Categorial grammars

Anna Chernilovskaya
Utrecht Institute of Linguistics OTS, Utrecht University, the Netherlands
anna.chernilovskaya@let.uu.nl

16 May 2008

General philosophy of categorial grammars: In a categorial grammar every natural language word is associated with a category, which describes it as either a function or an argument. In the case of functions, the category specifies the type and directionality of the arguments and the type of the result.

1 Non-associative Lambek calculus

- Basic categories: $BasCat = \{n,\ np,\ pp,\ s\}$ (n = noun, np = noun phrase, pp = prepositional phrase, s = sentence);
- Syntactic categories (logical formulae):

$$\begin{array}{rl} Cat = & A,\ \text{where } A \in BasCat \\ & A \otimes B,\ \text{where } A, B \in Cat \\ & A \backslash B, B/A,\ \text{where } A, B \in Cat \end{array}$$

  - a connective $\otimes$ should be intuitively understood as concatenation of two strings;
  - categories with division (so-called functional categories): combined with a category A from the left (right) produces a category B.

  Note that a connective $\otimes$ is not associative, i.e. $A \otimes (B \otimes C)$ is not the same as $(A \otimes B) \otimes C$. Therefore, parentheses are important. We will always omit the most external parentheses.

- Derivable objects: sequents of the form $A \Rightarrow B$, where A and B are categories;
- An axiom scheme: $A \Rightarrow A$, where A is a basic category;

- Inference rules:

Residuation rules:

$$\frac{A \otimes B \Rightarrow C}{A \Rightarrow C/B} \qquad \frac{A \otimes B \Rightarrow C}{B \Rightarrow A\backslash C}$$

(double-line rules)

Monotonicity rules:

$$\frac{A \Rightarrow C \quad B \Rightarrow D}{A \otimes B \Rightarrow C \otimes D}$$

$$\frac{C \Rightarrow A \quad B \Rightarrow D}{A\backslash B \Rightarrow C\backslash D}$$

$$\frac{C \Rightarrow A \quad B \Rightarrow D}{B/A \Rightarrow D/C}$$

- A lexicon: a set $Lex \subseteq$ 'words' $\times$ Cat ('words' here denotes a set of natural language words).

Exercise 1. Given a lexicon

$$\begin{array}{rlcl} Lex = \{ & \langle \text{Alice} &, & np \rangle \\ & \langle \text{John} &, & np \rangle \\ & \langle \text{boy} &, & n \rangle \\ & \langle \text{book} &, & n \rangle \\ & \langle \text{mathematician} &, & n \rangle \\ & \langle \text{a} &, & np/n \rangle \\ & \langle \text{the} &, & np/n \rangle \\ & \langle \text{annoying} &, & n/n \rangle \\ & \langle \text{to} &, & pp/np \rangle \\ & \langle \text{scolds} &, & (np\backslash s)/np \rangle \\ & \langle \text{teases} &, & (np\backslash s)/np \rangle \\ & \langle \text{dedicated} &, & ((np\backslash s)/pp)/np \rangle \} \end{array}$$

insert instead of every word its category, make an antecedent formula (combine every pair of words in parentheses with $\otimes$ and keep parentheses), put s as a succedent formula and derive that the following natural language phrases have a category s (give a derivation of a sequent: 'antecedent formula $\Rightarrow$ s'):

a) Alice (scolds John)

b) John (teases (the (annoying boy)))

c) John ((dedicated (a book)) (to (a mathematician)))

2 Context free grammar format

- comma (,) denotes concatenation;
- semicolon (;) denotes choice.

