

SENTENCE FRAGMENTS REGULAR STRUCTURES

Marcia C. Linebarger, Deborah A. Dahl, Lynette Hirschman, Rebecca J. Passonneau

Paoli Research Center
Unisys Corporation
P.O. Box 517
Paoli, PA

ABSTRACT

This paper describes an analysis of telegraphic fragments as regular structures (not errors) handled by minimal extensions to a system designed for processing the standard language. The modular approach which has been implemented in the Unisys natural language processing system PUNDIT is based on a division of labor in which syntax regulates the occurrence and distribution of elided elements, and semantics and pragmatics use the system's standard mechanisms to interpret them.

1. INTRODUCTION

In this paper we discuss the syntactic, semantic, and pragmatic analysis of fragmentary sentences in English. Our central claim is that these sentences, which have often been classified in the literature with truly erroneous input such as misspellings (see, for example, the work discussed in [Kwasny1980, Thompson1980, Kwasny1981, Sondheimer1983, Eastman1981, Jensen1983]), are regular structures which can be processed by adding a small number of rules to the grammar and other components of the system. The syntactic regularity of fragment structures has been demonstrated elsewhere, notably in [Marsh1983, Hirschman1983]; we will focus here upon the regularity of these structures across all levels of linguistic representation. Because the syntactic component regularises these structures into a form almost indistinguishable from full

assertions, the semantic and pragmatic components are able to interpret them with few or no extensions to existing mechanisms. This process of incremental regularisation of fragment structures is possible only within a linguistically modular system. Furthermore, we claim that although fragments may occur more frequently in specialised sublanguages than in the standard grammar, they do not provide evidence that sublanguages are based on grammatical principles fundamentally different from those underlying standard languages, as claimed by [Fitzpatrick1986], for example.

This paper is divided into five sections. The introductory section defines fragments and describes the scope of our work. In the second section, we consider certain properties of sentence fragments which motivate a modular approach. The third section describes our implementation of processing for fragments, to which each component of the system makes a distinct contribution. The fourth section describes the temporal analysis of fragments. Finally, the fifth section discusses the status of sublanguages characterised by these telegraphic constructions.

We define fragments as regular structures which are distinguished from full assertions by a missing element or elements which are normally syntactically obligatory. We distinguish them from errors on the basis of their regularity and consistency of interpretation, and because they appear to be generated intentionally. We are not denying the existence of true errors, nor that processing sentences containing true errors may require sophisticated techniques and deep reasoning. Rather, we are saying that fragments are distinct from errors, and can be handled in a quite general fashion, with minimal extensions to normal processing. Because we base the definition of *fragment* on the absence of a syntactically

¹This work has been supported in part by DARPA under contract N00014-85-C-0012, administered by the Office of Naval Research; by National Science Foundation contract DCR-85-02205; and by Independent R&D funding from System Development Corporation, now part of Unisys Corporation. Approved for public release, distribution unlimited.

obligatory element, noun phrases without articles are not considered to be fragmentary, since this omission is conditioned heavily by semantic factors such as the mass vs. count distinction. However, we have implemented a pragmatically based treatment of noun phrases without determiners, which is briefly discussed in Section 3.

Fragments, then, are defined here as elisions. We describe below the way in which these omissions are detected and subsequently 'filled in' by different modules of the system.

The problem of processing fragmentary sentences has arisen in the context of a large-scale natural language processing research project conducted at UNISYS over the past five years [Palmer1986, Hirschman1986, Dowding1987, Dahl1987]. We have developed a portable, broad-coverage text-processing system, PUNDIT.¹ Our initial applications have involved various message types, including: field engineering reports for maintenance of computers; Navy maintenance reports (Casualty Reports, or CASREPs) for starting air compressors; Navy intelligence reports (RAINFORMS); trouble and failure reports (TFRs) from Navy Vessels; and recently we have examined several medical domains (radiology reports, comments fields from a DNA sequence database). At least half the sentences in these corpora are fragments; Table 1 below gives a summary of the fragment content of three domains, showing the percent of centers which are classified as fragments. (Centers comprise all sentence types: assertions, questions, fragments, and so forth.)

Table 1. Fragments in three domains.

