Categorial Grammar*

Mark Steedman

Draft 2.0, January 16, 1998 (wordcount=1750)

*To appear in Rob Wilson and Frank Keil (eds.) *The MIT Encyclopedia of Cognitive Sciences*, Cambridge MA, MIT Press. The research was supported in part by NSF grant nos. IRI95-04372, ARPA grant no. N66001-94-C6043, and ARO grant no. DAAH04-94-G0426.

Categorial Grammar

The term Categorial Grammar (CG) names a group of theories of natural language syntax and semantics in which the main responsibility for defining syntactic form is borne by the lexicon. CG is therefore one of the oldest and purest examples of a class of "lexicalized" theories of grammar which also includes HPSG, LFG, TAG, Montague Grammar, Relational Grammar and certain recent versions of the Chomskean theory.

The various modern versions of CG are characterized by a much freer notion of derivational syntactic structure than is assumed under most other formal or generative theories of grammar. All forms of CG also follow Montague 1974 in sharing a strong commitment to the Principle of Compositionality—that is to the assumption that syntax and interpretation are homomorphically related, and may be derived in tandem. Significant contributions have been made by Categorial Grammarians to the study of semantics, syntax, morphology, intonational phonology, computational linguistics and human sentence processing.

Since the problem of formalizing the grammar of natural languages was first defined in its modern form in the 1950's, there have been two styles. Chomsky 1957 and much subsequent work in generative grammar begins by capturing the basic facts of English constituent order exemplified in (1) in a Context-free Phrase Structure Grammar (CFPSG) or system of rewrite rules or "productions" like (2), which have their origin in early work in recursion theory by Post, among others.

(1) Dexter likes Warren.

(2)
$$\begin{array}{lll} S & \rightarrow & NP\ \ VP \\ VP & \rightarrow & TV\ \ NP \\ TV & \rightarrow & \{\text{likes}, \text{sees}, \ldots\} \end{array}$$

Categorial Grammar (CG), together with its close cousin Dependency Grammar (which also originated in the 1950s, in work by Tesnière) stems from an alternative approach to context-free grammar pioneered by Bar-Hillel 1953 and Lambek 1958, with earlier antecedents in Ajdukiewicz 1935 and still earlier work by Husserl and Russell in category theory and the theory of types. Categorial Grammars capture the same information by associating a functional type or category with all grammatical entities. For example, all transitive verbs are associated via the lexicon with a category that can be written as follows:

(3) likes := $(S\backslash NP)/NP$

The notation here is the "result leftmost" notation according to which α/β and $\alpha\backslash\beta$ represent functions from β into α, where the slash determines that the argument β is respectively to the right ($/$) or to the left ($\backslash$) of the functor. Thus the transitive verb (3) is a functor over NPs to its right yielding predicates, or functors over NPs to the left, which in turn yield S. (There are several other notations for categorial grammars, including the widely-used "result on top" notation of Lambek 1958 and much subsequent work, according to which the above category is written $(np\backslash s)/np$. The advantage of the present notation for cognitive scientists is that semantic type can be read in a consistent left-right order, regardless of directionality.)

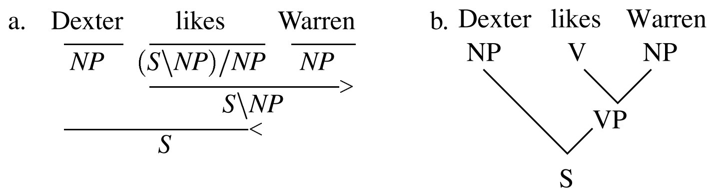
In "pure" context-free CG, categories can combine via two general function application rules, which in the present notation are written as in (4), to yield derivations, written as in (5a), in which underlines indexed with right and left arrows indicate the application of the two rules.

(4) *Functional application*

a. $X/Y \quad Y \quad \Rightarrow \quad X$

b. $Y \quad X\backslash Y \quad \Rightarrow \quad X$

(5)



Such derivations are equivalent to traditional trees like (5b) in CFPSG. However, diagrams like (5a) should be thought of as derivations, delivering a compositional interpretation directly, rather than a purely syntactic structure. The identification of derivation with interpretation becomes important when we consider the extensions of CG that take it beyond weak equivalence with CFPSG.

