FRAME BASED RECOGNITION OF THEME CONTINUITY

James T. Critz
Hewlett-Packard Company
Palo Alto, California
U.S.A

The paper describes a system which determines
continuity and shifts in English texts on the basis
of sentential themes. The theme object within the
thematic component of a sentence is determined and
a search is made to associate it with a frame
shared with the previous theme or themes. If the
theme object cannot be associated directly with one
of these frames, interpretive rules are applied to
do so indirectly through one of the frames normally
associated with the object but not yet with the text.

INTRODUCTION

This paper describes a system which interprets the semantic theme of
a text as a mapping of the semantic network of states associated in
the text with one or more theme objects onto an established semantic
frame. Theme objects are determined on the basis of the syntactic
distinction of thematic and rhematic (topic - comment) components of
a sentence. The concept of "a frame" as used here is that of a
semantic representation of experiences whose functional roles are
related, rather than of a psychological representation, where
physical properties of an object which do not bear any functional
relevance to the definition of the object and its relationship to
other components of the frame are included.

The problem of recognizing continuity and shifts in text analysis is
generally defined in terms of identifying the text's themes. It is
apparent from examples like (1 - 2) that a system which tries to
discover continuity on the sole basis of searching for antecedent
references for a pronoun or synonym is inadequate where a text is
developed through the techniques of elaboration, analogy, comparison
and other styles involving implied dependency.

(1) The store was crowded. The marketing campaign had been a
     success.

(2) Harold notified his lawyer. A storm was brewing.

In order to identify the implied theme object in text samples such
as these, the system described in this paper makes use of the
distinction of the syntactic categories "theme" and "rheme"
developed by the Prague School of Linguistics. Once it has
determined the sentential theme it further relates this theme to a

frame which can be reasonably associated with it in the given
context and links the frame with those extablished in the preceding
text. The objects which are the primary referents of a sentential
theme are assigned the role of being major nodes through which one
frame is linked directly or, where more abstract associations can be
inferred, indirectly with others relevant to the continuity of the
text. The concept of "frame" is that described in M. Minsky
(1977).

In the following sections the concepts of "theme" and "frame" are
defined more explicitly and the manner in which the system relates
them to each other is described. While the system is designed to
operate on English text, comparisons will be made to Czech.

THEMES

Within the grammatical theory of the Prague School of Linguistics a
sentence is composed of a minimum of two major constituents – a
"theme" ("topic") and a "rheme" ("comment"). (A sentence may also
contain components which function as a transition between these two,
but such components will not be discussed here.) The theme – rheme
division of a sentence typically conforms, in non-contrastive
environments, to the NP subject and VP major constituents of a
sentence, with the theme being the first of these to appear in the
surface word order of the sentence. Because of its more restrictive
word order, the theme of a sentence in English is normally its
subject. In Czech, however, a sentential subject may more easily
appear in the latter part of a sentence.

In its attempt to associate the theme of a sentence with a frame,
our system reduces the object of its operation to a "theme object",
defining this for English as the referent of the highest noun within
the theme constituent: normally, the subject head noun. Within VPs
functioning in English sentences as theme the theme object would be
the direct object's head noun, or, given other contrastive stress,
the head noun of the phrase containing the highest peak of
intonation were the sentence spoken.

It is important to note that the theoretical basis for labelling a
sentence constituent as theme does not necessarily mean that the
theme object has been previously mentioned in the text. The
immediate introduction of new theme objects in English is obvious
from the existence of non-definite themes in examples like (3).

(3)   The old hotel burned down last year. A new restaurant
      is to take its place.

However, unless some relation can be found between contextually new
themes and the existing context, sentences with non-definite themes
would be either make a text discontinuous or would maintain their
continuity only indirectly through the relation of the theme to
another noun in the same sentence which is anaphoric. In either
case there would be little justification for assigning a thematic
function to non-definite nouns, and the general definition of theme
proposed by the Prague School would have to be greatly altered or
dropped.

The association of themes with frames provides continuity without the requirement for explicit previous mention. The claim made in associating themes and frames is that themes are not referentially simple: their meaning involves reference not only to some individual, act, event, etc., but also reference to some structured set of these with which the noun referent is associated. Continuity lies in the existence of an assumed shared relation holding between a theme and other objects previously mentioned.

FRAMES

A frame may be generally defined as a limited complex of objects, acts, events, and/or states, whose association is determined by convention or experience and is assumed and expected by a speaker or author to be known by his or her audience. Simple frames include: the association of the objects "the bathroom" ("koupelna"), "the bedrooms" ("ložnice"), "the kitchen" ("kuchyně"), "the roof" ("střecha"), etc., with the frame "house", or the sequence "the morning" ("ráno" a "dopoledne"), "noon" ("poledne"), etc., as parts of the frame "day". Reference to one member of a frame suffices to make other members of the frame available as possible themes. For example, mention of "the kitchen and bathroom" implies the frame "house", and this mention, if themes are assumed to be bound to frames, suffices to allow the occurrence of "the house", "the garden" or other associated object as a theme in a later sentence.

