Tailoring Lexical Choice to the User's Vocabulary in Multimedia Explanation Generation

Kathleen McKeown
Jacques Robin
Michael Tanenblatt

Department of Computer Science
450 Computer Science Building
Columbia University
New York, N.Y. 10027
mckeown@cs.columbia.edu
robin@cs.columbia.edu
tanenbla@cs.columbia.edu

In this paper, we discuss the different strategies used in COMET (COordinated Multimedia Explanation Testbed) for selecting words with which the user is familiar. When pictures can not be used to disambiguate a word or phrase, COMET has four strategies for avoiding unknown words. We give examples for each of these strategies and show how they are implemented in COMET.

Topic area: Language Generation

Number of Words: 3222

Note: The 2 full page pictures at the end of the paper can be compressed into 1 inch figures using typesetting for camera ready copy, so we have not included them in the word count.

1 Introduction

Ideally, a language generation system should select words that its user knows, or at least, can deduce meaning for in context. While this would seem to involve simply substituting a known word for an unknown word (as is done, for example, in [Swartout 83]), in many cases avoiding an unknown word requires entirely rephrasing the original sentence. For example, in our domain of equipment maintenance and repair, if the user does not know the word “polarity” a sentence like “Check the polarity.” will be rephrased as “Make sure the plus on the battery lines up with the plus on the battery compartment.” Even when substitution can be made (e.g., substituting a referring expression for an object name), the substituted phrase can interact with other parts of the sentence requiring rewording in the original sentence.

In this paper, we discuss the different strategies used in COMET (COordinated Multimedia Explanation Testbed) for selecting words with which the user is familiar. Since COMET integrates both text and pictures in a single explanation (see [Feiner and McKeown 90a] for an overview of COMET and [Feiner and McKeown 90b, Feiner and McKeown 91] for details on media coordination in COMET), unknown words are frequently disambiguated through accompanying pictures. For example, COMET will use the most common object name even if the user is unfamiliar with the name, when the accompanying picture clearly shows the object and its location (this is similar to Appelt’s [Appelt 85] integration of language and physical actions for generating referring expressions). When pictures cannot be used to disambiguate a word or phrase, COMET has four strategies for avoiding unknown words:

1. Simple word or phrase substitution (e.g. replacing “a known good digital device” with “a working digital device”)
2. Rephrasing by reinvoking the content planner to provide conceptual definitions (e.g., replacing “Check the polarity” with “Make sure the plus on the battery lines up with the plus on the battery compartment.”)
3. Rephrasing by reinvoking the content planner to select attributes for referring expressions (e.g., replacing “the COMSEC cable” with “the cable that runs to the KY57”)
4. Using past discourse to construct a referring expression (e.g., replacing “Test the COMSEC cable.” with “Test the cable you just removed.”)

In the following sections, we first compare our work with previous work in the area. We then provide a overview of lexical choice in COMET, showing how and where it occurs in the overall system. Each of the strategies is then described in turn, prefaced by a brief discussion of disambiguation of unknown terms through pictures.

2 An Overview of Lexical Choice in COMET

An overview of COMET's architecture is shown in Figure . On receiving a request for an explanation via a menu interface, the **content planner** uses schemas [McKeown 85] to determine which information should be included from the underlying **knowledge sources** in the explanation. The explanation content, represented as a hierarchy of logical forms (LFs) [McKeown et. al. 90] is passed to the **media coordinator** [Feiner and McKeown 90b, Elhadad et al. 89], which refines the LFs by adding directives indicating which portions are to be produced by the **text generator** and which by the **graphics generator** [Seligmann and Feiner 89].

Figure 1: COMET System Architecture

The **Lexical Chooser** is part of the text generator [McKeown et. al. 90]. On receiving the annotated hierarchy of logical forms, it determines the overall grammatical form of each sentence based on the semantic structure of the LFs (e.g., conditional sentences are generated for precondition-action structures) and selects the words and phrases to be used for semantic concepts of the LF. The Lexical Chooser performs its task using a rewriting system implemented on top of the Functional Unification Formalism (FUF) [Elhadad 88, Elhadad 90], an extension of functional unification grammars [Kay 79]. The lexicon itself consists in a base of rules, with each rule rewriting a given set of semantic features into a corresponding set of lexical and syntactic features. The Lexical Chooser uses FUF extensions to access external knowledge sources from within FUF through the use of coroutines [Elhadad and Robin 92]. This is used to represent constraints on lexical choice from the user model, from the underlying knowledge bases, and from past discourse. Since the Lexical Chooser is situated within the text generator, it can also use constraints from the syntactic form of the sentence. Finally, as we will show later in the paper, COMET can reinvoke the content planner to replan sentence content when the original LF can not be lexicalized. Thus, COMET can use a wide variety of interacting constraints on lexical choice.

