ON REPRESENTING GOVERNED PREPOSITIONS AND HANDLING "INCORRECT" AND NOVEL PREPOSITIONS

Hatte R. Blejer, Sharon Flank, and Andrew Kehler

SRA Corporation
2000 15th St. North
Arlington, VA 22201, USA

ABSTRACT

NLP systems, in order to be robust, must handle novel and ill-formed input. One common type of error involves the use of non-standard prepositions to mark arguments. In this paper, we argue that such errors can be handled in a systematic fashion, and that a system designed to handle them offers other advantages. We offer a classification scheme for preposition usage errors. Further, we show how the knowledge representation employed in the SRA NLP system facilitates handling these data.

1.0 INTRODUCTION

It is well known that NLP systems, in order to be robust, must handle ill-formed input. One common type of error involves the use of non-standard prepositions to mark arguments. In this paper, we argue that such errors can be handled in a systematic fashion, and that a system designed to handle them offers other advantages.

The examples of non-standard prepositions we present in the paper are taken from colloquial language, both written and oral. The type of error these examples represent is quite frequent in colloquial written language. The frequency of such examples rises sharply in evolving sub-languages and in oral colloquial language. In developing an NLP system to be used by various U.S. government customers, we have been sensitized to the need to handle variation and innovation in preposition usage. Handling this type of variation or innovation is part of our overall capability to handle novel predicates, which are frequent in sub-language. Novel predicates created for sub-languages are less "stable" in how they mark arguments (ARGUMENT MAPPING) than general English "core" predicates which speakers learn as children. It can be expected that the eventual advent of successful speech understanding systems will further emphasize the need to handle this and other variation.

The NLP system under development at SRA incorporates a Natural Language Knowledge Base (NLKB), a major part of which consists of objects representing SEMANTIC PREDICATE CLASSES. The system uses hierarchical knowledge sources; all general "class-level" characteristics of a semantic predicate class, including the number, type, and marking of their arguments, are put in the NLKB. This leads to increased efficiency in a number of system aspects, e.g., the lexicon is more compact and easier to modify since it only contains idiosyncratic information. This representation allows us to distinguish between lexically and semantically determined ARGUMENT MAPPING and to formulate general class-level constraint relaxation mechanisms.

1.1 CLASSIFYING PREPOSITION USAGE

Preposition usage in English in positions governed by predicating elements, whether adjectival, verbal, or nominal, may be classified as (1) lexically determined, (2) syntactically determined, or (3) semantically determined. Examples are:

LEXICALLY DETERMINED:
   laugh at, afraid of
SYNTACTICALLY DETERMINED:
   by in passive sentences
SEMANTICALLY DETERMINED:
   move to/from

Preposition usage in idiomatic phrases is also considered to be lexically determined, e.g., *with respect to*.

1.2 A TYPOLOGY OF ERRORS IN PREPOSITION USAGE

We have classified our corpus of examples of the use of non-standard

prepositions into the following categories: (1) substitution of a semantically appropriate preposition -- either from the same class or another -- for a semantically determined one, (2) substitution of a semantically appropriate preposition for a lexically determined one, (3) false starts, (4) blends, and (5) substitution of a semantically appropriate preposition for a syntactically determined one. A small percentage of the non-standard use of prepositions appears to be random.

1.3 COMPUTATIONAL APPLICATIONS OF THIS WORK

In a theoretical linguistics forum (Blejer and Flank 1988), we argued that these examples of the use of non-standard prepositions to mark arguments (1) represent the kind of principled variation that underlies language change, and (2) support a semantic analysis of government that utilizes thematic roles, citing other evidence for the semantic basis of prepositional case marking from studies of language dysfunction (Aitchison 1987:103), language acquisition (Pinker 1982:678; Menyuk 1969:56), and typological, cross-linguistic studies on case-marking systems.

More theoretical aspects of our work (including diachronic change and arguments for and against particular linguistic theories) were covered in that paper; here we concentrate on issues of interest to a computational linguistics forum. First, our natural language knowledge representation and processing strategies take into account the semantic basis of prepositional case marking, and thus facilitate handling non-standard and novel use of prepositions to mark arguments. The second contribution is our typology of errors in preposition usage. We claim that an NLP system which accepts naturally occurring input must recognize the type of the error to know how to compensate for it. Furthermore, the knowledge representation scheme we have implemented is an efficient representation for English and lends itself to adaptation to representing non-English case-marking as well.

