CONSTRAINTS ON NOUN PHRASE CONJUNCTION: A DOMAIN-INDEPENDENT MECHANISM

Lynette Hirschman

Research and Development, Federal and Special Systems Group, Burroughs Corporation, Paoli, Pennsylvania, USA

Coordinate conjunction is any important device for expressing complex information. To provide a habitable language processing system, most applications require some treatment of conjunction. When a rich conjunction mechanism is introduced to account for the wide variety of conjoinings found in continuous text (as opposed to queries), it also becomes crucial to restrict the parses obtained for the conjoinings, since conjunction is a major source of ambiguity in parsing. For example, the sequence of syntactic classes:

NOUN PREPOSITION NOUN and NOUN

has two possible parses:

NOUN	PREPOSITION	(NOUN	and	NOUN)
swelling	of	(hands	and	feet)

(NOUN	PREPOSITION	NOUN)	and	NOUN
(swelling	of	hands)	and	fever

The choice of the correct parse depends on applying domain-specific constraints to determine which words "go together" best. The concept of conjunctional compatibility depends on semantic parallelism between the conjoined elements; thus, in the first example, hands and feet are more semantically parallel than swelling and feet. Although it is possible to write special rules to define allowable conjoinings for a particular domain of application, this would mean that the

system would not be portable: the conjunction constraints would have to be rewritten for each new domain. It is, however, possible to formulate a general domain-independent principle to capture conjunction constraints. This paper will describe such a principle, implemented as "conjunction restrictions" within the general treatment of conjunction in the natural language processing system of the Linguistic String Project parser. The illustrative examples are drawn from actual text occurrences in hospital discharge summaries, processed using the mechanisms described below, as well as from some building code specifications, also processed with the LSP system.

> The first rule of conjunctional compatibility is simply:
>
> Two conjoined nouns must belong to the same (domain-specific) semantic class.

This rule will produce the correct parse for the two previous examples, swelling of hands and feet and swelling of hands and fever. Hands and feet are both BODY-PART words in the first example, while swelling is not; in the second example, swelling and fever are both SIGN-SYMPTOM words.

In order to determine conjunctional compatibility, the noun conjuncts are compared pairwise; each conjunct is initially assigned a list consisting of its semantic class(es). (Most words belong to just one semantic class; however, certain lexically ambiguous words belong to two or more classes, as illustrated below.) The intersection of the two lists of semantic classes is then computed. If the intersection is not empty, then the conjoining is allowed, and the intersection replaces the original list of classes associated with each conjunct. This serves as a record of any disambiguation that occurs as a result of computing conjunctional compatibility. For example, discharge is lexically ambiguous in medical narrative: it belongs to the MEDICAL-ACT class, as in:

	hospital	admission	and discharge
INITIAL LISTS	(INST)	(MEDICAL-ACT)	(MEDICAL-ACT,SIGN-SYMPTOM)
AFTER CONJ.	(INST)	(MEDICAL-ACT)	(MEDICAL-ACT)

whereas it has the SIGN-SYMPTOM reading in the phrase:

	no bleeding	or discharge
INITIAL LISTS	(SIGN-SYMPTOM)	(MEDICAL-ACT,SIGN-SYMPTOM)
AFTER CONJ.	(SIGN-SYMPTOM)	(SIGN-SYMPTOM)

Thus computing conjunction compatibility also provides a record of the correct reading when a conjunct is lexically ambiguous.

The conjunction rule does not account for the fact that certain noun phrases have a semantic class different from that of their head noun. Consider the following phrase from a building code text:

buildings	and	portions	of	buildings
BLDG		PART		BLDG

In this case, the head noun portion is "transparent" to its prepositional modifier, of buildings. That is, the phrase portion of buildings really has a distribution characteristic of the word building. If we could disregard the word portions of and look instead at the noun in the prepositional phrase, namely building with class BLDG, then the conjunction rule would apply correctly. We call this type of construction the "computed attribute" construction, because, in its most general form, the semantic class for an entire phrase is computed from the interaction of the semantic class of the head noun and the class(es) of its modifiers.

The computed attribute construction turns out to be characteristic of a set of English semantic classes common to many domains. These classes include the classes AMOUNT, PERIOD, BEGINNING, ENDING, FREQUENCY, etc. To handle these

constructions, an additional mechanism has been added as a restriction to the grammar. It operates on a word belonging to a transparent word class modified by an appropriate pre-nominal noun or adjective modifier (e.g. building portion), or by an appropriate prepositional phrase modifier (e.g., portions of buildings). It assigns to the phrase as a whole the class of the modifier and binds the modifier to the head noun, so that the modifier does not distribute over the other conjoined noun(s). This produces only the parse

```
buildings  and  (portions of buildings)
.BLDG           PART       BLDG
                     BLDG          <= COMPUTED ATTRIBUTE
```

and eliminates the parse (buildings and portions) of buildings.

The computer attribute is recorded as a special COMPUTED-ATTRIBUTE list on the head noun; it lists the semantic class(es) associated with the entire phrase. The COMPUTED-ATTRIBUTE list is assigned before conjunctional compatibility is checked, which allows the conjunction mechanism to make use of a COMPUTED-ATTRIBUTE, if one is present. On the basis of a computed attribute and conjunction compatibility, the following sentence received exactly one analysis:

```
There was no cyanosis      and no history  of prior seizures
          (SIGN-SYMPTOM)      (PERIOD)          (SIGN-SYMP-
                                                    TOM)
                  COMPUTED-ATTRIBUTE => (SIGN-SYMPTOM)
```

A preliminary test on a 22-sentence paragraph from a hospital discharge summary was run to compare the parses obtained with and without the above conjunction and computed attribute mechanisms. Five sentences in the paragraph contained conjoined noun phrases. These mechanisms reduced the total number of parses obtained for the five sentences almost in half: 8 parses total with the conjunction and computed attribute mechanisms compared to 14 parses total without them.

The conjunction and computed attribute mechanisms do not resolve all of the ambiguities of conjoined noun phrases. In particular, they do not always resolve the issue of distribution of right and left modifiers. However, they substantially reduce the number of incorrect parses and they require no modification for application to new domains.