	Total centers	Percent fragments
CASREPS	153	53%
RAINFORM	41	78%
TFR	35	51%

The PUNDIT system is highly modular: it consists of a syntactic component, based on string grammar and restriction grammar [Sager1981, Hirschman1985]; a semantic component, based on inference-driven mapping, which decomposes predicating expressions into predicates and thematic roles [Palmer1983, Palmer1985]; and a pragmatics component which processes both referring expressions [Dahl1986], and temporal expressions [Passonneau1987, Passonneau1988].

¹ Prolog UNDERstanding of Integrated Text.

2. DIVISION OF LABOR AMONG SYNTAX, SEMANTICS, AND PRAGMATICS

We argue here that sentence fragments provide a strong case for linguistically modular systems such as PUNDIT, because such elisions have distinct consequences at different levels of linguistic description. Our approach to fragments can be summarised by saying that syntax detects 'holes' in surface structure and creates dummy elements as placeholders for the missing elements; semantics and pragmatics interpret these placeholders at the appropriate point in sentence processing, utilising the same mechanisms for fragments as for full assertions.

Syntax regulates the holes. Fragment elisions cannot be accounted for in purely semantic/pragmatic terms. This is evidenced by the fact that there are syntactic restrictions on omissions; the acceptability of a sentence fragment hinges on grammatical factors rather than, e.g., how readily the elided material can be inferred from context. For example, the discourse *Old house too small. *New one will be larger than _ was* (where the elided object of *than* is understood to be *old house*) is ill-formed, whereas a comparable discourse *First repairman ordered new air conditioner. Second repairman will install _* (where the elided object of *install* is understood to be *air conditioner*) is acceptable. In both cases above, the referent of the elided element is available from context, and yet only the second ellipsis sounds well-formed. Thus an appreciation of where such ellipses may occur is part of the linguistic knowledge of speakers of English and not simply a function of the contextual salience of elided elements. Since these restrictions concern structure rather than content, they would be difficult or impossible to state in a system such as a 'pure' semantic grammar which only recognised such omissions at the level of semantic/pragmatic representation.

Furthermore, it matters to semantics and pragmatics HOW an argument is omitted. The syntactic component must tell semantics whether a verb argument is missing because the verb is used intransitively (as in *The tiger was eating*, where the patient argument is not specified) or because of a fragment ellipsis (as in *Eaten by a tiger*, where the patient argument is missing because the subject of a passive sentence has been elided). Only in the latter case does the missing argument of *eat* function as an

antecedent subsequently in the discourse: compare *Eaten by a tiger. Had screamed bloody murder right before the attack* (where the victim and the screamer are the same) vs. *The tiger was eating. Had screamed bloody murder right before the attack* (where it is difficult or impossible to get the reading in which the victim and the screamer are the same).

Semantics and pragmatics fill the holes. In PUNDIT's treatment of fragments, each component contributes exactly what is appropriate to the specification of elided elements. Thus the syntax does not attempt to 'fill in' the holes that it discovers, unless that information is completely predictable given the structure at hand. Instead, it creates a dummy element. If the missing element is an elided subject, then the dummy element created by the syntactic component is assigned a referent by the pragmatics component. This referent is then assigned a thematic role by the semantics component like any other referent, and is subject to any selectional restrictions associated with the thematic role assigned to it. If the missing element is a verb, it is specified in either the syntactic or the semantic component, depending upon the fragment type.

3. PROCESSING FRAGMENTS IN PUNDIT

Although the initial PUNDIT system was designed to handle full, as opposed to fragmentary, sentences, one of the interesting results of our work is that it has required only very minor changes to the system to handle the basic fragment types introduced below. These included the additions of: 6 fragment BNF definitions to the grammar (a 5% increase in grammar size) and 7 context-sensitive restrictions (a 12% increase in the number of restrictions); one semantic rule for the interpretation of the dummy element inserted for missing verbs; a minor modification to the reference resolution mechanism to treat elided noun phrases like pronouns; and a small addition to the temporal processing mechanism to handle tenseless fragments. The small number of changes to the semantic and pragmatic components reflects the fact that these components are not 'aware' that they are interpreting fragmentary structures, because the regularisation performed by the syntactic component renders them structurally indistinguishable from full assertions.

