Phenogrammatical Labelling in Convergent Grammar: the Case of Wrap

Lia Mansfield, Scott Martin, Carl Pollard, and Chris Worth
Department of Linguistics
Ohio State University
{lia,scott,pollard,worth}@ling.ohio-state.edu

March 15, 2009

Introduction

We propose a treatment of phenogrammatical (ϕ-) labelling for multimodal categorial grammars (MMCGs) that combines the advantages of earlier methods due to Oehrle (1994) and to Morrill and Solias (1993). We illustrate with a remarkably simple and straightforward analysis of wrap within the framework of convergent grammar (CVG), a kind of MMCG where the tectogrammatical (τ-)modes correspond to traditional notions of grammatical function. The paper is organized as follows. Sections 1 and 2 review the systems of Oehrle and of Morrill and Solias respectively. Section 3 presents our proposed version of ϕ-calculus. Section 4 sketches the basics of (applicative) CVG with ϕ-labelling. And section 5 illustrates our handling of wrap with analyses of English complement wrap, Chamorro VSO order, and English particle 'movement'.

1 Oehrle

Oehrle (1994) proposed simultaneous labelling of (nondirectional, unimodal) CG derivations by both semantic and phenogrammatical[1] terms (hereafter, ϕ-terms). Like familiar (extensional) meaning-term calculi, Oehrle's ϕ-calculus is a standard higher-order logic (HOL),

[1] We use Curry's term *phenogrammar* for what Oehrle, Morrill, Muskens, and de Groote refer to, respectively, as 'phonology', 'prosodics', 'syntax', and 'concrete syntax'; and his term *tectogrammar* for what is usually called 'combinatorics' or 'abstract syntax'. Throughout we abbreviate 'pheno-' and 'tecto-' by ϕ and τ respectively. Roughly speaking, phenogrammar has to do with (1) the temporal ordering of (the phonological realizations of) words and other minimal units of tectogrammar (abstract syntax) such as clitics, phrasal affixes, particles, prosodically realized morphemes with information-structural content, etc.; and (2) the 'flavors' or 'strengths' of the bonds (or boundaries, or junctures) between them.

equipped with one base type (here called St) and the (cartesian) product type constructor ($\times$) in addition to the usual (intuitionistic) implication ($\rightarrow$). Constants of type St are interpreted as phonological words (except for ϵ, which denotes the null phonology); and a constant $\circ$ of type St $\rightarrow$ St $\rightarrow$ St (written infix) is interpreted as concatenation. The usual monoid axioms ensure that in the intended model, St is interpreted as the free monoid generated by the interpretations of the constants. Because the ϕ-term labelling of a derivation parallels the semantic term labelling, it follows that e.g. verbs with n syntactic (and semantic) arguments have n-ary string functions as their ϕ-interpretations. 'Wrappers', such as *took ... to task* are handled by abstracting over a string value occupying the 'insertion point': $\lambda x \lambda y(y \circ \mathit{take} \circ x \circ \mathit{to} \circ \mathit{task})$. This is a useful technique, e.g. for 'quantifier lowering' where 'the gap is really filled', but not for cases (discussed below) where the insertion point has to be 'remembered' after something has been inserted. Variants of Oehrle's general setup have been incorporated into abstract categorial grammar (ACG, de Groote (2001)) and λ-grammar Muskens (2001).

2 Morrill and Solias

Morrill and Solias's (1993) proposal is in the context of a version of CG where each of the two tectogrammatical modes (continuous and discontinuous) has a product type constructor with left and right residuals. To each tectogrammatical mode corresponds a binary *partial* operation (concatenation ($\circ$) and tupling ($\langle\cdot,\cdot\rangle$) respectively) of the algebra into which the ϕ-terms are interpreted (the algebra is untyped by design); thus these are *algebraic* terms, not the terms of a lambda calculus (to say nothing of a HOL). In particular, ϕ-terms never denote functions. As expected, concatenation has a two-sided identity. Because the 'chunks' of a wrapper (i.e. the components of a tuple) must be wrapped around whatever is inserted, additional *projection* operations 1 and 2 are required that recover the components of a tuple. It does not seem entirely clear whether tupling can be iterated to produce ϕ-entities with multiple insertion points. The presence of an untyped partial algebra within an otherwise typed system is somewhat disconcerting. Moreover, the ϕ-labelling of the Left rule for the discontinuous product shares with Oehrle's approach the defect that location of the insertion point is not remembered after the wrapping takes place.

