

Parsing Intentions of Speech

Ke Wang¹, Gerald Penn², Rongpei Wang³

^{1,3} Dalian University of Technology, 116024 Dalian, China

²University of Toronto, 10 King's College Road, M5S 3G4, Canada

¹wang.coco.ke@gmail.com, ²gpenn@cs.toronto.edu, ³rpwang1@hotmail.com

Abstract. This paper is aimed at revealing the inner structures of intentions of speech through demonstrating a method of parsing intentions of speech. The intentions of speech are parsed separately, which is inspired by Searle's notation $F(p)$. In this research, both the propositional content and the illocutionary forces are represented as semantic frames, the categories of which are combined together with syntactic rules of Categorical Grammar. This method is implemented in a grammar system programed in ALE a logic programming language based on the logic of Typed Feature Structures.

1 Introduction

The studies of intentions are aimed at designing intelligent agents which can understand human's intentions through communication. In the 80s and 90s of last century, many scientists coming from many fields began to research in this area. And some results of those researches have been applied to very practical usages. James Allen has developed his ideas in his early study of intention [1] into a language processing model TRIPS [2] which is used as a very important section in the intelligent task learning system PLOW [3]. Barbara Grosz and Candace Sidner developed their theory of shared plan [10] into a model of collaboration COLLAGEN [11] which is later used as the most critical part of a series of robotic systems [12]. Philip R. Cohen and Hector Levesque [7, 8] considered the intention of a conversation as a persistent goal over the whole joint activity, which is formalized as a true proposition in some possible world. Later, Levesque developed his idea about possible worlds into a logic programming language GOLOG [14].

However, there is still one problem left for the researchers to conquer; that is the inner structures of intentions of speech. In Allen et al.'s paper [2], though they "need to produce a detailed semantic representation of what was said-something that captures what the user meant by the utterance"; however, at that time, "the only way to get such system is to build it by hand". They couldn't make it because there has not yet been available a complete theory of the role of intentions in communication: how smaller intentions combine to form composite ones.

This paper is an attempt to reveal the syntactic-semantic structure of intentions of speech, and hence how the small intentions of an utterance are composed as a whole. This research is inspired by Searle's notation $F(p)$, i.e. the illocutionary force and propositional content are parsed separately. Both the illocutionary force and propositional content are represented as semantic frames; and the combination rules

are borrowed from Categorical Grammar. In order to implement our ideas, we constructed a grammar system which is programmed in ALE, a logic programming language based on the logic of typed feature structure. We will introduce it in detail in section 3. And in section 4, we will discuss about how to broaden the coverage of the utterances with probability-based methods.

2 Methods used in this paper

The main idea of the study of intentions of speech can be traced back to J. L. Austin's theory about speech act [4]. He describes speech act as three levels, i.e., locution, illocution, and perlocution. For example, a speaker said "it's dark" meaning that the light is not bright enough. Austin calls this kind of "saying something" locutionary act. By saying "it's dark", the speaker might be conveying his view about the situation, and may indirectly convince the hearer to turn on the light. This is called illocutionary act by Austin. The result of "saying so" is that the hearer turned on the light; and this is exactly what the speaker wants him to do. The final immediate result of "saying so" is named perlocutionary act. John R. Searle employs the term "illocution", but he does not accept Austin's distinction. He [17] argues that there is a distinction between illocutionary force and propositional content of the illocutionary act, as shown in the following sentences:

- (1) Sam smokes habitually.
- (2) Does Sam smoke habitually?
- (3) Sam, smoke habitually!
- (4) Would that Sam smoked habitually.

The first sentence is an assertion; the second one is a question; the third is an order; and the fourth a wish or desire. They are of different illocutionary acts, though they share the same propositional content. Searle uses $F(p)$ to represent the relationship between illocutionary force and propositional content, where F represents the illocutionary force, and P represents the propositional content. In this paper, we adopt Searle's notation; and we assume intention of speech could be rewritten as $F(p)$.

Searle's study opens up a possibility for us to look into the detail structure of intentions of speech. Searle's notation $F(p)$ inspired us, that we could parse the intentions of speech by recognizing the illocutionary force and the propositional content separately. In this paper the propositional content is represented by Semantic Frame [9], for usually propositions could be roughly described as a relationship held by the predicate and its arguments. For example, in (5), <bring> is the predicate; and "you", "me", and "book" are the arguments.