There is wide variation in computational strategies for mapping from the actual natural language expression to some sort of PREDICATE-ARGUMENT representation. At issue is how the system recognizes the arguments of the predicate.

At one end of the spectrum is an approach which allows any marking of arguments if the type of the argument is correct for that predicate. This approach is inadequate because it ignores vital information carried by the preposition. At the other extreme is a semantically constrained syntactic parse, in many ways a highly desirable strategy. This latter method, however, constrains more strictly than what humans actually produce and understand. Our strategy has been to use the latter method, allowing relaxation of those constraints, under certain well-specified circumstances.

Constraint relaxation has been recognized as a viable strategy for handling ill-formed input. Most discussion centers around orthographic errors and errors in subject-verb agreement. Jensen, Heidorn, Miller, and Ravin (1983:158) note the importance of "relaxing restrictions in the grammar rules in some principled way." Knowing which constraints to relax and avoiding a proliferation of incorrect parses however, is a non-trivial task. Weischedel and Sondheimer (1983:163ff) offer cautionary advice on this subject.

There has been some discussion of errors similar to those cited in our paper. Carbonell and Hayes (1983:132) observed that "problems created by the absence of expected case markers can be overcome by the application of domain knowledge" using case frame instantiation. We agree with these authors that the use of domain knowledge is an important element in understanding ill-formed input. However, in instances where the preposition is not omitted, but rather replaced by a non-standard preposition, we claim that an understanding of the linguistic principles involved in the substitution is necessary.

To explain how constraint relaxation is accomplished, a brief system description is needed. Our system uses a parser based on Tomita (1986), with modifications to allow constraints and structure-building. It uses context-free phrase structure rules, augmented with morphological, contextual, and semantic constraints. Application of the phrase structure rules results in a parse tree, similar to a Lexical-Functional Grammar (LFG) "c-structure" (Bresnan 1982). The constraints are unified at parse time to produce a functionally labelled template (FLT). The FLT is then input to a semantic translation module. Using ARGUMENT

111

MAPPING rules and other operator-operand semantic rules, semantic translation creates situation frames (SF). SFs consist of a predicate and entity frames (EF), whose semantic roles in the situation are labeled. Other semantic objects are relational frames (e.g. prepositional phrases), property frames (e.g. adjective phrases), and unit frames (measure phrases). During the semantic interpretation and discourse analysis phase, the situation frame is interpreted, resulting in one or more instantiated knowledge base (KB) objects, which are state or event descriptions with entity participants.

2.0 REPRESENTING ARGUMENT MAPPING IN AN NLP SYSTEM

In our lexicons, verbs and adjectives are linked to one or more predicate classes which are defined in the Natural Language Knowledge Base (NLKB). Predicates typically govern one or more arguments or thematic roles. All general, class-level information about the thematic roles which a given predicate governs is represented at the highest possible level. Only idiosyncratic information is represented in the lexicon. When lexicons are loaded the idiosyncratic information in the lexicon is unified with the general information in the NLKB. Our representation scheme has certain implementational advantages: lexicons are less error-prone and easier to modify, the data are more compact, constraint relaxation is facilitated, etc. More importantly, we claim that such semantic classes are psychologically valid.

Our representation scheme is based on the principle that ARGUMENT MAPPING is generally determined at the class-level, i.e., predicates group along semantic lines as to the type of ARGUMENT MAPPING they take. Our work draws from theoretical linguistic studies of thematic relations (e.g., Gruber 1976, Jackendoff 1983, and Ostler 1980). We do not accept the "strong" version of localism, i.e., that all form mirrors function -- that ARGUMENT MAPPING classes arise from metaphors based on spatial relations. Unlike case grammar, we limit the number of cases or roles to a small set, based on how they are manifested in surface syntax. We subsequently "interpret" roles based on the semantic class of the predicate, e.g., the GOAL of an ATTITUDE is generally an animate "experiencer".

For example, in the NLKB the ARGUMENT MAPPING of predicates which denote a change in spatial relation specifies a GOAL argument, marked with prepositions which posit a GOAL relation (*to*, *into*, and *onto*) and a SOURCE argument, marked with prepositions which posit a SOURCE relation (*from, out of, off of*). A sub-class of these predicates, namely Vendler's (1967) achievements, mark the GOAL argument with prepositions which posit an OVERLAP relation (*at*, *in*). Compare:

> **MOVE** to/into/onto
> from/out of/off of

> **ARRIVE** at/in
> from

The entries for these verbs in SRA's lexicon merely specify which semantic class they belong to (e.g., SPATIAL-RELATION), whether they are stative or dynamic, whether they allow an agent, and whether they denote an achievement. Their ARGUMENT MAPPING is not entered explicitly in the lexicon. The verb *reach*, on the other hand, which marks its GOAL idiosyncratically, as a direct object, would have this fact in its lexical entry.

