii) *The English Caused-Motion Construction* (see (2)), which has the following syntactic active structure: [SUBJ [V OBJ OBL]],

iii) *The English Resultative Construction* (see (3)), which has the following syntactic active structure: [SUBJ [V OBJ OBL]], and

iv) *The Way Construction* (see (4)), which has the following syntactic active structure: [SUBJ$_i$ [V [POSS$_i$ *way*] OBL]]

(1) Sally baked her sister a cake. (Goldberg, 1995, 141)

[1] I use the term *antecedent* (taken from Croft (1991)) as a collection term for roles like instrument, comitative, manner and source.

(2) They laughed the poor guy out of the room. (Goldberg, 1995, 152)

(3) He talked himself blue in the face. (Goldberg, 1995, 189)

(4) Frank dug his way out of the prison. (Goldberg, 1995, 199)

Typical for verbs appearing in these constructions is that their argument frames are not necessarily predictable from the verb's semantics. In Construction Grammar, the argument frames can be contributed by the constructions, and the meaning is composed by the verb's semantics and the construction it appears in. There is no need to assume several verb meanings for the same stem in order to account for a verb with more than one possible argument frame.

Müller (2006) points out a problem with phrasal Construction Grammar as presented in Goldberg (1995), namely that for example 218 constructions are required in order to account for resultatives in connection with permutations of SUBJ, OBJ and OBL, verb initial/verb final position, passive, middle, modal infinitives and free datives in German. And this leaves out the treatment of adjuncts and complex predicates, which could make the number of constructions needed infinite. Müller's criticism presupposes that the phrasal constructions either are flat, or that they necessitate constraints trees of a depth greater than one. For the German subordinate clauses in (5), he assigns the structures in (6):

(5) a. daß so grün selbst Jan die Tür nicht streicht
that that green even Jan the door not paints
'that not even Jan would paint the door that green'

b. daß so grün die Tür selbst Jan nicht streicht
that that green the door even Jan not paints

c. daß Jan so grün selbst die Tür nicht streicht
that Jan that green even the door not paints

d. daß eine solche Tür so grün niemand streicht
that a such door that green nobody paints
'that nobody paints such a door that green'

(6) a. [OBL SUBJ OBJ V]
b. [OBL OBJ SUBJ V]
c. [SUBJ OBL OBJ V]
d. [OBJ OBL SUBJ V]

What is new in the approach that I am going to suggest here, is that constructions are decomposed into sub-constructions. This makes it possible to maintain binary structures without constraints on threes of a depth greater than one, and at the same time have a phrasal approach to constructions. The examples in (5) can be given the (binary) structures in (7):[2]

[2]COMPL stands for *complementizer*

(7) a. [[[[COMPL ARG4] ARG1] ARG2] V]
b. [[[[COMPL ARG4] ARG2] ARG1] V]
c. [[[[COMPL ARG1] ARG4] ARG2] V]
d. [[[[COMPL ARG2] ARG4] ARG1] V]

Before I explain how this can be achieved, I will discuss the argument roles I am assuming.

3 Argument roles

The five argument roles can have different syntactic realizations, as the examples (8)–(12) illustrate. I here exemplify how the argument roles are realized in English.

Arg1-role: The agent or source. The arg1-role can be realized as an NP subject (see (8a)), as the passive auxiliary (see (8b)) or as the infinitival marker (see (8c)). If the arg1-role is realized as the passive morphology, it cannot be a source.

(8) a. **John** smashed the ball.
b. The ball **was** smashed.
c. (John tried) **to** smash the ball.

Arg2-role: The patient. This role is usually realized as the direct object (see (9a)), but if the sentence is unaccusative or passive, it can be realized as subject (see (9b) and (9c), respectively). The role can also be realized as the infinitival marker (see (9d)). When realized as subject or direct object, the argument can be an NP (see (9a) and (9b)), a subordinate clause (see 9e) or an infinitival clause (see (9f)).

(9) a. John smashed **the ball**.
b. **The boat** arrived.
c. **The ball** was smashed.
d. (The car needed) **to** be washed.
e. John said **that Mary smashed the ball**.
f. John promised **to smile**.

Arg3-role: The benefactive or recipient. This role is usually realized as indirect object (see (10a)), but if the sentence is passive, the role can be realized as subject (see (10b)). It can also be realized as the infinitival marker (see (10c)).

(10) a. John gave **Mary** a book.
b. **Mary** was given the book.
c. (Mary wanted) **to** be given a book.

Arg4-role: The goal. This is either a resultative or an end-of-path, and is realized as a PP, AP or NP complement (see (11a)–(11c)).