- (5) Could [[you]^{GIVER} <bring> [the book]^{THEME} [to me]^{RECEIVER}]^{PROPOSITION?}

Searle also suggests to "...distinguish two elements in the syntactical structure of the sentence, which we might call the propositional indicator and the illocutionary force indicator." And he views the illocutionary indicators as a kind of devices, by which the illocutionary act is performed. In modern English, those indicators includes at least: "word order, stress, intonation contour, punctuation, the mood of the verb", and so on. For example, (5) is a question. We recognize it as a question because the auxiliary "could" was moved to the front of the utterance and because we saw a question mark at the end of the utterance. In other words, those illocutionary force indicators activated a question frame in our minds, thus we could recognize it as a question. For example, this question could then be rewritten into (6):

(6) [[Could]^{AUXILIARY} [you bring the book to me]^{PROPOSITION}] [?]^{Q-MARKER} QUESTION

However, Frame Semantics does not provide any composition rules. In this paper, we take advantage of the syntactic rules of Categorical Grammar [13, 20] to combine the categories. For instance, the semantic structure of the proposition of (5) could be obtained by the way of (7):

(7) Could you bring the book to me ?
 GIVER <bring>GIVER/RECEIVER/THEME THEME RECEIVER
 <bring>\GIVER/RECEIVER
 <bring>\GIVER
 <bring>

The semantic structure of illocutionary force of (6) could be obtained by the way of (8), where "[... ..]" represents the very abstract question frame (PROP=PROPOSITION). We say that the auxiliary "could" evokes the question frame, or in other words the question frame is assumed when we meet "could"; and then the rest of the frame are tested through the following steps of parsing.

(8) Could [... ..] you bring the book to me ?
 AUX query/Q-MARKER/PROP/AUX PROPOSITION Q-MARKER
 query/Q-MARKER/PROP
 query/Q-MARKER
 query

Note that PROPOSITION in (8) should be further structured as the frame labeled with <bring> in (7). They are combined together by an operation called "node-unification"; refer to the details in section 3.3.

3 Grammar System

The grammar system includes a type hierarchy and a set of grammar rules. The type hierarchy composes of a taxonomy of intentions of speech (refer to 3.1), and some

linguistic knowledge, such as classification of verbs (over 600 English verbs), definitions of illocutionary force indicators, definitions of semantic roles (over 140 semantic roles), and so on. In the type hierarchy, each type is in accord with a set of features. The classification of verbs is aimed at assigning appropriate arguments to the predicate through treating the arguments as the features of verbs. The grammar rule is an integration of Frame Semantics and Categorical Grammar (see Section 2). This grammar system is only fit for parsing simple sentences, i.e., compounded sentences, like if-then-construction, are not covered.

3.1 Taxonomy of Intentions of Speech

Searle classified illocutionary acts into five types, i.e., assertive, directive, expressive, commissive, and declaration. According to Searle [18, p12-15], an assertive is to “commit the speaker ... to the truth of the expressed proposition”, like (1) above; directive is “to get the hearer to do something”, such as (3); expressive is “to express the psychological state ... specified in the propositional content”, such as (4); declaration is about how a “successful performance (of a speech act, noted by authors) guarantees that the propositional content corresponds to the world”, like “you are fired”; and commissive is to “commit the speaker to some future course of action”, for instance, “I will come”. Apparently, Searle’s classification is not specific enough if we want to use it for parsing. Besides, they are not concerned with the linguistic features of utterances, such as, for which features that an utterance is to be recognized as a query, and for which features an utterance is to be identified as an imperative. In this research, we classified intentions of speech into more specific types keeping both their linguistic manifestations and their philosophical significance into perspective, such as Searle’s philosophy about social construction, see Fig. 1.

The reasons for that the intentions of speech are classified into “to describe the world”, “query”, and “imperative” are simple. We are driven to speak, because we want to transmit information, or we want to ask for information, or we want to give an order. Roughly, “to describe the world” corresponds to assertive; and “query” and “imperative” are equal to Searle’s directive. But we have some disagreements with Searle on the classification of assertive, expressive, and commissive, and on the explanation of declaration force. We think both expressive and commissive are assertive, when we look at the inner world from outside. If assertive is “to commit the speaker to something’s being the case, to the truth of the expressed proposition”, then commissive is “to commit the speaker to some future course of action” by truthfully stating the inner world of the speaker. For example, when a speaker said “I will come”, he must already have had a plan in his mind. Similarly, an expressive is to “express the psychological state” of the speaker. Searle didn’t think of them as assertive, probably because that it is hard to detect the truth condition of people’s inner world. But we assume that all of these happen under the truthfulness conditions, so it is possible and feasible for an observer to detect people’s inner world. This is one of the reasons that we separate “third personal behavior” from “brute facts”. The other disagreement is about the force of declaration. Searle thinks the truth condition of this kind of illocutionary act depends on some institutions, so he calls the facts caused by declarations, such as nominating, “institutional fact”. Institutional fact is also called