The primary characteristics of a frame, as defined in our natural language system, and distinguishing frames from other relational structures, are that each member of the frame must be uniquely identifiable (This uniqueness also specifies the member's role in the frame.) and that the members share a common relation or property which uniquely characterizes the frame to which they belong. Recognition that a given object is a unique member of a frame is the attribution of the property of "definiteness" to that object. A non-definite theme is interpreted as a description of a frame object, but the reference itself as definite.

DETERMINING CONTINUITY

Themes and frames are related to each other through the following process. The main components of the system include:

- a parser
- a lexicon, providing not only the possible syntactic and semantic representations of each lexical item and conditions on their use, but also an index to the frames normally associated with the entry
- a frame dictionary, stating the set types of the possible members for each frame and their relationship to each other within the frame
- a set of logical inference and interpretive rules, which can be used to relate frames in the construction of a frame system representing changes in the continuity and development of a text
- a list of frames currently activated by theme reference, interpretation or inference

During the parse of a sentence the head noun of the sentence is
selected as the most likely candidate for being its theme and,
provided that tests of the sentence and the one preceding it do not
indicate a contrastive environment requiring the selection of
another noun as theme, declares it to be the theme. The frames
indexed in the lexical entry for the theme noun are then searched
and compared for identity with or inclusion within those frames
currently active. Those frames most recently used and those which
are associated with anaphoric noun phrases within the sentence are
examined first. If the sentence being processed is the first
sentence in the text or if a search of all active frames fails, the
theme is associated with a "base" frame. The base frame for a text
includes as its members the set types {AUTHOR}, {AUTHOR'S GOAL},
 READER and other nodes which may be filled by objects or persons
these as part of their immediate world.

The frame system for the text is built of the relationships holding
between all the activated frames and the base frame.

The process is illustrated for (4). For reasons of space, the
hierarchical relations which hold between members of each frame is
not shown. An AGENT frame is automatically assigned to persons
mentioned in a text. The LOCATION, BUILDING and LOCK frames in (4)
are interpreted.

(4)   John turned the key. The door still would not move.

        BASE   {... [John] [author] [reader] ...}

        AGENT {... [John] [LOCATION > BUILDING] [GOAL] ...}

        BUILDING {... [door] ...}

        LOCK   {... [key] [bolt] ...}

        DOOR   {... [lock] [handle] ...}

COMPARING ENGLISH AND CZECH TEXTS

Comparing the processing of English and Czech texts, it is important
to note that while most frames can be expected to be the same for
both languages, frame membership and relations between frame members
which are dependent on cultural, political, linguistic and similar
differences will cause frames to differ at least at the lower node
levels. The "day" frame mentioned earlier, in which English
"morning" is paralleled by "ráno" and "dopoledne" in Czech, is one
example.

An even more significant difference is found in the effect of
varying restrictions on word order in the two languages. Czech
seems to indicate shifts to new frames by moving subject nouns which
would be thematic in the English equivalent of the sentence to the end
of the sentence. Where the same frame or same samll set of frames
can be associated with consequetive subjects in the text, the
subject is thematic. An example of this can be seen in (5),
where "cesta" ("trip") is associated with the general frame of a
person's daily life and with Karel's in particular. In (6),

however, where a shift is made from an AGENT frame to introduce an
object which is not, on the basis of previous-themes, expected,
although an association could be found, Czech makes the subject
rhematic, and uses as its theme an object more easily associated
with existing frames. English may optionally keep "envelope" as the
theme of its equivalent for (6) and allows interpretive rules to
determine the frame it is associated with.

(5)   Karel se včera přestěhoval do nového bytu.   Cesta
      Karel himself yesterday moved to new apartment.   trip

      do práce už bude kratší.
      to work now will-be shorter.

      Karel moved to a new apartment yesterday. His trip
      to work now will be shorter.

(6)   Marie otevřela dveře. Na zemi ležela velká obálka.
      Marie opened doors.   on ground lay large envelope.

      Marie opened the door. { A large envelope lay on the ground.
                              { There was a large envelope laying
                              { on the ground.

REFERENCES

[1] Critz, J., Definiteness and Knowledge Frames, Paper presented
     to the Annual Meeting of the Linguistic Society of America,
     (December 1981).

[2] Danes, F., Papers on Functional Sentence Perspective (Cesko-
     slovensko akademie ved, Prague, 1974).

[3] Hirst, G., Discourse-Oriented Anaphora Resolution, American
     Journal of Computational Linguistics 7.2 (1981) 85-98.

[4] Johnson-Laird, P. and Wason, P. (eds.), Thinking: readings
     in cognitive science (Cambridge University, New York, 1977)

[5] Minsky, M., Frame System Theory, in: Johnson-Laird, P. and
     Wason, P. (eds.), Thinking: readings in cognitive science
     (Cambridge University, New York, 1977)