Fragments present parsing problems because the ellipsis creates degenerate structures. For example, a sequence such as *chest negative* can be analysed as a 'sero-copula' fragment meaning *the chest X-ray is negative*. or a noun compound like *the negative of the chest*. This is compounded by the lack of derivational and inflectional morphology in English, so that in many cases it may not be possible to distinguish a noun from a verb (*repair parts*) or a past tense from a past participle (*decreased medication*). Adding fragment definitions to the grammar (especially if determiner omission is also allowed) results in an explosion of ambiguity. This problem has been noted and discussed by Kwasny and Sondheimer [Kwasny1981]. Their solution to the problem is to suggest special relaxation techniques for the analysis of fragments. However, in keeping with our thesis that fragments are normal constructions, we have chosen the alternative of constraining the explosion of parses in two ways. The first is the addition of a control structure to implement a limited form of preference via 'unbacktrackable' or (xor). This binary operator tries its second argument only if its first argument does not lead to a parse. In the grammar, this is used to prefer "the most structured" alternative. That is, full assertions are preferred over fragments -- if an assertion or other non-fragment parse is obtained, the parser does not try for a fragment parse.

The second mechanism that helps to control generation of incorrect parses is selection. PUNDIT applies surface selectional constraints incrementally, as the parse is built up [Lang1988]. For example, the phrase *air compressor* would NOT be allowed as a serocopula because the construction *air is compressor* would fail selection.²

3.1. Fragment Types

The fragment types currently treated in PUNDIT include the following:

Zerocopula: a subject followed by a predicate, differing from a full clause only in the absence of a verb, as in *Impellor blade tip erosion evident*;

Two (tensed verb + object): a sentence missing its subject, as in *Believe the coupling from diesel to sac lube oil pump to be sheared*;

² Similarly, the assertion parse for the title of this paper would fail selection (sentences don't fragment structures), permitting the serocopula fragment parse.

Nstg_frag: an isolated noun phrase (noun-string fragment), as in *Loss of oil pump pressure*.

Objbe_frag (object-of-be fragment): an isolated complement appropriate to the main verb *be*, as in *Unable to consistently start nr 1b gas turbine*;

Predicate: an isolated complement appropriate to auxiliary *be*, as in *Believed due to worn bushings*, where the full sentence counterpart is *Failure is believed (to be) due to worn bushings*;³

Obj_gap_fragment: a center (assertion, question, or other fragment structure) missing an obligatory noun phrase object, as in *Field engineer will replace _*

Note that we do not address here the processing of *response fragments* which occur in interactive discourse, typically as responses to questions.

The relative frequency of these six fragment types (expressed as a percentage of the total fragment content of each corpus) is summarised below.⁴

Table 2. Breakdown of fragments by type.

	CASREPS	RAINFORM	TFR
TVO	17.5%	40.6%	61%
ZC	52.5%	50%	16.6%
NF	25%	6.2%	16.6%
OBJBE	3.7%	0%	5.5%
PRED	1.2%	3.1%	0%
OBJ_GAP	0%	3.1%	0%

The processing of these basic fragment types can be summarised briefly as follows: a detailed surface parse tree is provided which represents the overt lexical content in its surface order. At this level, fragments bear very little resemblance to full assertions. But at the level of the *Intermediate Syntactic Representation* (ISR),

³ It is interesting to note that at least some of these types of fragments resemble non-fragmentary structures in other languages. *tvo* fragments, for example, can be compared to zero-subject sentences in Japanese, zero-copulas resemble copular sentences in Arabic and Russian, and structures similar to predicate can be found in Cantonese (our thanks to K. Fu for the Cantonese data). This being the case, it is not surprising that analogous sentences in English can be processed without resorting to extragrammatical mechanisms.

⁴ ZC = zero-copula; NF = *nstg_fragment*; PRED = predicate; OBJBE = *objbe_frag*; OBJ_GAP = *obj_gap_fragment*.

which is a regularised representation of syntactic structure [Dahl1987..], fragments are regularised to parallel full assertions by the use of dummy elements standing in for the missing subject or verb. The CONTENT of these dummy elements, however, is left unspecified in most cases, to be filled in by the semantic or pragmatic components of the system.