3 A synthesis

As a new candidate ϕ-calculus, we propose the *typed lambda theory of the free monoid*. This is just the positive ($\times$, $\rightarrow$) typed lambda calculus (TLC) with the same signature (base type St and constants for phonological words, null phonology, and concatenation) as Oehrle's calculus. However, our calculus is *only* a TLC, not an HOL, so the monoid axioms belong to the equational theory (term equivalence definition), not to the object language.

Unlike Oehrle's approach to wrap, ours does not abstract over the insertion point. Instead it incorporates the fundamental insight of Morrill and Solias that a wrapper is a tuple.

However, we sidestep the problematic technicalities of their (merely) algebraic approach by treating tupling and projections not as a partial operations on an untyped algebra of ϕ-entities, but rather as the pairing term constructor provided by the TLC itself. That is, wrappers constitute a separate type from St, namely Wr, which is defined to be $\mathrm{St} \times \mathrm{St}$. As a matter of notation, hereafter we use s and t as (metavariables over) terms of type St; w as a term of type Wr; x and y as object-language variables of type St; and u and v as object-language variables of type Wr. Also P is used as a metavariable over ϕ-types.

As a reminder that we are doing phenogrammar not semantics, we write the wrapper $\langle s, t \rangle$ as $s \circ_{\uparrow} t$. We now define $\mathsf{rwrap} : \mathrm{Wr} \rightarrow \mathrm{St} \rightarrow \mathrm{Wr}$ to be the term $\lambda u \lambda x(\pi(u) \circ_{\uparrow} (x \circ \pi'(u)))$. This is interpreted as the the *right wrap* operation (Bach (1979)) that wraps a wrapper around a string by concatenating the string to the front of the right-hand chunk of the wrapper while retaining the insertion point. Likewise Bach's *left wrap* is denoted by the term $\mathsf{lwrap} : \mathrm{Wr} \rightarrow \mathrm{St} \rightarrow \mathrm{Wr}$, defined to be the term $\lambda u \lambda x((\pi(u) \circ x) \circ_{\uparrow} \pi'(u))$.

Note that strings and wrappers are not squeezed into a partial algebra, nor are they (merely) the sorts of a two-sorted algebra. Rather they are two of the domains (type interpretations) in a Henkin model (typed applicative structure) that interprets the ϕ-calculus. As in Oehrle's system, the only operation symbol in the signature is concatenation ($\circ$). The counterparts of Morrill and Solias's tupling and projections come for free courtesy of the positive TLC; and the wrap operations are internally definable.

Among the closed ϕ-terms, only those of type St are considered to be actually sayable. An important role will be played in our grammars by the type-schematized family of functional terms $\mathsf{say}_P : P \rightarrow \mathrm{St}$, which collectively denote a polymorphic function from (certain) ϕ-entities to strings. This function is characterized by the equations[2]:

$$\mathsf{say}_{\mathrm{St}} = \lambda x x$$
$$\mathsf{say}_{\mathrm{Wr}} = \lambda u(\pi(u) \circ \pi'(u))$$

4 CVG Overview

Convergent Grammar (CVG, Pollard (2008)) is a form of MMCG, presented in the Gentzen-sequent style of natural deduction, where the implicative type constructors in the tectogrammar correspond to traditional notions of grammatical function rather than to directionality. Like ACG, CVG has distinct calculi for ϕ-grammar, τ-grammar, and semantics; but for expository simplicity, we suppress the τ-terms and the semantics, displaying only the τ-types and the typed ϕ-terms. Also for the sake of simplicity, we ignore CVG's mechanisms for handling unbounded dependencies and scoping of *in situ* scopal expressions, and consider only the applicative fragment. Thus, for present purposes, typing judgments derived by the grammar are of the form:

$$\vdash p : P, A$$

[2]For a larger fragment that incorporates hypothetical reasoning, additional equations are needed for certain functional ϕ-types.

where $p : P$ is a typed ϕ-term and A is a τ-type[3].