2.1 GROUPING SEMANTIC ROLES

Both on implementational and on theoretical grounds, we have grouped certain semantic roles into superclasses. Such groupings are common in the literature on case and valency (see Somers 1987) and are also supported by cross-linguistic evidence. Our grouping of roles follows previous work. For example, the AGENT SUPERCLASS covers both animate agents as well as inanimate instruments. A GROUND SUPERCLASS (as discussed in Talmy 1985) includes both SOURCE and GOAL, and a GOAL SUPERCLASS includes GOAL, PURPOSE, and DIRECTION.

Certain semantic roles, like GOAL and SOURCE, as well as being sisters are "privatives", that is, opposites semantically.

Our representation scheme differentiates between lexically and semantically determined prepositions. We will show how this representation facilitates recognition of the type of error, and therefore principled relaxation of the constraints. Furthermore, a principled

relaxation of the constraints depends in many instances on knowing the relationship between the non-standard and the expected prepositions: are they sisters, privatives, or is the non-standard preposition a parent of the expected preposition.

In the following section we present examples of the five types of preposition usage errors. In the subsequent section, we discuss how our system presently handles these errors, or how it might eventually handle them.

3.0 THE DATA

We have classified the variation data according to the type of substitution. The main types are:

(1) semantic for semantic (Section 3.1),
(2) semantic for lexical (Section 3.2),
(3) blends (Section 3.3),
(4) false starts (Section 3.4), and
(5) semantic for syntactic (Section 3.5).

The data presented below are a representative sample of a larger group of examples. The current paper covers the classifications which we have encountered so far; we expect that analysis of additional data will provide further types of substitutions within each class.

3.1 SEMANTIC FOR SEMANTIC
3.1.1 To/From

The substitution of the goal marker for the source marker cross-linguistically is recognized in the case literature (e.g., Ikegami 1987). In English, this appears to be more pronounced in certain regional dialects. Common source/goal alternations cited by Ikegami (1987:125) include: *averse from/to, different from/to, immune from/to,* and *distinction from/to.* The majority of examples involve *to* substituting for *from* in lexical items which incorporate a negation of the predicate; the standard marker of GROUND in this class of predicates is a SOURCE marker, e.g., *different from.* The "positive" counterparts mark the GROUND with GOAL, e.g., *similar to,* as discussed in detail in Gruber (1976). Variation between *to* and *from* can only occur with verbs which incorporate a negative, otherwise the semantic distinction which these prepositions denote is necessary.

(1) *The way that he came on to that bereaved brother completely alienated me TO Mr. Bush.* 9/26/88 MCS

(2) *At this moment I'm different TO primitive man.* 10/12/88 The Mind, PBS

3.1.2 To/With

Communication and transfer of knowledge can be expressed either as a process with multiple, equally involved participants, or as an asymmetric process with one of the participants as the "agent" of the transfer of information. Our data document the substitution of the GOAL marker for the CO-THEME marker; this may reflect the tendency of English to prefer "agent" focussing. The participants in a COMMUNICATION situation are similar in their semantic roles, the only difference being one of "viewpoint." By no means all communication predicates operate in this way: e.g., EXPLANATION, TRANSFER OF KNOWLEDGE are more clearly asymmetric. The system differentiates between "mutual" and "asymmetric" communication predicates.

(3) *The only reason they'll chat TO you is, you're either pretty, or they need something from your husband.* 9/30/88 MCS

(4) *I'll have to sit down and explore this TO you.* 10/16/88

3.2 SEMANTIC FOR LEXICAL
3.2.1 Goal Superclass (Goal/ Purpose/ Direction)

Goal and purpose are frequently expressed by the same case-marking, with the DIRECTION marker alternating with these at times. The standard preposition in these examples is lexically determined. In example (6), instead of the lexically determined *to,* which also marks the semantic role GOAL, another preposition within the same superclass is chosen. In example (5) the phrasally determined *for* is replaced by the GOAL marker. There is abundant cross-linguistic evidence for a GOAL SUPERCLASS which includes GOAL and PURPOSE; to a lesser extent DIRECTION also patterns with these cross-linguistically.