Tvo. We consider first the *tvo*, a subject-less tensed clause such as *Operates normally*. This is parsed as a sequence of tensed verb and object: no subject is inferred at the level of surface structure. In the ISR, the missing subject is filled in by the dummy element elided. At the level of the ISR, then, the fragment *operates normally* differs from a full assertion such as *It operates normally* only by virtue of the element elided in place of an overt pronoun. The element elided is assigned a referent which subsequently fills a thematic role, exactly as if it were a pronoun; thus these two sentences get the same treatment from semantics and reference resolution [Dahl1986, Palmer1986].

Elided subjects in the domains we have looked at often refer to the writer of the report, so one strategy for interpreting them might be simply to assume that the filler of the elided subject is the writer of the report. This simple strategy is not sufficient in all cases. For example, in the CASREPS corpus we observe sequences such as the following, where the filler of the elided subject is provided by the previous sentence, and is clearly not the writer of the report.

- (1) Problem appears to be caused by one or more of two hydraulic valves. Requires disassembly and investigation.
- (2) Sac lube oil pressure decreases below alarm point approximately seven minutes after engagement. Believed due to worn bushings.

Thus, it is necessary to be able to treat elided subjects as pronouns in order to handle these sentences.

The effect of an elided subject on subsequent focusing is the same as that of an overt pronoun. We demonstrated in section 2 that elided subjects, but not semantically implicit arguments, are expected foci (or forward-looking centers [Gross1986]) for later sentences.

The basic assumption underlying this treatment is that the pragmatic analysis for elided subjects should be as similar to that of pronouns as possible. One piece of supporting evidence for this assumption is that in many languages, such as Japanese [Gundel1980, Hinds1983, Kameyama1985] the functional equivalent of unstressed pronouns in English is a sero, or elided noun phrase.⁵ If seros in other languages can correspond to unstressed pronouns in English, then we hypothesise that seros in a sublanguage of English can correspond functionally to pronouns in standard English. In addition, since processing of pronouns is independently motivated, it is *a priori* simpler to try to fit elision into the pronominal paradigm, if possible, than to create an entirely separate component for handling elision. Under this hypothesis, then, two fragments represent simply a realisation of a grammatical strategy that is generally available to languages of the world.⁶

Zerocopula. For a zerocopula (e.g., *Disk bad*), the surface parse tree rather than the ISR inserts a dummy verb, in order to enforce subcategorisation constraints on the object. And in the ISR, this null verb is 'filled in' as the verb *be*. It is possible to fill in the verb at this level because no further semantic or pragmatic information is required in order to determine its content.⁷ Hence the representation for *Disk bad* is nearly indistinguishable from that assigned to the corresponding *Disk is bad*; the only difference is in the absence of tense from the former. If the null verb represents auxiliary *be*, then, like an overt auxiliary, it does not appear in the regularised form. *Sac failing* thus receives a regularisation with *fail* as the main verb. Thus the null verb inserted in the syntax is treated in the ISR in a fashion exactly parallel to the treatment of overt

⁵ Stressed pronouns in English correspond to overt pronouns in languages like Japanese, as discussed in [Gundel1980, Gundel1981], and [Dahl1982].

⁶ An interesting hypothesis, discussed by Gundel and Kameyama, is that the more topic prominent a language is, the more likely it is to have sero-NP's. Perhaps the fact that sublanguage messages are characterized by rigid, contextually supplied, topics contributes to the availability of the two fragment type in English.

⁷ In some restricted subdomains, however, other verbs may be omitted: for example, in certain radiology reports an omitted verb may be interpreted as *show* rather than *be*. Hence we find *Chest films 1/10 little change*, paraphrasable as *Chest films show little change*.

occurrences of *be*.

Nstg_frag. The syntactic parse tree for this fragment type contains no empty elements; it is a regular noun phrase, labeled as an *nstg_frag*. The ISR transforms it into a VSO sequence. This is done by treating it as the subject of an element *empty_verb*; in the semantic component, the subject of *empty_verb* is treated as the sole argument of a predicate existential(λ). As a result, the *nstg_frag* *Failure of sac* and a synonymous assertion such as *Failure of sac occurred* are eventually mapped onto similar final representations by virtue of the temporal semantics of *empty_verb* and of the head of the noun phrase.