We take the precise inventory of grammatical functions (flavors of implication $\multimap_{\mathrm{F}}$) to be a parameter of cross-linguistic variation. For the purposes of this paper, we take $\textsc{f}$ to range over the set F of grammatical function names $\{\textsc{s}, \textsc{c}, \textsc{p}\}$ (mnemonic for *subject*, *complement*, and *particle* respectively). Thus, besides the ϕ-calculus in the preceding section, a CVG consists of (1) lexical entries such as:

$$\vdash \textit{Fido} : \mathrm{St}, \mathrm{NP}$$

$$\vdash \textit{barked} : \mathrm{St}, \mathrm{NP} \multimap_{\mathrm{S}} \mathrm{S}$$

and (2) elimination rules for the implications, labelled by ϕ-'recipes' that specify the ϕ-term of the conclusion in terms of the ϕ-terms of the premisses. For English, one of the latter is the following, which captures the fact that subjects, unlike other arguments, always linearize to the left of the verb:

Schema $\mathbf{M}_{\mathrm{S}}$ (English Subject Modus Ponens)
If $\vdash p : P, A$ and $\vdash p' : P', A \multimap_{\mathrm{S}} B$, then $\vdash \mathsf{say}(p) \circ \mathsf{say}(p') : \mathrm{St}, B$

Here the ϕ-recipe is to concatenate the ϕ-entities of the subject and the predicate in that order, first making sure that both have been 'spelled out' (converted into something sayable by removing the insertion point if any). This latter precaution is necessary, in light of sentences such as *to err is human*, given that (as we will see directly), English VPs formed from a verb and its complements still 'contain an insertion point' and are therefore *not* sayable.

This allows us to account for idiosyncrasies of English word order without interfering with cross-linguistic generalizations about subjecthood based on tecto-structure. In fact, there is a long and copious literature aimed at defining grammatical functions in terms of combinatoric order of selection by the verb (or obliqueness), best exemplified within the CG tradition by Dowty (1982a,b). A subject is defined tectogrammatically as the last argument selected by a verb, or, alternatively, the least oblique argument of the verb. This is as useful a definition for English, where subjects linearize immediately to the left of the verb, as it is for the VSO language Chamorro, in which they customarily linearize immediately to the right of the verb.

Likewise, within this approach to grammatical functions, a first object is defined as the second least oblique argument, or penultimately selected argument of the verb, a second object as the antepenultimately selected argument, and so on for as many arguments as needed for a given language. Our elimination rules allow us to capture cross-linguistic differences in the way these various types of arguments are realized in surface word orders while preserving cross-linguistic generalizations about the obliqueness hierarchy based on the order in which they are selected tectogrammatically.

[3] In particular, since the implication introduction rules are absent from the fragment, the typing judgments have no contexts.

5 The Grammar of Wrap

5.1 Complement Wrap in English

In English, the first object of a verb is the argument that gets ‘promoted’ to subjecthood in passive constructions. By looking at their passive counterparts, Dowty showed that the first objects of ditransitive verbs are precisely those which linearize adjacent to the verb, as in:

(1) a. John gave Mary the book
b. Mary was given the book (by John)
c. * The book was given Mary (by John)

but, on the other hand,

(2) a. John gave the book to Mary
b. * To Mary was given the book (by John)
c. The book was given to Mary (by John)

This and other evidence was used to motivate the existence of a transitive verb phrase (TVP) – that is, a sign that combines with its first object to yield a (intransitive) verb phrase – which is a constituent in the tectogrammar but discontinuous in the phenogrammar. Crucially, this means that complements in English are linearized in the reverse of the order in which they are ‘selected’, so that the last complement selected is pronounced closest to the verb which selects it.