(5) *It's changing TO the better.* 8/3/88 MCS

(6) *Mr. Raspberry is almost 200 years behind Washingtonians aspiring FOR full citizenship.* 10/13/88 WP

113

3.2.2 On/Of

Several examples involve lexical items expressing knowledge or cognition, for which the standard preposition is lexically determined. This preposition is uniformly replaced by *on*, also a marker of the semantic role of REFERENT. Examples include *abreast of, grasp of, an idea of,* and *knowledge of*. We claim that the association of the role REFERENT with knowledge and cognition (as well as with transfer-of-information predicates) is among the more salient associations that language learners encounter.

(7) *Terry Brown, 47, a truck driver, agreed; "with eight years in the White House," he said, "Bush ought to have a better grasp ON the details."* 9/27/88 NYT p. B8

(8) *I did get an idea ON the importance of consistency as far as reward and penalty are concerned.* 11/88 ETM journal

3.2.3 With/From/To

In this class, we believe that "mutual action verbs" such as *marry* and *divorce* routinely show a CO-THEME marker *with* being substituted for either *to* or *from*. Such predicates have a SECONDARY-MAPPING of PLURAL-THEME in the NLKB. Communication predicates are another class which allows a PLURAL-THEME and show alternation of GOAL and CO-THEME (Section 3.1.2).

(9) *Today Robin Givens said she won't ask for any money in her divorce WITH Mike Tyson.* 10/19/88 ATC

3.3 FALSE STARTS

The next set of examples suggests that the speaker has "retrieved" a preposition from a different ARGUMENT MAPPING for the verb or for a different argument than the one which is eventually produced. For example, *confused with* replaces *confused by* in (10), and *say to* replaces *say about* in (11). Such examples are more prevalent in oral language. Handling these examples is difficult since all sorts of contextual information -- linguistic and non-linguistic -- goes into detecting the error.

(10) *They didn't want to be confused WITH the facts.* 11/14/88 DRS

(11) *The memorial service was really well done. The rabbi did a good job. What do you say TO a kid who died like that?* 11/14/88

3.4 BLENDS

Here, a lexically or phrasally determined preposition is replaced by a preposition associated with a semantically similar lexical item. In (12) Quayle says he was *smitten about* Marilyn, possibly thinking of *crazy about*. In (13) he may be thinking of *on the subject/topic of*. The questioner in (14) may have *in support/favor of* in mind. In (15) Quayle may have meant *we learn by making mistakes*. In (16), the idiomatic phrase *in support of* is confused with the ARGUMENT MAPPING of the noun *support*, e.g., "he showed his support for the president".

(12) *I was very smitten ABOUT her... I saw a good thing and I responded rather quickly and she did too.* 10/20/88 WP, p. C8

(13) *ON the area of the federal budget deficit, ...* 10/5/88 Sen. Quayle in VP debate (& NYT 10/7/88 p. B6)

(14) *You made one of the most eloquent speeches IN behalf of contra aid.* 10/5/88 Questioner in VP debate (& NYT 10/7/88 p.B6)

(15) *We learn BY our mistakes.* 10/5/88 Sen. Quayle in VP debate (& NYT 10/7/88 p. B6)

(16) *We testified in support FOR medical leave.* 10/22/88 FFS

3.5 SEMANTIC FOR SYNTACTIC -- WITH/BY

In the majority of the following examples, the syntactically governed *by* marking passives is replaced by WITH. This alternation of *with* and *by* in passives has been attested for hundreds of years, and we hypothesize that English may be in the process of reinterpreting *by*, as well as replacing it with *with* in certain contexts. On the one hand, *by* is being reinterpreted as a marker of "archetypal" agents, i.e, those high on the scale of AGENTIVITY (i.e., speaker > human > animate > potent > non-animate, non-potent). On the other hand, a semantically appropriate marker is being

substituted for *by*.

We analyze the WITH in these examples either as the less agentive AGENT (namely the INSTRUMENT) in example (18), or the less agentive CO-THEME in example (17). The substitutions are semantically appropriate and the substitutes are semantically related to AGENT.

(17) *All of Russian life was accompanied WITH some kind of singing.* 8/5/88 ATC

(18) *Audiences here are especially enthused WITH Dukakis's description of the Reagan-Bush economic policies.* 11/5/88 ATC

4.0 THE COMPUTATIONAL IMPLEMENTATION

Of the five types of errors cited in Section 3, substitutions of semantic for semantic (Section 3.1), semantic for lexical (Section 3.2), and semantic for syntactic (Section 3.5) are the simplest to handle computationally.