social fact by Searle in [19]. However, the truth condition of many facts does not obviously depend on the power of an institution, such as “snow is white”. The reason that W-H-I-T-E represents the meaning white is that it is recognized by the social members who use it as a symbol to describe colors. Or in other words it is comprehensible because the social members reached an agreement that this word represents the color of snow. So, we define the intentions of this kind of utterances as to describe “social fact”. The discussion of linguistic philosophy may need a book-length, so we just stop at here. We hope that we could have another opportunity to discuss it in a later paper.

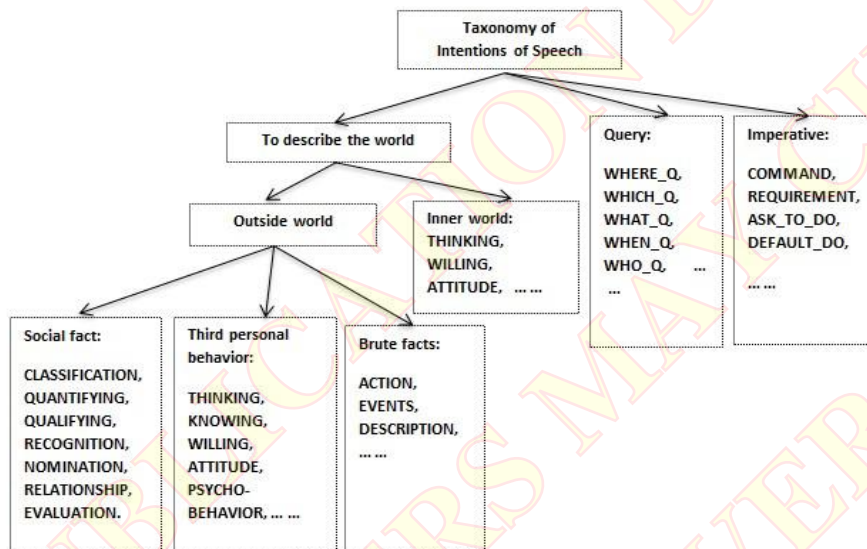


Fig. 1: Taxonomy of Intentions of Speech.

Table 1 shows some syntactic features of utterances, according to which intentions of speech are classified into types. As it is shown, each type of utterances is described as a distinctive chunk labeled with a semantic role, such as CLASS in “classification”, and a set of syntactic elements, such as “be”, “action_verb”, “?”, and so on. Some of the distinctive chunks are realized by a specific set of words, such as ATTRIBUTE is realized by attribute adjectives. Some distinctive chunks are realized by specific sets of intentions of speech. For example, CLUE in “recognition” is realized by the implementation of “brute fact”, and the WILL in “willing” is realized by the implementation of “future fact”. This is the way that some complex intentions of speech are composed; and this is also the way how the grammar rules are compiled, such as (5) through (8).

Table 1: Syntactic features of utterances.

Intentions & Distinctive chunks	Linguistic features	Examples
---------------------------------	---------------------	----------

classification	TARGET + be + CLASS	Apple is a kind of fruit.
CLASS	class words	fruit, teacher, thief, ...
quantifying	NUMBER + TARGET	... ten dollars
qualifying	ATTRIBUTE + TARGET	red shirt
ATTRIBUTE	attribute adjectives	red, tall, good, bad, ...
recognition	DEMONSTRATIVE + be + TARGET + CLUE	... This is the book I bought yesterday
CLUE	brute_fact	I bought yesterday
nomination	TARGET/DEMONSTRATIVE + NAME	This/He is Jack.
NAME	proper names	Jack, Harvard University
relationship	ITEM_1 + relation_verb/be + ITEM_2	Three plus three is six.
evaluation	TARGET + ATTRIBUTE + than + STANDARD	He is taller than me.
	TARGET + ATTRIBUTE+CLASS + SCOPE	He is the tallest student in his class.
brute_fact	PROPOSITION (past and present tense)	Here comes the bus.
PROPOSITION	frame	
3rd_thinking	Third person+think/... .. + THINK	He thought this is correct/I like it.
THINK	social_fact/third_personal_behavior	this is correct/I like it
3rd_knowing	Third person+know/... ..+KNOWN	He knows that guy is Jack/she loves chocolate/she bought the book... ..
KNOWN	“to describe the world”	That guy is jack/she loves chocolate/she bought the book
willing	COGNIZER + promise/wish/... .. + WILL	I wish he will come to the party.
WILL	future_fact	He will come to the party
Wh_query	Wh_pronoun + QUERY + ?	When could you bring me the book?
QUERY	query	
query	auxiliary + PROPOSITION + ?	Could you bring me the book?
imperative	PROPOSITION (action_verb in present tense)	Listen! Let's go!