(11) a. John smashed the ball **out of the room**.
b. John hammered the metal **flat**.
c. He painted the car **a brilliant red**.[3]

Arg5-role: The antecedent. This is a participant which precedes the patient in the chain of events. It can be instrument, comitative, manner or source. It is realized as a PP complement (see (12)).[4]

(12) John punctured the balloon **with a needle**.

4 Argument frames and valence alternations

I assume that argument frames are made up of constellations of the five argument roles above. Some of the argument frames are exemplified in (13). (13a) has one argument role, the arg1-role, which constitutes an arg1-frame. (13b) has two argument roles, the arg1-role and the arg2-role, and the roles together constitute an arg12-frame. (13c) has one argument role, the arg2-role, which constitutes an arg2-frame. (13d) has three argument roles, an arg1-role, an arg2-role and an arg3-role, and these three roles constitute an arg123-frame. (13e) has three argument roles, an arg1-role, an arg2-role and an arg4-role. The three roles constitute an arg124-frame. (13f) has the three roles arg1-role, arg2-role and arg5-role, which constitute an arg125-frame.

(13) a. John smiles. (arg1-frame)
b. John smashed the ball. (arg12-frame)
c. The boat arrived. (arg2-frame)
d. John gave Mary a book. (arg123-frame)
e. John gave a book to Mary. (arg124-frame)
f. John punctured a balloon with a needle. (arg125-frame)

In this account, valence alternations can be explained in terms of verbs entering different syntactic argument frames that are made up of sub-constructions. Examples (14)–(20) are taken from Levin (1993). I have equipped each example with the corresponding argument frame (in parenthesis).

(14) *Causative/Inchoative Alternation*
a. Janet broke the cup. (arg12-frame)
b. The cup broke. (arg2-frame)

(15) *Unexpressed Object Alternation*

[3]This example is taken from Rothstein (1985, 83)

[4]The distinction between participants that precede the object in the causal chain (what here is referred to as the arg5-role) and participants that follow (the arg4-role) is found in (Croft, 1991, 183-240).

a. Mike ate the cake. (arg12-frame)
b. Mike ate. (arg1-frame)

(16) *Conative Alternation*
a. Paula hit the fence. (arg12-frame)
b. Paula hit at the fence. (arg14-frame)

(17) *Preposition Drop Alternation*
a. Martha climbed up the mountain. (arg14-frame)
b. Martha climbed the mountain. (arg12-frame)

(18) *Dative Alternation*
a. Bill sold a car to Tom. (arg124-frame)
b. Bill sold Tom a car. (arg123-frame)

(19) *Locative Alternation*
a. Jack sprayed paint on the wall. (arg124-frame)
b. Jack sprayed the wall with paint. (arg125-frame)

(20) *Instrument Subject Alternation*
a. David broke the window with a hammer. (arg125-frame)
b. The hammer broke the window. (arg12-frame)

I see the argument frames to constitute general construction types that more specific constructions can inherit from. The arg12-frame in (14a) is for example different from the arg12-frame in (20b) in that (20b) is not agentive. The arg124-frame can be seen to have several subtypes, namely the Caused-Motion Construction ((2)), the Resultative Construction ((3)) and the *Way* Construction ((4)).

Some verbs, like *drip*, can enter a great number of argument frames, as illustrated in (21). Here 14 different argument frames are listed. 8 of them have passive counterparts. If one uses a lexical approach, as suggested by Müller, the number of lexical constructions becomes quite large. It is possible to do with only one lexical entry for *drip* here, since the verb is treated more like a modifier of the syntactic argument frame it appears in, rather than as a head with full control of its syntactic environment.

(21) a. *arg1-frame*:
The roof drips.
b. *arg14-frame*:
The doctor drips into the eyes.
c. *arg15-frame*:
The doctor drips with water.
d. *arg145-frame*:
The doctor drips into the eyes with water.

e. *arg12-frame*:
The roof drips water.

f. *arg124-frame*:
The roof drips water into the bucket.

g. *arg125-frame*:
The doctor dripped the eyes with water.

h. *arg145-frame*:
The doctor dripped into the eyes with water.

i. *arg123-frame*:
John dripped himself two drops of water.

j. *arg1234-frame*:
John dripped himself two drops of water into his eyes.

k. *arg12345-frame*:
John dripped himself two drops of water into his eyes with a drop counter.

l. *arg2-frame*:
Water dripped.

m. *arg24-frame*:
Water dripped into the bucket.

n. *arg0-frame*:
It drips.