A central problem for any theory of grammar is to capture the fact that elements of sentences which belong together at the level of semantics or interpretation may be separated by unboundedly much intervening material in sentences, the most obvious example in English arising from the relative clause construction. All theories of grammar respond to this problem by adding something such as the transformationalists' Wh-movement, GPSG Feature-passing, ATN HOLD registers or whatever to a context-free core. Usually, such additions increase automata-theoretic power. To the extent that the constructions involved seem to be quite severely constrained, and that certain kinds of long range dependency seem to be universally prohibited, there is clearly some explanatory value in keeping such power to a minimum.

All of the generalizations of Categorial Grammar respond to this problem by adding various *type-driven combinatory operators* to pure CG. The many different proposals for how to do this fall under two quite distinct approaches. The first, *rule-based*, approach, pioneered by Lyons 1968, Bach 1976, Dowty 1979, among other linguists, and by Lewis 1970 and Geach 1972,, among philosophical logicians, starts from the pure CG of Bar-Hillel, and adds rules corresponding to simple operations over categories, such as "wrap" (or commutation of arguments), "type-raising," (which resembles the application of traditional nominative, accusative etc. *case* to NPs etc.) and functional composition. One possible derivation of a complex relative clause comes out as follows in one fairly typical version, "Combinatory" Categorial Grammar (CCG), discussed at length by the present author (see "Further Reading"), in which type-raising and composition are for historical reasons indicated by **T** and **B**, respectively.

(6) a woman whom Dexter thinks that Warren likes

$$\begin{array}{lllllll}
\text{whom} & \text{Dexter} & \text{thinks} & \text{that} & \text{Warren} & \text{likes} \\
(N\backslash N)/(S/NP) & S/(S\backslash NP)\ {>}\mathbf{T} & (S\backslash NP)/S' & S'/S & S/(S\backslash NP)\ {>}\mathbf{T} & (S\backslash NP)/NP \\
 & \multicolumn{2}{c}{S/S'\ {>}\mathbf{B}} & & & \\
 & \multicolumn{3}{c}{S/S\ {>}\mathbf{B}} & & \\
 & \multicolumn{4}{c}{S/(S\backslash NP)\ {>}\mathbf{B}} & \\
 & \multicolumn{5}{c}{S/NP'\ {>}\mathbf{B}} \\
\multicolumn{6}{c}{N\backslash N\ {>}}
\end{array}$$

Notice that this analysis bears no resemblance to a traditional right-branching clause structure modified by structure-preserving movement transformations.

The alternative, *deductive*, style of Categorial Grammar, pioneered by van Benthem 1986 and Moortgat 1988 takes as its starting point Lambek's syntactic calculus. The Lambek system embodies a view of the categorial slash as a form of logical implication for which a number of axioms or inference rules define a proof theory. (For example, functional application corresponds to the familiar classical rule of *Modus Ponens* under this view). A number of further axioms give rise to a deductive calculus in which many but not all of the rules deployed by the alternative rule-based generalizations of CG are theorems. For example, the derivation (6) corresponds to a proof in the Lambek calculus using type-raising and composition as lemmas.

The differences between these approaches make themselves felt when the grammars in question are extended beyond the weak context-free power of the Lambek calculus and the combinatory rules that are theorems thereof, as they must be to capture natural language in an explanatory fashion. The problem is that almost any addition of axioms corresponding to the non-Lambek combinatory rules that have been proposed in the rule-based framework causes a collapse of the calculus into "permutation completeness"—that is, into a grammar that accepts all permutations of the words of any sentence it accepts. This forces the advocates of the Lambek calculus into the "multi-modal" systems involving many distinct slashes encoding multiple notions of implication (Morrill 1994), and forces the advocates of rule based systems to impose type restrictions on their rules. (Nevertheless, Joshi et al. 1991 show that certain rule-based CGs remain of low automata-theoretic power.)

These two styles of CG are reviewed and compared at length by Moortgat, (with a deductive bias) and Wood, (with a rule-based bias)—see "Further Reading". To some extent the same biases are respectively exhibited in the selection made in two important collections of papers edited by Buszkowski *et al.* and Oehrle *et al.* (see "Further Reading"), which include several of the papers cited here.