We capture this by writing the elimination rule for complements in English to maintain an insertion point immediately to the right of the verb. That is to say, all complements in English are right-wrapped in the phenogrammar:

Schema $\mathbf{M}_{\mathrm{CE}}$ (English Complement Modus Ponens)
If $\vdash p : \mathrm{Wr}, A \multimap_{\mathrm{C}} B$ and $\vdash p' : P, A$, then $\vdash \mathsf{rwrap}(p)(\mathsf{say}(p')) : \mathrm{Wr}, B$

Here the ϕ-recipe is to right-wrap the ϕ-entity of the functor around that of the complement (while retaining its insertion point), first making sure that the latter has been rendered sayable (by removing its insertion point if any).

Inasmuch as there is no grammar rule that introduces insertion points into ϕ-entities, these must be already present in the lexical entries of all words that select complements, e.g.:

$$\vdash \mathit{bit} \circ_{\uparrow} \epsilon : \mathrm{Wr}, \mathrm{NP} \multimap_{\mathrm{C}} \mathrm{NP} \multimap_{\mathrm{S}} \mathrm{S}$$

$$\vdash \mathit{gave} \circ_{\uparrow} \epsilon : \mathrm{Wr}, \mathrm{NP} \multimap_{\mathrm{C}} \mathrm{NP} \multimap_{\mathrm{C}} \mathrm{NP} \multimap_{\mathrm{S}} \mathrm{S}$$

The use of wrap and the decoupling of τ- and ϕ-terms therefore allow us to account for idiosyncrasies of English word order while maintaining generalizations about grammatical function that may have cross-linguistic scope.

5.2 Chamorro VSO Linearization

Another longstanding puzzle easily solved by our handling of Wrap is that of VSO word order in languages such as Welsh, Tongan, Syrian Arabic, and (considered here) Chamorro (Chung (1998)). Traditionally, such languages have been considered to display a tension between linear word order and semantic accessibility. This is because verbs form constituents with their objects, not with their subjects, based on the obliqueness hierarchy described above. However, in VSO languages, the verb and object are non-adjacent. We resolve this tension by saying that the verb and object do form a constituent tectogrammatically, but are realized discontinuously in the phenogrammar. In other words, verbs in Chamorro combine with their arguments via right-wrap, whether subject or object.

Schema $\mathrm{M_{CC}}$ (Chamorro Complement Modus Ponens)
If $\vdash p : \mathrm{Wr}, A \multimap_{\mathrm{C}} B$ and $\vdash p' : P, A$, then $\vdash \mathsf{rwrap}(p)(\mathsf{say}(p')) : \mathrm{Wr}, B$

This is the same as the English complement rule, but the role it plays in Chamorro is larger: unlike English, Chamorro typically displays no combinatorial distinction between subjects and complements. The subject combines with the verb via the same mode of implication and therefore, as the last argument selected, is realized between the verb and the object. Moreover, because both VPs and TVPs are are tectogrammatical constituents, we are able to give a straightforward account of VP-modifying or TV-modifying adverbs.[4]

5.3 English Particle Verbs

Particle verbs are among the few constructions in English that display a variable word order with no corresponding difference in meaning. However, Farrell (2005) shows that the two word orders are syntactically distinct. We capture the distinction in question by treating particles that are separated from the verb as normal PP complements consisting of an intransitive preposition. Adjacent particles, however, are joined more tightly to the verb than normal complementation, and no other linguistic material can intervene between them (phenogrammatically speaking). We therefore use a different mode of tectogrammatical combination for this case – Particle rather than Complement. The two lexical entries for a particle verb like *hung* as in the sentences *Harvey hung the picture up* vs. *Harvey hung up the picture* are therefore:

$$\vdash \textit{hung} \circ_{\uparrow} \epsilon : \mathrm{PP} \multimap_{\mathrm{C}} \mathrm{NP} \multimap_{\mathrm{C}} \mathrm{NP} \multimap_{\mathrm{S}} \mathrm{S}$$

for the separated case and

$$\vdash \textit{hung} \circ_{\uparrow} \epsilon : \mathrm{P} \multimap_{\mathrm{P}} \mathrm{NP} \multimap_{\mathrm{C}} \mathrm{NP} \multimap_{\mathrm{S}} \mathrm{S}$$

for the adjacent case.