3.2 Parsing

The grammar system is implemented in ALE [5], a logic programming language based on the logic of typed feature structure [6, 15]. A typed feature structure is a directed graph possibly with cycles. The nodes on the graph are labeled with types; and the edges between nodes are labeled with features, see Fig. 2 A typed feature structure could be represented either by a graph, like (a), or by an AVM, like (b).

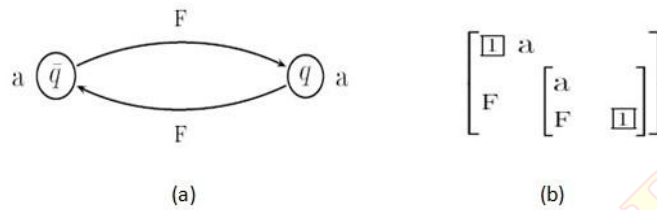


Fig. 2: The graphic and AVM representation of feature structure

We chose to describe semantic frame with typed feature structure; for one reason, it is more similar to the concept of frame, the other reason is that dependence-tree and phrase-structure-tree are not suitable for representing intentions of speech, as shown in Fig. 3.

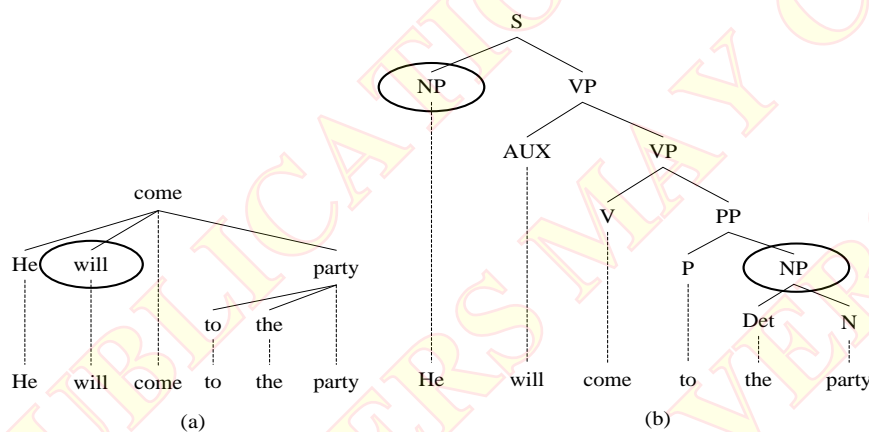


Fig. 3: The dependence-tree and phrase-structure tree of “he will come to the party”.

In Fig. 3 above, (a) and (b) are the dependence tree and the phrase-structure tree respectively. Apparently, in (a) “will” is not argument of “come”, though it syntactically depends on it. In (b), the two NPs are not on the same level, but semantically, they are equally treated as the arguments of the predicate. If either dependence-tree or phrase-structure tree were employed, the parsing task would be made too complicated to be accomplished.

The only one operation in the logic of typed feature structure is Unification. This is very similar to the unification operation in PROLOG except that it is constrained by “type consistency”. For example, in (9), whether the two feature structures are unifiable depends on if “b” and “c” are consistent. Carpenter [6, p12] used to give a definition of consistency. According to him, they are unifiable, only when *b* is a subtype or supertype of *c*.

$$(9) \begin{bmatrix} a \\ F: [b] \end{bmatrix} \sqcup \begin{bmatrix} a \\ F: [c] \end{bmatrix} = \text{true or false?}$$

In order to parse the semantic structures of intentions of speech, we need to translate the application rules of Categorical Grammar into unification operation. We replace “/” and “\” in categorical grammar with “ \sqcup ” the unification operation in the logic of typed feature structure. For example, the semantic structure of (8) could be obtained through a series of operations as the followings¹. The order of the feature structures on the left side of “=” indicates the parsing order.