The differences are less important for the present purpose than the fact that all of these theories have the effect of engendering derivational structures that are much freer that traditional surface structures, while nevertheless guaranteeing that the non-standard derivations deliver the same semantic interpretation as the standard ones. For example, since all of these theories allow the residue of relativization *Dexter thinks that Warren likes* in example (6) to be a derivational constituent of type S/NP, they also all allow a non-standard analysis of the canonical sentence *Dexter thinks that Warren likes these flowers* in terms of an identically derived constituent followed by an object NP:

(7) [[Dexter thinks that Warren likes]$_{S/NP}$[these flowers]$_{NP}$]$_{S}$

This is a surprising property, because it seems to flout all received opinion concerning the surface constituency of English sentences, suggesting that a structure in which objects—even embedded one—dominate subjects is as valid as the standard one in which subjects dominate objects. The implication is that the "binding theory" (which must explain such facts as that in every language in the world you can say the equivalent of *Warren and Dexter shave each other* but not **Each other shave Dexter and Warren*) must be regarded as a property of semantic interpretation or logical form rather than of surface structure as such (cf. Dowty 1979, Szabolcsi 1989,, Chierchia 1988, Hepple 1990, Jacobson 1992).

These proposals also imply that there are many semantically equivalent surface derivations for every traditional one, a problem that is sometimes misleadingly referred to as "spurious ambiguity", and which appears to make parsing more laborious. However, this problem can be eliminated using standard chart-parsing techniques with an equivalence check on logical forms associated with constituents, as proposed by Karttunen 1989 and other advocates of unification-based computational realizations of CG—see Carpenter 1997 for a review.

Flexible or combinatory Categorial Grammars of all kinds have real advantages for capturing a number of phenomena that are problematic for more traditional theories of grammar. For example, as soon as the analysis in (7) is admitted, we explain why similar fragments can behave like constituents for purposes of coordination:

(8) [[I dislike]$_{S/NP}$, but [Dexter thinks that Warren likes]$_{S/NP}$[these flowers]$_{NP}$]$_{S}$

(Other even more spectacular coordinating nonstandard fragments are discussed by Dowty 1988,.)

We also explain why intonation seems similarly able to treat such fragments as phrasal units in examples like the following, in which % marks an intonational boundary or break, and capitalization indicates stress (cf. Oehrle 1988, and Prevost 1995):

(9) Q: I know who YOU like, but who does DEXTER like?
A: [DEXTER likes]$_{S/NP}$ % [WARREN]$_{NP}$

Moreover, the availability of semantic interpretations for such non-standard constituents appears under certain plausible assumptions about the relation of the competence grammar to the processor to simplify the problem of explaining the availability to human sentence processors of semantic interpretations for fragments like *the flowers sent for*, as evidenced by the effect of this content in b below in eliminating the "garden-path" effect of the ambiguity in a, discussed by Crain and the present author 1985 and Altmann 1988.

(10) a. The doctor sent for the patient died.
b. The flowers sent for the patient died.

All of these phenomena imply that the extra structural ambiguity engendered by generalized categorial grammars is not "spurious," but a property of competence grammar itself.

References

Ajdukiewicz, Kazimierz. 1935. Die syntaktische Konnexität. In *Polish Logic 1920-1939*, ed. Storrs McCall. Oxford University Press, 207-231. translated from *Studia Philosophica*, 1, 1-27.

Altmann, Gerry. 1988. Ambiguity, Parsing Strategies, and Computational Models. *Language and Cognitive Processes* 3:73–98.

Bach, Emmon. 1976. An Extension of Classical Transformational Grammar. In *Problems in Linguistic Metatheory: Proceedings of the 1976 Conference at Michigan State University*, 183–224.

Bar-Hillel, Yehoshua. 1953. A Quasi-arithmetical Notation for Syntactic Description. *Language* 29:47–58.

Carpenter, Robert. 1997. *Type-Logical Semantics*. Cambridge MA: MIT Press. In press.

Chierchia, Gennaro. 1988. Aspects of a Categorial Theory of Binding. In *Categorial Grammars and Natural Language Structures*, ed. Richard T. Oehrle, Emmon Bach, and Deirdre Wheeler. 153–98. Dordrecht: Reidel. Proceedings of the Conference on Categorial Grammar,Tucson, AR, June 1985.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Crain, Stephen, and Mark Steedman. 1985. On not being led up the garden path: the use of context by the psychological parser. In *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*, ed. L. Kartunnen D. Dowty and A. Zwicky. 320–358. ACL Studies in Natural Language Processing. Cambridge University Press.