Objbe_frag and predicate. These are isolated complements; the same devices described above are utilised in their processing. The surface parse tree of these fragment types contains no empty elements; as with zerocopula, the untensed verb *be* is inserted into the ISR; as with two, the dummy subject elided is also inserted in the ISR, to be filled in by reference resolution. Thus the simple adjective *Inoperative* will receive an ISR quite similar to that of *He/she/it is inoperative*.

Obj_gap_fragment. The final fragment type to be considered here is the elided noun phrase object. Such object elisions occur more widely in English in the context of instructions, as in *Handle _ with care*. Cookbooks are especially well-known repositories of elided objects, presumably because they are filled with instructions. Object elision also occurs in telegrammatic sublanguages generally, as in *Took _ under fire with missiles* from the Navy sighting messages. If these omissions occurred only in direct object position following the verb, one might argue for a lexical treatment; that is, such omissions could be treated as a lexical process of intransitivization rather than by explicitly representing gaps in the syntactic structure. However, noun phrase objects of prepositions may also be omitted, as in *Fragile. Do not tamper with _*. Thus we have chosen to represent such elisions with an explicit surface structure gap. This gap is permitted in most contexts where *nstgo* (noun phrase object) is found: as a direct object of the verb and as an object of a preposition.⁸ In PUNDIT, elided objects are

⁸ Note, however, that there are some restrictions on the occurrence of these elements. They seem not to occur in

permitted only in a fragment type called *obj_gap_fragment*, which, like other fragment types, may be attempted only if an assertion parse has failed. Thus a sentence such as *Pressure was decreasing rapidly* will never be analysed as containing an elided object, because there is a semantically acceptable assertion parse. In contrast, *John was decreasing gradually* will receive an elided object analysis, paraphrasable as *John was decreasing IT gradually*, because *John* is not an acceptable subject of intransitive *decrease*; only pressure or some equally measurable entity may be said to decrease. This selectional failure of the assertion parse permits the elided object analysis.

Our working hypothesis for determining the reference of object gaps is that they are, just like subject gaps, appropriately treated as pronouns. However, we have not as yet seen extensive data relevant to this hypothesis, and it remains subject to further testing.

These, then, are the fragment types currently implemented in PUNDIT. As mentioned above, we do not consider noun phrases without determiners to be fragments, because it is not clear that the missing element is *syntactically* obligatory. The interpretation of these noun phrases is treated as a pragmatic problem. In the style of speech characteristic of the CASREPs, determiners are nearly always omitted. Their function must therefore be replaced by other mechanisms. One possible approach to this problem would be to have the system try to determine what the determiner would have been, had there been one, insert it, and then resume processing as if the determiner had been there all along. This approach was taken by [Marsh1981]. However, it was rejected here for two reasons. The first is that it was judged to be more error-prone than simply equipping the reference resolution component with the ability to handle noun phrases without determiners directly.⁹ The second reason

predicative objects, in double dative constructions, and, perhaps, in sentence adjuncts rather than arguments of the verb. (Thus compare *Patient very ill. Do not operate on ... with Operating room closed on Sunday. Do not perform surgery on...*) One possibility is that these expressions can occur only where a definite pronoun would also be acceptable. In general, object gaps seem most acceptable where they represent an argument of a verb, either as direct object or as object of a preposition selected for by a verb.

⁹ This ability would be required in any case, should the system be extended to process languages which do not have

for not selecting this approach is that it would eliminate the distinction between noun phrases which originally had a determiner and those which did not. At some point in the development of the system it may become necessary to use this information.

The basic approach currently taken is to assume that the noun phrase is definite, that is, it triggers a search through the discourse context for a previously mentioned referent. If the search succeeds, the noun phrase is assumed to refer to that entity. If the search fails, a new discourse entity is created.

In summary, then, these fragment types are parsed 'as is' at the surface level; dummy elements are inserted into the ISR to bring fragments into close parallelism with full assertions. Because of the resulting structural similarity between these two sentence types, the semantic and pragmatic components can apply exactly the same interpretive processes to both fragments and assertions, using pre-existing mechanisms to 'fill in' the holes detected by syntax.