Rewriting rules:

Syntactic rules:	Lexical insertion rules:
$s \to np, vp$	$n \to$ girl; mathematician; linguist; book; piece
$np \to pn$	$pn \to$ Bill; John; Molière
$np \to det, np$	$det \to$ a; an; the
$np \to adj, n$	$adj \to$ boring; interesting
$adj \to adv, adj$	$adv \to$ very
$pp \to prep, np$	$prep \to$ about
$vp \to v1$	$v1 \to$ runs
$vp \to v2, np$	$v2 \to$ hates; wrote
$vp \to v3, np, pp$	$v3 \to$ gave
$vp \to v4, pp$	$v4 \to$ talks
$vp \to v5, sbar$	$v5 \to$ thinks
$sbar \to \text{that}, s$	

Exercise 2. Find categories for all words of the following sentences. Do this as follows: first draw a tree of a context free grammar derivation, then for every pair of words designate a functor, which would get a functional category.

a) Bill runs.

b) Bill hates the girl.

c) Bill gave a very boring book to John.

d) A mathematician talks about a linguist.

e) The girl thinks that Molière wrote an interesting piece.

Exercise 3. Here there are some Dutch prepositional phrases, i.e. of a basic category *pp* (under every word you can see its translation to English to help you understand). Assign categories to all words using basic categories *np* ('de boterham', 'de rijen') and *n* ('boterham', 'rijen'). Every word has a single category, so having found a category for a word, you can use it for the rest of the sentences of this exercise.

(1) *bij de boterham*
with the bread
(to go) with the bread

(2) *er bij*
it with
(to go) with it

(3) *voor bij de boterham*
for with the bread

(4) *tussen de rijen*
between the rows
between the rows (static interpretation)

(5) *tussen de rijen door*
between the rows through
passing between the rows (dynamic interpretation)

(6) *er tussen door*
them between through
passing between them (dynamic interpretation)

Exercise 4. This exercise is about English personal pronouns. If you assign a category np to the words 'she' and 'her', you would be able to derive that ungrammatical sentences like 'Her hates the girl' and 'The girl hates she' have a category s. With a category assignment $she \to s/(np\backslash s)$ you can still derive 'She hates the girl', but you block 'The girl hates she'. Your task:

1) Check all derivations mentioned (impose bracketing so that you can use old categories for 'the girl' and 'hates');

2) Find a category for 'her' such that 'The girl hates her' has an s category, but 'Her hates the girl' does not.

Exercise 5. This is more or less the same exercise, but about the Dutch language. As in English, the third person singular masculine and feminine pronouns have a subject form ('hij' = he, 'zij' = she) and an oblique form ('hem' = him, 'haar' = her). The neuter personal pronoun 'het' = it can occur as a subject, but it cannot take the position of a prepositional object (in English you can say: 'The girl talks *about it*', but in Dutch no). Look at the example (8) to see how it works.

To express third person singular neuter pronominalisation in prepositional phrases, Dutch has the pronominal form 'er', which precedes the preposition it combines with rather than following it as a regular prepositional object. Look at the examples (9) and (10) in order to understand. Slash '/' here denotes choice; the star '*' denotes that you cannot insert this word, otherwise a sentence becomes ungrammatical.

(7) *Jan lijkt op Karel.*
Jan looks like Charles.
John resembles Charles.

(8) *Hij/zij/het lijkt op hem/haar/*het.*
He/she/it looks like him/her/it.
He/she/it resembles him/her/it.

(9) *Hij/zij/het lijkt er op.*
He/she/it looks 'er' like.
He/she/it resembles it.

(10) *Hij/zij/het lijkt *op er.*
He/she/it looks *like 'er'.
He/she/it resembles it.

Using basic types s, np and pp, provide categories for all pronouns. Fill in the table below.

WORD	CATEGORY
Jan, Karel	np
lijkt	$(np\backslash s)/pp$
op	pp/np
hij, zij	
hem, haar	
het	
er	

3 Combinatory Categorial Grammar

Remember (analogously to the Exercise 1) the categories:

WORD	CATEGORY
Anna, Manny	np
married, met	$(np\backslash s)/np$ which we will abbreviate as vp/np, where vp stands for a verb phrase, and further as tv (transitive verb)

Note that in this system you do not need parentheses.

The CCG system comprises four group of rules.