4.1 SEMANTIC FOR SEMANTIC OR LEXICAL

The representation scheme described above (Section 2) facilitates handling the semantic for semantic and semantic for lexical substitutions.

Semantic for semantic substitutions are allowed if

(i) the predicate belongs to the communication class and the standard CO-THEME marker is replaced by a GOAL marker, or

(ii) the predicate incorporates a negative and GOAL is substituted for a standard SOURCE, or vice versa.

Semantic for lexical substitutions are allowed if

(iii) the non-standard preposition is a non-privative sister of the standard preposition (e.g., in the GOAL SUPERCLASS),

(iv) the non-standard preposition is the NLKB-inherited, "default" preposition for the predicate (e.g., REFERENT for predicates of cognition and knowledge), or

(v) in the NLKB the predicate allows a SECONDARY-MAPPING of PLURAL-THEME (e.g., marital predicates as in the *divorce with* example).

Handling the use of a non-standard preposition marking an argument crucially involves "type-checking", wherein the "type" of the noun phrase is checked, e.g. for membership in an NLKB class such as animate-creature, time, etc. Type-checking is also used to narrow the possible senses of the preposition in a prepositional phrase, as well as to prefer certain modifier attachments.

Prepositional phrases can have two relations to predicating expressions, i.e., a governed argument (PREP-ARG) or an ADJUNCT. During parsing, the system accesses the ARGUMENT MAPPING for the predicate; once the preposition is recognized as the standard marker of an argument, an ADJUNCT reading is disallowed. The rule for PREP-ARG is a separate rule in the grammar. When the preposition does not match the expected preposition, the system checks whether any of the above conditions (i-v) hold; if so, the parse is accepted, but is assigned a lower likelihood. If a parse of the PP as an ADJUNCT is also accepted, it will be preferred over the ill-formed PREP-ARG.

4.2 SEMANTIC FOR SYNTACTIC

The substitution of semantic marking for syntactic (WITH for BY) is easily handled: during semantic mapping *by* phrases in the ADJUNCTS are mapped to the role of the active subject, assuming that "type checking" allows that interpretation of the noun phrase. It is also possible for such a sentence to be ambiguous, e.g., "he was seated by the man". We treat *with* phrases similarly, except that ambiguity between CO-THEME and PASSIVE SUBJECT is not allowed, based on our observation that *with* for *by* is used for noun phrases low on the animacy scale. Thus, only the CO-THEME interpretation is valid if the noun phrase is animate.

4.3 FALSE STARTS AND BLENDS

False starts are more difficult, requiring an approach similar to that of case grammar. In these examples, the preposition is acceptable with the verb, but not to mark that particular argument. The

115

type of the argument marked with the "incorrect" preposition must be quite inconsistent with that sense of the predicate for the error even to be noticed, since the preposition is acceptable with some other sense. We are assessing the frequency of false starts in the various genres in which our system is being used, to determine whether we need to implement a strategy to handle these examples. We predict that future systems for understanding spoken language will need to accomodate this phenomenon.

We do not handle blends currently. They involve a form of analogy, i.e., *smitten* is like *mad*, syntactically, semantically, and even stylistically; they may shed some light on language storage and retrieval. Recognizing the similarity in order to allow a principled handling seems very difficult.

In addition, blends may provide evidence for a "top down" language production strategy, in which the argument structure is determined before the lexical items are chosen/inserted. Our data suggest that some people may be more prone to making this type of error than are others. Finally, blends are more frequent in genres in which people attempt to use a style that they do not command (e.g., student papers, radio talk shows).

5.0 DIRECTIONS FOR FUTURE WORK

In this paper we have described a frequent type of ill-formed input which NLP systems must handle, involving the use of non-standard prepositions to mark arguments. We presented a classification of these errors and described our algorithm for handling some of these error types. The importance of handling such non-standard input will increase as speech recognition becomes more reliable, because spoken input is less formal.

In the near term, planned enhancements include adjusting the weighting scheme to more accurately reflect the empirical data. A frequency-based model of preposition usage, based on a much larger and broader sampling of text will improve system handling of those errors.

REFERENCES

1. Aitchison, Jean. 1987. *Words in the Mind.* Blackwell, NY.

2. Blejer, Hatte and Sharon Flank. 1988. More Evidence for the Semantic Basis of Prepositional Case Marking, delivered December 28, 1988, Linguistic Society of America Annual Meeting, New Orleans.