4. TEMPORAL ANALYSIS OF FRAGMENTS

Temporal processing of fragmentary sentences further supports the efficacy of a modular approach to the analysis of these strings.¹⁰ In PUNDIT's current message domains, a single assumption leads to assignment of present or past tense in untensed fragments, depending on the aspectual properties of the fragment.¹¹ This assumption is that the messages report on actual situations which are of present relevance. Consequently, the default tense assignment is *present* unless this prevents assigning an actual time.¹²

For sentences having progressive grammatical aspect or stative lexical aspect, the assignment of present tense always permits interpreting

articles.

¹⁰For a discussion of the temporal component, cf. [Passonneau1987, Passonneau1988].

¹¹Since the two fragment is tensed, its input to the time component is indistinguishable from that of a full sentence.

¹²Pundit does not currently take full advantage of modifier information that could indicate whether a situation has real time associated with it (e.g., *potential sac failure*), or whether a situation is past or present (e.g., *sac failure yesterday; pump now operating normally*).

a situation as having an actual time [Passonneau1987]. Thus, a present tense reading is always assigned to an untensed progressive fragment, such as *pressure decreasing*; or an untensed serocopula with a non-participial complement, such as *pump inoperative*.

A non-progressive serocopula fragment containing a cognitive state verb, as in *failure believed due to worn bushings*, is assigned a present tense reading. However, if the lexical verb has non-stative aspect,¹³ e.g., *tests conducted* (process) or *new sac received* (transition event) then assignment of present tense conflicts with the assumption that the mentioned situation has occurred or is occurring. The simple present tense form of verbs in this class is given a habitual or iterative reading. That is, the corresponding full sentences in the present, *tests are conducted* and *new sac is received*, are interpreted as referring to types of situations that tend to occur, rather than to situations that have occurred. In order to permit actual temporal reference, these fragments are assigned a past tense reading.

nstg_frag represents another case where present tense may conflict with lexical aspect. If an *nstg_frag* refers to a non-stative situation, the situation is interpreted as having an actual past time. This can be the case if the head of the noun phrase is a nominalisation, and is derived from a verb in the process or transition event aspectual class. Thus, *investigation of problem* would be interpreted as an actual process which took place prior to the report time, and similarly, *sac failure* would be interpreted as a past transition event. On the other hand, an *nstg_frag* which refers to a stative situation, as in *inoperative pump*, is assigned present tense.

5. RELATION OF FRAGMENTS TO THE LARGER GRAMMAR

An important finding which has emerged from the investigation of sentence fragments in a variety of sublanguage domains is that the linguistic properties of these constructions are largely domain-independent. Assuming that these sentence fragments remain constant across different sublanguages, what is their relationship to the language at large? As indicated above, we

believe that fragments should not be regarded as ERRORS, a position taken also by [Lehrberger1982, Marsh1983], and others. Fragments do occur with disproportionate frequency in some domains, such as field reports of mechanical failure or newspaper headlines. However, despite this frequency variation, it appears that the parser's preferences remain constant across domains. Therefore, even in telegraphic domains the preference is for a full assertion parse, if one is available. As discussed above, we have enforced this preference by means of the xor ('unbacktrackable' or) connective. Thus despite the greater frequency of fragments we do not require either a grammar or a preference structure different from that of standard English in order to apply the stable system grammar to these telegraphic messages.

Others have argued against this view of the relationship between sublanguages and the language at large. For example, Fitzpatrick et al. [Fitzpatrick1986] propose that fragments are subject to a constraint quite unlike any found in English generally. Their Transitivity Constraint (TC) requires that if a verb occurs as a transitive in a sublanguage with fragmentary messages, then it may not also occur in an intransitive form, even if the verb is ambiguous in the language at large. This constraint, they argue, provides evidence that sublanguage grammars have "a life of their own", since there is no such principle governing standard languages. The TC would also cut down on ambiguities arising out of object deletion, since a verb would be permitted to occur transitively or intransitively in a given subdomain, but not both.