$$(10) \begin{bmatrix} giver \\ GIVER_E: [giver_element] \end{bmatrix} \sqcup [you] = \begin{bmatrix} giver \\ GIVER_E: [you] \end{bmatrix}$$

$$(11) \begin{bmatrix} receiver \\ RECEIVER_E: [receiver_element] \\ SYN: [receiver_particle] \end{bmatrix} \sqcup [to] \sqcup [me] = \begin{bmatrix} receiver \\ RECEIVER_E: [me] \\ SYN: [to] \end{bmatrix}$$

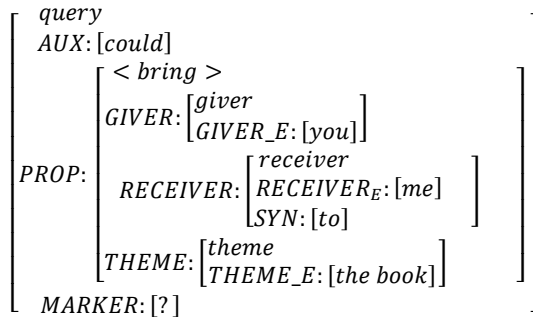
$$(12) \begin{bmatrix} theme \\ THEME_E: [theme_element] \end{bmatrix} \sqcup [the\ book] = \begin{bmatrix} theme \\ THEME_E: [the\ book] \end{bmatrix}$$

$$(13) \begin{bmatrix} giver \\ GIVER_E: [you] \end{bmatrix} \sqcup \begin{bmatrix} < bring > \\ GIVER: [giver] \\ RECEIVER: [receiver] \\ THEME: [theme] \end{bmatrix} \sqcup \begin{bmatrix} theme \\ THEME_E: [the\ book] \end{bmatrix} \sqcup$$

$$\begin{bmatrix} receiver \\ RECEIVER_E: [me] \\ SYN: [to] \end{bmatrix} = \begin{bmatrix} < bring > \\ GIVER: \begin{bmatrix} giver \\ GIVER_E: [you] \end{bmatrix} \\ RECEIVER: \begin{bmatrix} receiver \\ RECEIVER_E: [me] \\ SYN: [to] \end{bmatrix} \\ THEME: \begin{bmatrix} theme \\ THEME_E: [the\ book] \end{bmatrix} \end{bmatrix}$$

$$(14) [could] \sqcup \begin{bmatrix} query \\ AUX: [auxiliary] \\ PROP: [frame] \\ MARKER: [q_marker] \end{bmatrix} \sqcup \begin{bmatrix} < bring > \\ GIVER: \begin{bmatrix} giver \\ GIVER_E: [you] \end{bmatrix} \\ RECEIVER: \begin{bmatrix} receiver \\ RECEIVER_E: [me] \\ SYN: [to] \end{bmatrix} \\ THEME: \begin{bmatrix} theme \\ THEME_E: [the\ book] \end{bmatrix} \end{bmatrix} \sqcup [?] =$$

¹ Considering the convenience of reading and the length of this paper, because the actual semantic structure of an utterance could be very complex, we only show some simple examples as the illustration of parsing.



(10) through (12) are to label each category with a semantic role. It is assumed that what semantic role a category is to have depends on the category's lexical meaning and on the category's syntactic environment. In (10), as long as "you" is predefined as a subtype of "giver_element", it is unifiable with the "node" [giver_element]. In (11), "me" is labeled with "receiver", for it is preceded by particle "to". Usually, in similar constructions, "me" is replaceable, such as "bring it to Jack", "bring it to him", etc., but "to" is relatively irreplaceable. So, it is reasonable to say that "to" is functioned as a kind of "semantic role indicator" (refer to Appendix). (13) is to parse the propositional content. Each argument is fulfilled with a typed category by the way of "node-unification". (14) is intended for parsing the intentions of speech, i.e., the propositional content and other syntactic elements are to be merged with the nodes of the predefined illocutionary force frame. "Could", "?", these illocutionary force indicators that are considered as the features of "query", are planted at "AUX" and "MARKER" respectively; and the proposition <bring> is unified with the node [frame].