3. Bresnan, Joan, ed. 1982. *The Mental Representation of Grammatical Relations.* MIT Press, Cambridge.

4. Carbonell, Jaime and Philip Hayes. 1983. Recovery Strategies for Parsing Extragrammatical Language. *American Journal of Computational Linguistics* 9(3-4): 123-146.

5. Chierchia, Gennaro, Barbara Partee, and Raymond Turner, eds. 1989. *Properties, Types and Meaning.* Kluwer, Dordrecht.

6. Chomsky, Noam. 1981. *Lectures on Government and Binding.* Foris, Dordrecht.

7. Croft, William. 1986. *Categories and Relations in Syntax: The Clause-Level Organization of Information.* Ph.D. Dissertation, Stanford University.

8. Dahlgren, Kathleen. 1988. *Naive Semantics for Natural Language Understanding.* Kluwer, Boston.

9. Dirven, Rene and Gunter Radden, eds. 1987. *Concepts of Case.* Gunter Narr, Tubingen.

10. Dowty, David. 1989. On the Semantic Content of the Notion of 'Thematic Role'. In Chierchia, et al. II:69-129.

11. Foley, William and Robert Van Valin Jr. 1984. *Functional Syntax and Universal Grammar.* Cambridge Univ. Press, Cambridge.

12. Gawron, Jean Mark. 1988. *Lexical Representations and the Semantics of Complementation*. Garland, NY.

13. Gazdar, Gerald, Ewan Klein, Geoffrey Pullum, and Ivan Sag. (GKPS) 1985. *Generalized Phrase Structure Grammar*. Harvard Univ. Press, Cambridge.

14. Gruber, Jeffrey. 1976. *Lexical Structures in Syntax and Semantics*. North-Holland, Amsterdam.

15. Haiman, John. 1985. *Natural Syntax: Iconicity and Erosion*. Cambridge University Press, Cambridge.

16. Hirst, Graeme. 1987. *Semantic Interpretation and the Resolution of Ambiguity*. Cambridge University Press, Cambridge.

17. Ikegami, Yoshihiko. 1987. 'Source' vs. 'Goal': a Case of Linguistic Dissymetry, in Dirven and Radden 122-146.

18. Jackendoff, Ray. 1983. *Semantics and Cognition*. MIT Press, Cambridge.

19. Jensen, Karen, George Heidorn, Lance Miller and Yael Ravin. 1983. Parse Fitting and Prose Fixing: Getting a Hold on Ill-formedness. *American Journal of Computational Linguistics* 9(3-4): 147-160.

20. Menyuk, Paula. 1969. *Sentences Children Use*. MIT Press, Cambridge.

21. Miller, Glenn and Philip Johnson-Laird. 1976. *Language and Perception*. Harvard University Press, Cambridge.

22. Ostler, Nicholas. 1980. A Theory of Case Linking and Agreement. Indiana University Linguistics Club.

23. Pinker, Steven. 1982. A Theory of the Acquisition of Lexical Interpretive Grammars, in Bresnan 655-726.

24. Shopen, Timothy, ed. 1985. *Language Typology and Syntactic Description*. Cambridge University Press, Cambridge.

25. Somers, H. L. 1987. *Valency and Case in Computational Linguistics*. Edinburgh University Press, Edinburgh.

26. Talmy, Leonard. 1985. Lexicalization Patterns: Semantic Structure in Lexical Forms. In Shopen III:57-149.

27. Tomita, Masuru. 1986. *Efficient Parsing for Natural Language*. Kluwer, Boston.

28. Vendler, Zeno. 1967. *Linguistics in Philosophy*. Cornell University Press, Ithaca.

29. Weischedel, Ralph and Norman Sondheimer. 1983. Meta-rules as a Basis for Processing Ill-Formed Input. *American Journal of Computational Linguistics* 9(3-4):161-177.

APPENDIX A. DATA SOURCES

ATC: National Public Radio news program, "All Things Considered"
ME: National Public Radio news program, "Morning Edition"
WE: National Public Radio news program, "Weekend Edition"
MCS: WAMU radio, Washington D.C., interview program, "The Mike Cuthbert Show"
DRS: WAMU radio, Washington D.C., interview program, "Diane Rehm Show"
FFS: WAMU radio, Washington D.C., interview program, "Fred Fiske Saturday"
AIH: Canadian Broadcasting Company radio news program, "As It Happens"
NYT: The New York Times
WP: The Washington Post
ETM: Student journal for "Effective Teaching Methods," a junior undergraduate course

117