As the authors recognise, this hypothesis runs into difficulty in the face of verbs such as *resume* (we find both *Sac resumed normal operation* and *Noise has resumed*), since *resume* occurs both transitively and intransitively in these cases. For these cases, the authors are forced to appeal to a problematic analysis of *resume* as syntactically transitive in both cases; they analyse *The noise has resumed*, for example, as deriving from a structure of the form (*Someone/something resumed the noise*); that is, it is analyzed as underlyingly transitive. Other transitivity alternations which present potential counter-examples are treated as syntactic gapping processes. In fact, with these two mechanisms available, it is not clear what COULD provide a counter-example to

¹³ Mourelatos' class of occurrences [Mourelatos1981].

the TC. The effect of all this insulation is to render the Transitivity Constraint vacuous. If all transitive/intransitive alternations can be treated as underlyingly transitive, then of course there will be no counter-examples to the transitivity constraint. Therefore we see no evidence that sublanguage grammars are subject to additional constraints of this nature.

In summary, this supports the view that fragmentary constructions in English are regular, grammatically constrained ellipses differing minimally from the standard language, rather than ill-formed, unpredictable sublanguage exotica. Within a modular system such as PUNDIT this regularity can be captured with the limited augmentations of the grammar described above.

ACKNOWLEDGMENTS

The system described in this paper has been developed by the entire natural language group at Unisys. In particular, we wish to acknowledge the contributions of John Dowding, who developed the ISR in conjunction with Deborah Dahl; and Martha Palmer's work on the semantics component. The ISR is based upon the work of Mark Gawron.

We thank Tim Finin and Martha Palmer as well as the anonymous reviewers for useful comments on an earlier version of this paper.

References

- [Dahl1987] Deborah A. Dahl, John Dowding, Lynette Hirschman, Francois Lang, Marcia Linebarger, Martha Palmer, Rebecca Passonneau, and Leslie Riley, Integrating Syntax, Semantics, and Discourse: DARPA Natural Language Understanding Program, R&D Status Report, Paoli Research Center, Unisys Defense Systems, May 14, 1987.
- [Dahl1986] Deborah A. Dahl, Focusing and Reference Resolution in PUNDIT, Presented at AAI, Philadelphia, PA, 1986.
- [Dahl1982] Deborah A. Dahl and Jeanette K. Gundel, Identifying Referents for two kinds of Pronouns. In *Minnesota Working Papers in Linguistics and Philosophy of Language*, Kathleen Houlihan (ed.), 1982, pp. 10-29.
- [Dahl1987] Deborah A. Dahl, Martha S. Palmer, and Rebecca J. Passonneau, Nominalizations in PUNDIT, Proceedings of the 25th Annual Meeting of the ACL, Stanford, CA, July, 1987.
- [Dowding1987] John Dowding and Lynette Hirschman, Dynamic Translation for Rule Pruning in Restriction Grammar. In *Proc. of the 2nd International Workshop on Natural Language Understanding and Logic Programming*, Vancouver, B.C., Canada, 1987.
- [Eastman1981] C.M Eastman and D.S. McLean, On the Need for Parsing Ill-Formed Input. *American Journal of Computational Linguistics* 7, 1981.
- [Fitspatrick1986] E. Fitspatrick, J. Bachenko, and D. Hindle, The Status of Telegraphic Sublanguages. In *Analyzing language in Restricted Domains*, R. Grishman and R. Kittredge (ed.), Lawrence Erlbaum Associates, Hillsdale, NY, 1986.
- [Gross1986] Barbara J. Gross, Aravind K. Joshi, and Scott Weinstein, Towards a Computational Theory of Discourse Interpretation, Mss., 1986.
- [Gundel1981] Jeanette K. Gundel and Deborah A. Dahl, The Comprehension of Focussed and Non-Focussed Pronouns, Proceedings of the Third Annual Meeting of the Cognitive Science Society, Berkeley, CA, August, 1981.