[4] The rule for modification, not given here, wraps the modifiee around the modifier, preserving the insertion point of the former. This shows that it is not always the tectogrammatical (or semantic) functor, but rather the constituent traditionally viewed as the *head*, that retains its insertion point.

The elimination rule for particles resembles that for subjects and complements, with the crucial difference that now the functor *left*-wraps around the particle:

Schema $\mathbf{M}_{\mathrm{P}}$ (English Particle Modus Ponens)
If $\vdash p : \mathrm{Wr}, A \multimap_{\mathrm{P}} B$ and $\vdash p' : P, A$, then $\vdash \mathsf{lwrap}(p)(p') : \mathrm{Wr}, B$

Thus the particle is inseparably joined to the verb, with the insertion point to its right. This analysis captures the well-known tighter connection between verb and (adjacent) particle than between verb and complement along with the characteristic word-order alternation of English verb-particle constructions.

6 Conclusion

We present a new calculus for ϕ-terms based on positive typed lambda calculus and formalized in the Convergent Grammar of Pollard (2008). This calculus uses the same base type, constants, and concatenation operation as Oehrle (1994), and preserves the basic insight of Morrill and Solias (1993) that wrappers are tuples. However, unlike Oehrle's, our calculus does not provide for abstraction over the insertion point in a ϕ-term and avoids the technical problems encountered by Morrill and Solias's purely algebraic approach. Since it enjoys the additional benefit that the insertion point is 'remembered' in ϕ-terms even after the insertion has taken place, the calculus presented here compares favorably with both of these previous approaches. This remembering of the insertion point proves essential in the cases of wrap we analyze in Chamorro and English, in which we adopt the *right-* and *left wrap* operations of Bach (1979) to analyze discontinuous phenogrammatical constituents.

References

Emmon Bach. Control in Montague Grammar. *Linguistic Inquiry*, 10(4):515–531, 1979.

Sandra Chung. *The Design of Agreement: Evidence from Chamorro.* University of Chicago Press, 1998.

Philippe de Groote. Towards Abstract Categorial Grammars. In *Association for Computational Linguistics, 39th Annual Meetings and 10th Conference of the European Chapter, Proceedings of the Conference*, pages 148–155, 2001.

David Dowty. Grammatical Relations and Montague Grammar. In Pauline Jacobson and Geoffrey K. Pullum, editors, *The nature of syntactic representation.* Reidel, Dordrecht, 1982a.

David Dowty. More on the Categorial Analysis of Grammatical Relations. In A. Zaenen, editor, *Subjects and Other Subjects.* Indiana University Linguistics Club, Bloomington, 1982b.

Patrick Farrell. English verb-preposition constructions: Constituency and Order. *Language*, 81(1):96–137, 2005.

Glyn Morrill and Teresa Solias. Tuples, Discontinuity, and Gapping in Categorial Grammar. In *Proceedings of the sixth conference of the European chapter of the Association for Computational Linguistics*, pages 287–296, Morristown, NJ, USA, 1993. Association for Computational Linguistics.

Reinhard Muskens. Lambda Grammars and the Syntax-Semantics Interface. In R. van Rooy and M. Stokhof, editors, *Proceedings of the Thirteenth Amsterdam Colloquium*, pages 150–155, 2001.

Richard Oehrle. Term-labelled categorial type systems. *Linguistics and Philosophy*, 17: 633–678, 1994.

Carl Pollard. An introduction to convergent grammar. Presented at the Séminaire Calligramme, Nancy, France, June 2008. Available at `http://www.ling.ohio-state.edu/~pollard/cvg/calliho.pdf`.