5 Analysis

The basic argument frame of a clause is arrived at by letting the morpho-syntactic functional elements in the clause (phrase structure rules, function words and inflections) contribute information about which sub-constructions that have applied by means of types. An item that realizes the arg1-role, will contribute the type *arg1+*, an item that realizes the arg2-role contributes the type *arg2+*, and so on. The argument roles that are not realized will be registered with negative types. When a clause is processed, the argument role types are unified. A transitive clause will have the argument role types *arg1+*, *arg2+*, *arg3–* and *arg4–*.[5] As is shown in the type hierarchy in Figure 1, the unification of the types *arg1+*, *arg2+*, *arg3–* and *arg4–* yields the type *arg12*. Similarly, a ditransitive clause will contribute the argument role types *arg1+*, *arg2+*, *arg3+* and *arg4–*, which unifies as the type *arg123*.

The argument role types, that the morpho-syntactic items contribute, together with the hierarchy of argument frames, account for the possible argument frames. The system allows one to constrain a verb to only enter a specific frame. An unergative intransitive verb will for example be constrained to have an arg1-frame. This constraint is only compatible with the following constellation of argument role

[5] I am not including the arg5-role for expository reasons.

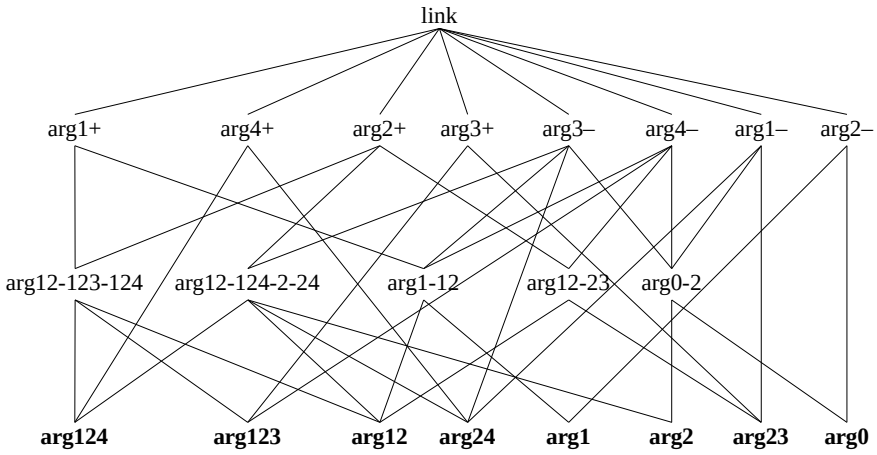

Figure 1: The hierarchy of *argument frame* types

types: *arg1+*, *arg2–*, *arg3–* and *arg4–*. A verb can also be allowed to enter more than one frame. Unexpressed object alternation verbs like *eat* (see (15)) can be constrained to have the argument frame type *arg1-12*. It will then be compatible with two constellations of argument role types, namely *arg1+*, *arg2–*, *arg3–* and *arg4–* and *arg1+*, *arg2+*, *arg3–* and *arg4–* (see Figure 1).[6]

In the approach that I have suggested, permutations and adjunct attachment in German do not pose a problem, since the structures are binary, and there is no need (as Müller claims) to posit constraints on trees of a depth greater than one. The different sub-constructions apply independently, and it is only after the whole clause is processed that it is clear what kind of construction they were a part of.

References

Blake, Barry. 1990. *Relational Grammar*. London and New York: Routledge.

Borer, Hagit. 2005. *Structuring Sense. An Exo-Skeletal Triology. Volume I. In Name Only*. New York: Oxford University Press.

Croft, William. 1991. *Syntactic Categories and Grammatical Relations*. Chicago: University of Chicago Press.

[6]Such an analysis is utilized in a broad-coverage HPSG-like grammar that I have developed for Norwegian, Norsyg. Norsyg parses 40.9% of the items of an article on concrete. The article, which has 313 grammatical items, was taken from Norwegian Wikipedia articles marked as excellent, and no changes were made to the grammar in order to adapt it to the data. A manual inspection of all the items that parsed, using the [incr tsdb()] treebanking tool (Oepen and Flickinger, 1998; Oepen, 2001), revealed that 95 out of 128 items (74.2%) had the intended analysis. This means that Norsyg has a coverage of 30.4% of the grammatical items in the article.

Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.

Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Müller, Stefan. 2006. Phrasal or Lexical Constructions? *Language* 82(4), 850–883.

Oepen, Stephan. 2001. [incr tsdb()] — Competence and Performance Laboratory. User Manual. Technical report, Computational Linguistics, Saarland University, Saarbrücken, Germany, in preparation.

Oepen, Stephan and Flickinger, Daniel P. 1998. Towards Systematic Grammar Profiling. Test Suite Technology Ten Years After. *Journal of Computer Speech and Language* 12(4) (Special Issue on Evaluation), 411 – 436.

Rothstein, Susan D. 1985. *The syntactic forms of predication*. Ph. D.thesis, Bar-Ilan University.