1) Function application rules:

$$\begin{array}{ll} X/Y \quad Y \Rightarrow X & (>) \\ Y \quad Y\backslash X \Rightarrow X & (<) \\ (X \text{ and } Y \text{ are categories}) & \end{array}$$

In the course of a derivation, we will use a usual tree format rather than a context free one, e.g. $\dfrac{X/Y \quad Y}{X}>$. With these rules one can derive that *Anna married Manny* is of a category s. Note that there is no need to impose bracketing.

$$\dfrac{\dfrac{Anna}{np} \quad \dfrac{\dfrac{married}{(np\backslash s)/np} \quad \dfrac{Manny}{np}}{np\backslash s}>}{s}<$$

2) Coordination rule:

$$X \quad CONJ \quad X \Rightarrow X \quad \langle\Phi\rangle$$

$CONJ$ is a category for conjunction words like 'and', 'or', 'but', etc.

Now you can derive: *Anna met and married Manny* is of a category s.

3) Composition rules:

$$\begin{array}{llll} X/Y \quad Y/Z \Rightarrow X/Z & (>B) & \text{forward composition} \\ Z\backslash Y \quad Y\backslash X \Rightarrow Z\backslash X & (<B) & \text{backward composition} \end{array}$$

Exercise 6.

a) Derive that *Anna met and might marry Manny* is of a category s with vp/vp as a category for 'might'. Where would you apply forward composition?

b) Show that the following coordination does not work (i.e. there is no way to derive that the following 'sentence' is of a category s): **Anna met Manny and might marry.*

4) Type raising:

$$\begin{array}{ll} X \Rightarrow T/(X\backslash T) & (T) \\ X \Rightarrow (T/X)\backslash T & (T) \end{array}$$

where $X\backslash T$, T/X are *valid* categories of English.

Remark. For example, the following proof is not valid in a categorial grammar of English:

$$\dfrac{s/(np\backslash s) \quad \dfrac{np}{T/(np\backslash T)}\,T}{s/(np\backslash(np\backslash s))}\,>B$$

because $np\backslash(np\backslash s)$ is not a valid English category.

Exercise 7.

a) Show that you can also use type raising rules to derive that *Anna married Manny* is of a category s. However, it is not really important here because as you have shown, you do not need to use type raising here.

b) Derive that *Anna married and I detest Manny* with the categories tv for 'detest' and $s/(np\backslash s)$ for 'I'. This example of coordination makes the use of T explicit.

c) Show that the following 'sentence' cannot be of a category s: **Anna married Manny and I detest.*

d) Let us abbreviate a ditransitive verb: $dtv := (vp/np)/np = tv/np$. For example, a verb 'gave' has this category. How would you use the forward/backward composition and type raising rules to derive the sentence *John gave a teacher an apple and a policeman a flower*?

Exercise 8.

a) Suppose that the following two rules are valid in CCG:

$$\begin{array}{lcl} np/np & \Rightarrow & (s/(np\backslash s))/np \\ s/s & \Rightarrow & (s/(np\backslash s))/np \end{array}$$

Having this in mind, derive that a phrase *[(\[(*The (brother of)), and (John (believes that))\], Pete) sleeps]* has a category s, even though it is not an English sentence. Use the following categories:

the	np/n
brother	n
John, Pete	np
of	$(n\backslash n)/np$
believes	$(np\backslash s)/s'$
that	s'/s
and	$(X\backslash X)/X$ (for some X)

The correct bracketing is given for the sake of simplicity.

b) Come back to the section 1, where the Lambek calculus was discussed. Let us add associativity to its set of rules: $\dfrac{(A \otimes B) \otimes C \Rightarrow D}{A \otimes (B \otimes C) \Rightarrow D}$. Now bracketing does not play any role any more. The result system is called the Associative Lambek calculus.

Derive the two sequents in the Associative Lambek calclus:

$$np/np \Rightarrow (s/(np\backslash s))/np$$
$$s/s \Rightarrow (s/(np\backslash s))/np$$

I.e. rules used in the part (a) above are derivable. The goal of the part (b) is to show that the Lambek calculus with imposed associativity does not describe the English language adequately.