Dowty, David. 1979. Dative Movement and Thomason's Extensions of Montague Grammar. In *Linguistics, Philosophy, and Montague Grammar*, ed. S. Davis and M. Mithun. Austin: University of Texas Press.

Dowty, David. 1988. Type-raising, Functional Composition, and Non-constituent Coordination. In *Categorial Grammars and Natural Language Structures*, ed. Richard T. Oehrle, Emmon Bach, and Deirdre Wheeler. 153–198. Dordrecht: Reidel. Proceedings of the Conference on Categorial Grammar,Tucson, AR, June 1985.

Geach, Peter. 1972. A Program for Syntax. In *Semantics of Natural Language*, ed. Donald Davidson and Gilbert Harman. 483–497. Dordrecht: Reidel.

Hepple, Mark. 1990. *The Grammar and Processing of Order and Dependency: a Categorial Aproach*. Doctoral dissertation, University of Edinburgh. Tech Report, Centre for Cognitive Science, University of Edinburgh.

Jacobson, Pauline. 1992. The Lexical Entailment Theory of Control and the Tough Construction. In *Lexical Matters*, ed. Ivan Sag and Anna Szabolcsi. 269–300. Chicago: CSLI/Chicago UP.

Joshi, Aravind, K.Vijay-Shanker, and David Weir. 1991. The Convergence of Mildly Context-sensitive Formalisms. In *Processing of Linguistic Structure*, ed. Peter Sells, Stuart Shieber, and Tom Wasow. 31–81. Cambridge MA: MIT Press.

Karttunen, Lauri. 1989. Radical Lexicalism. In *Alternative Conceptions of Phrase Structure*, ed. M.R. Baltin and A.S. Kroch. Chicago University Press, Chicago.

Lambek, Joachim. 1958. The Mathematics of Sentence Structure. *American Mathematical Monthly* 65:154–170.

Lewis, David. 1970. General semantics. *Synthèse* 22:18–67.

Lyons, John. 1968. *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.

Montague, Richard. 1974. *Formal philosophy: papers of Richard Montague*. New Haven: Yale University Press. Edited by Richmond H. Thomason.

Moortgat, Michael. 1988. *Categorial Investigations*. Doctoral dissertation, Universiteit van Amsterdam. publ. Foris, Dordrecht, 1989.

Morrill, Glyn. 1994. *Type-logical Grammar*. Kluwer, Dordrecht.

Oehrle, Richard T. 1988. Multidimensional Compositional Functions as a Basis for Grammatical Analysis. In *Categorial Grammars and Natural Language Structures*, ed. Richard T. Oehrle, Emmon Bach, and Deirdre Wheeler. 349–390. Dordrecht: Reidel. Proceedings of the Conference on Categorial Grammar,Tucson, AR, June 1985.

Prevost, Scott. 1995. *A Semantics of Contrast and Information Structure for Specifying Intonation in Spoken Language Generation*. Doctoral dissertation, University of Pennsylvania, Philadelphia, PA.

Szabolcsi, Anna. 1989. Bound Variables in Syntax: Are there any? In *Semantics and Contextual Expression*, ed. Renate Bartsch, Johan van Benthem, and P. van Emde Boas. 295–318. Dordrecht: Foris.

van Benthem, Johan. 1986. *Essays in Logical Semantics*. Dordrecht: Reidel.

Further Reading

Buszkowski, Wojciech, Witold Marciszewski, and Johan van Benthem (eds). 1988. *Categorial Grammar*. Amsterdam: John Benjamins.

Moortgat, Michael. 1997. Categorial Type Logics. In *Handbook of Logic and Language*, ed. Johan van Benthem and Alice ter Meulen. 93–177. Amsterdam: North Holland.

Oehrle, Richard, Emmon Bach, and Deirdre Wheeler. 1988. *Categorial Grammars and Natural Language Structures*. Dordrecht: Reidel.

Steedman, Mark. 1996. *Surface Structure and Interpretation*. Cambridge Mass.: MIT Press. Linguistic Inquiry Monograph, 30.

Wood, Mary McGee. 1993. *Categorial Grammar*. Routledge.

Index