4 Discussion and Future Works

In this paper, we presented an idea and a method of parsing intentions of speech. The idea to parse the illocutionary force and the propositional content separately is inspired by Searle's notation $F(P)$. In order to implement this idea, we developed a grammar system. Theoretically, it is very hard to test the coverage of the grammar system. It at present includes about 50 rules, more than 600 verbs, and hundreds of nouns; however, this doesn't mean it could only process 50 utterances. The limitation of the grammar system is that each utterance to be parsed has to correspond to a set of grammar rules. But, we believe it is hopeful to develop methods to broaden the coverage of this grammar system. For example, since it is assumed that each language with infinite utterances could be described by a set of finite grammar rules; therefore, it is possible and reasonable to cut a compounded sentence into clauses and then send these clauses to the grammar system. With these parsed clauses, we could, with manual labor, obtain a probability table which could be about how possible it is to predict two clauses to be joined together. Besides, we are planning to construct an intelligent agent to test our hypothesis, i.e. finer semantic representation will promote the quality of understanding intentions of humans.

References

1. J. F. Allen, and C. R. Perrault, Analyzing intention in Utterances. *Artificial Intelligence*, Vol. 15, p143-178 (1980)
2. J. F. Allen et al., Towards Conversational Human-computer Interaction, *AI Magazine* (2001)
3. J. F. Allen et al., PLOW: A Collaborative Task Learning Agent. *National Conference on Artificial Intelligence*, Vancouver, BC (2007)
4. J. L. Austin, *How to Do Things with Words*, Cambridge University Press (1962)
5. B. Carpenter, and G. Penn, The Attribute Logic Engine: User's Guide. This is downloadable from www.cs.toronto.edu/~gpenn/ale.html
6. B. Carpenter, *the Logic of Typed Feature Structures*, Cambridge University Press (1992)
7. P. Cohen, and H. Levesque, Persistence, Intention, and Commitment in Intentions in Communication edited by Philip R. Cohen et al., the MIT Press (1990)
8. P. Cohen, and H. Levesque, Rational Interaction as the Basis for Communication in Intentions in Communication edited by Philip R. Cohen et al, the MIT Press (1990)
9. C. Fillmore, The Case for Case in Bach and Harms (edit): *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston, page 1-881 (1968)
10. B. Grosz, and C. Sidner, Plans for Discourse in Intentions in Communication edited by Cohen, Philip R. et al. The MIT Press, London, England (1990)
11. C. Rich, et al, COLLAGEN Applying Collaborative Discourse Theory to Human-computer Interaction. *AI Magazine*, Vol. 22, No. 4 (2001)
12. C. Rich, and C. Sidner, Robots and Avatars as Hosts, Advisors, Companions, and Jesters, *AAAI Spring symposium* (2009)
13. J. Lambek, The Mathematics of Sentence Structure, *American Math Monthly*, Vol. 65, No. 3 (1958)
14. H. Levesque, et al., GOLOG: A Logic Programming Language for dynamic domains, *Journal of Logic Programming*, Vol. 31, p59-84 (1997)
15. G. Penn, The Algebraic Structure of Attributed type Signatures, doctoral thesis of Carnegie Mellon University (2000)
16. R. Perrault, An Application of Default Logic to Speech Act Theory in Intentions in Communication edited by Philip R. Cohen et al, the MIT Press (1990)
17. J. R. Searle, *Speech Acts: An Essay in the philosophy of Language*. Cambridge University Press (1969)
18. J. R. Searle, *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge University Press (1979)
19. J. R. Searle, *the Construction of Social Reality*. Allen Lane the Penguin Press (1995)
20. M. M. Wood, *Categorial Grammars*, Routledge, London and New York (1993)

Appendix:

Table 2: Semantic role indicators

Semantic roles	Semantic role indicators	Semantic roles	Semantic role indicators
PLACE	at, above, after, around, beyond, by,	GROUND	from

	in, inside, on, outside, within, round	ATTITUDE	against
SCOPE	near, under, up, upon, before, behind, below, between, over, to	SUPPORT	against, on
POSITION	across, among, between, beyond, in, within, under	REFERENCE	than, like, against
PATH	out, back	AGENT	by
TIME	at, after, ago, around, before, between, beyond, by, during, in, inside, on, outside, over, within, for, near, past, round, since, through, till, until	COGNIZER	to
DIRECTION	at, in, to, for, towards, into	REASON MEANS	since, by by, through
PURPOSE	at, for, to	COMPANY	with, and
ASPECT	at	CAUSE	at, by, with, from, through, under
CAUSER	with, for, about	TOOLS	with
ORIGIN	from	WAY	at, by, in, on
ADDRESSEE	to	MANNER	with, in,
NAME	under	TOPIC	on, over, about
		DISTANCE	for
		INGREDIENT	from
		RECEIVER	to