Consider other theoretical possibilities for functional composition rules. In the following exercises we will see that not all of them work for English, but they do so for some other languages.

$$\begin{array}{llcll} X/Y & Y/Z & \Rightarrow & X/Z & (>B) \\ X/Y & Z\backslash Y & \Rightarrow & Z\backslash X & (>B_\times) \\ Z\backslash Y & Y\backslash X & \Rightarrow & Z\backslash X & (<B) \\ Y/Z & X\backslash Y & \Rightarrow & X/Z & (<B_\times) \end{array}$$

Exercise 9. The aim of this exercise is to show that $(>B_\times)$ is not allowed in English.

a) Using the table of categories

think	$(np\backslash s)/s'$
likes	tv
that	s'/s

derive that an (ill-formed) phrase **I John think that likes Manny* is a sentence (which would mean the same as a well-formed *I think John likes Manny*). Note that you have to use $(>B_\times)$ rule.

b) Object extraction: Derive that the following phrase is a valid noun phrase (np) using a category $(n\backslash n)/(s/np)$ for 'who(m)': *a man who(m) I think that John likes.* In this derivation you do not use $(>B_\times)$.

c) Subject extraction: Show that the following phrase does not have a category np having taken a category $(n\backslash n)/(np\backslash s)$ for 'who': **a man who I think that likes John.* However, you can derive this np if you agree to use $(> B_{\times})$.

Exercise 10. This exercise concerns another type of composition rule $(< B_{\times})$. A conclusion is that the categorial grammar of English needs the backward crossed composition rule $Y/Z \quad Y\backslash X \Rightarrow X/Z \quad (< B_{\times})$ in its restricted version, namely where X and Y are of the shape s, $A\backslash s$, or s/A for some category A.

a) Using the following table of categories

buy, cook	vp/np
shall	vp/vp
today, tomorrow	$vp\backslash vp$
mushrooms	n
the	np/n

derive a sentence *John shall buy today and cook tomorrow the mushrooms.* Note where you have to apply the $(< B_{\times})$ rule and check if the condition is satisfied.

b) Now derive that you cannot get a sentence from **[The friend of] [smiled] [the man in the grey suit]* unless you violate the condition imposed on the rule $(< B_{\times})$.

friend, man, suit	n
the	np/n
in, of	$(n\backslash n)/np$
grey	n/n
smiled	vp
shall	vp/vp
today, tomorrow	$vp\backslash vp$

c) The rule $(< B_{\times})$ also helps to derive so-called heavy np-shift: *I shall give to my sister a picture of Rembrandt* with $(vp/pp)/np$ for 'give'. Provide the derivation.

4 References

- Moortgat, M., Categorial grammar and formal semantics. Encyclopedia of Cognitive Science. Nature Publishing Group, Macmillan. 2002.
- Moortgat, M., Categorial Type Logic. In van Benthem and ter Meulen (eds) Handbook of Logic and Language. Elsevier / MIT Press, 1997, pp. 93-177.
- A book that I mentioned (with a lot of linguistics in it):

  Steedman, M., The syntactic process. Cambridge, MA: MIT Press, 2000.
- You could also check a home page of Mark Steedman: http://www.cogsci.ed.ac.uk/~steedman/, where you will find some of his papers and slides of presentations among which I would advise 'Coordination and the Theory of Grammar' (2000) and 'Intonation, Grammar, and Spoken Language Processing' (2002).

There are also other kinds of categorial grammars, for example, Discontinuous Lambek Caculus:

- Morrill, G., M. Fadda and O.Valentín. Nondeterministic Discontinuous Lambek Calculus. Proceedings of the Seventh International Workshop on Computational Semantics, IWCS7, Tilburg, 2007.
- Glyn Morrill's home page: http://www.lsi.upc.edu/˜morrill/

and Symmetric Categorial grammar, which you can study using the web page

- http://symcg.pbwiki.com.
- And the last, but not the least: use wikpedia (preferably in English) for the articles on categorial grammar and combinatory categorial grammar.