- [Gundel1980]
Jeanette K. Gundel, Zero-NP Anaphora in Russian. *Chicago Linguistic Society Parasession on Pronouns and Anaphora*, 1980.
- [Hinds1983]
John Hinds, Topic Continuity in Japanese. In *Topic Continuity in Discourse*, T. Givon (ed.), John Benjamins Publishing Company, Philadelphia, 1983.
- [Hirschman1983]
Lynette Hirschman and Naomi Sager, Automatic Information Formatting of a Medical Sublanguage. In *Sublanguage: Studies of Language in Restricted Semantic Domains*, R. Kittredge and J. Lehrberger (ed.), Series of Foundations of Communications, Walter de Gruyter, Berlin, 1983, pp. 27-80.
- [Hirschman1986]
L. Hirschman, Conjunction in Meta-Restriction Grammar. *J. of Logic Programming*(4), 1986, pp. 299-328.
- [Hirschman1985]
L. Hirschman and K. Puder, Restriction Grammar: A Prolog Implementation. In *Logic Programming and its Applications*, D.H.D. Warren and M. VanCaneghem (ed.), 1985.
- [Jensen1983]
K. Jensen, G.E. Heidorn, L.A. Miller, and Y. Ravin, Parse Fitting and Prose Fixing: Getting a Hold on Ill-Formedness. *American Journal of Computational Linguistics* 9, 1983.
- [Kameyama1985]
Megumi Kameyama, Zero Anaphora: The Case of Japanese, Ph.D. thesis, Stanford University, 1985.
- [Kwasny1981]
S.C. Kwasny and N.K. Sondheimer, Relaxation Techniques for Parsing Ill-Formed Input. *Am J. of Computational Linguistics* 7, 1981, pp. 99-108.
- [Kwasny1980]
Stan C. Kwasny, *Treatment of Ungrammatical and Extra-Grammatical Phenomena in Natural Language Understanding Systems*. Indiana University Linguistics Club, 1980.
- [Lang1988]
Francois Lang and Lynette Hirschman, Improved Portability and Parsing Through Interactive Acquisition of Semantic Information, Proc. of the Second Conference on Applied Natural Language Processing, Austin, TX, February, 1988.
- [Lehrberger1982]
J. Lehrberger, Automatic Translation and the Concept of Sublanguage. In *Sublanguage: Studies of Language in Restricted Semantic Domains*, R. Kittredge and J. Lehrberger (ed.), de Gruyter, Berlin, 1982.
- [Marsh1983]
Elaine Marsh, Utilising Domain-Specific Information for Processing Compact Text. In *Proceedings of the Conference on Applied Natural Language Processing*, Santa Monica, CA, February, 1983, pp. 99-103.
- [Marsh1981]
Elaine Marsh, A or THE? Reconstruction of Omitted Articles in Medical Notes, Mss., 1981.
- [Mourelatos1981]
Alexander P. D. Mourelatos, Events, Processes and States. In *Syntax and Semantics: Tense and Aspect*, P. J. Tedeschi and A. Zaenen (ed.), Academic Press, New York, 1981, pp. 191-212.
- [Palmer1983]
M. Palmer, Inference Driven Semantic Analysis. In *Proceedings of the National Conference on Artificial Intelligence (AAAI-83)*, Washington, D.C., 1983.

[Palmer1986]

Martha S. Palmer, Deborah A. Dahl, Rebecca J. [Passonneau] Schiffman, Lynette Hirschman, Marcia Linebarger, and John Dowding, *Recovering Implicit Information*, Presented at the 24th Annual Meeting of the Association for Computational Linguistics, Columbia University, New York, August 1986.

[Palmer1985]

Martha S. Palmer, *Driving Semantics for a Limited Domain*, Ph.D. thesis, University of Edinburgh, 1985.

[Passonneau1988]

Rebecca J. Passonneau, *A Computational Model of the Semantics of Tense and Aspect*. *Computational Linguistics*, 1988.

[Passonneau1987]

Rebecca J. Passonneau, *Situations and Intervals*, Presented at the 25th Annual Meeting of the Association for Computational Linguistics, Stanford University, California, July 1987.

[Sager1981]

N. Sager, *Natural Language Information Processing: A Computer Grammar of English and Its Applications*. Addison-Wesley, Reading, Mass., 1981.

[Sondheimer1983]

N. K. Sondheimer and R. M. Weischedel, *Meta-rules as a Basis for Processing Ill-Formed Input*. *American Journal of Computational Linguistics* 9(3-4), 1983.

[Thompson1980]

Bosena H. Thompson, *Linguistic Analysis of Natural Language Communication with Computers*. In *Proceedings of the 8th International Conference on Computational Linguistics*, Tokyo, 1980.