Each lexicon entry has a semantic concept as its key and lists different possible words that can be used to realize the concept. Constraints are used as tests to determine which of the alternative words or phrases should be selected. The user model indicates both the reading level of the current user[1], any individual words that COMET knows the user does not understand, and any wording preferences (e.g., the user knows abbreviations, the user is familiar with military terminology). We make no claims about which of these forms of user models is better (e.g., is easier to acquire), but simply show that if this information is available, COMET is able to use it in selecting words for the explanation. In the following sections, we show how this information is used to constrain word choice.

3 Multimedia Disambiguation

An accompanying picture often makes clear what the referent of a referring expression is. If the user is unfamiliar with a term, the accompanying picture will define it. For example, Figure 6 shows one step of an explanation generated by COMET for loading frequency into the radio. The text refers to a “FCTN knob” and the accompanying picture clearly shows the knob and its identifying labels on the front panel of the radio. Location is conveyed by showing surrounding context of the knob and highlighting is used to clearly identify the knob [Feiner and McKeown 91].

COMET can generate an explicit reference to the accompanying illustration (called a *cross reference*) in cases where the explanation would not otherwise clarify a referent. By querying the layout representation maintained by the graphics side of the system, the content planner can construct object descrip-

---

[1]We currently use two levels for a poor and good reader

tions that explicitly refer to the graphic illustrations. These descriptions can take into account special graphical devices of the illustrations such as cutaway views, ghost images, highlights and arrows. The cross reference shown in Figure 7 would be generated if the user does not understand the term “holding battery”. Currently, cross referencing is implemented in a separate module and is not part of the integrated COMET demo. Generating cross references requires issuing queries to IBIS’ illustration representation, full interaction with IBIS to request generation of accompanying pictures, and selection of one of several different possible forms of cross referencing. For full details on cross referencing in COMET, see [McKeown et al 92].

4 Word/phrase Substitution

Whenever possible COMET simply replaces an unknown word by another word or phrases. Some paired sentences showing alternative wordings are shown below in Figure 2. The first word is generated if the user’s vocabulary level is above a certain reading level or if the word is explicitly listed as not known. A lexicon entry for one of these word pairs is shown in Figure 3. It shows that for any concept under the concept c-disconnect in the knowledge base taxonomy, COMET will use the word “disconnect” if the user’s vocabulary level is above level 6[2]. Otherwise, COMET uses the word “remove”. COMET also checks whether the user knows abbreviations or not and if so, will use a referring expression such as “FCTN knob” as shown in Figure 6. If not, COMET uses the full name (“function knob”) and if COMET has no information about the user, it generates the abbreviation and relies on the accompanying illustration to clarify the referent.

1. Replace the digital device with a *known good/working* digital device.
2. Screw the new manpack antenna onto the RT and tighten until the manpack antenna is *snug/tight*.
3. *Disconnect/Remove* the COMSEC cable from the KY57 audio connector.
4. This will cause the display to shown *an arbitrary/some* number.

Figure 2: Word Substitutions Generated by COMET

[2]this is an arbitrary level that divides users into poor readers and good readers. Any division could have been used.

```
(lhs ((concept #(under c-disconnect))))
(rhs ((process ((cat verb-group)
                (alt (((CONTROL (OK-Lex-UM 'c-disconnect 6))
                       (lex "disconnect"))
                      ((lex "remove")))))))))
```

Figure 3: Lexicon Entry Selecting Disconnect or Remove

5 Rephrasing through Replanning

In constructing an alternate description of a semantic concept that uses only user-known words, simple word replacement is not always a viable strategy. COMET handles two distinct cases. The first involves describing the concept at a more detailed semantic level of abstraction and is done by retrieving alternate knowledge base definitions. The second involves creating referring descriptions when object names are not known. In this case, the content planner is reinvoked to provide attributes of the object that can be used in the description.

5.1 Retrieving alternate knowledge base definitions

Sometimes the original text uses a word or phrase that abstracts the details of an action to allow generation of a very concise expression. If unfamiliar with the word or phrase, the user will be unable to infer the specifics of the action needed to perform the task. Alternative wordings require choosing a less abstract level of semantic decomposition at which to describe the action. In COMET, this problem is handled by having the lexical chooser reinvoke the content planner to retrieve an alternate definition of the concept from the knowledge base.

This strategy is used for rephrasing the request “Check the polarity” which COMET issues when providing instructions for installing a new holding battery. Alternate methods for checking the polarity are stored as different tokens of the concept **c-polarity** in the knowledge base[3] and can be used if the user is unfamiliar with the word “polarity.” Figure 4 shows two tokens represented for the concept polarity in the LOOM knowledge base [Mac Gregor & Brill 89]. Figure 5 shows the original LF and Figure 6 shows the new **Phenomenon** role filler that replaces the **phenomenon** role filler in the old LF. The remainder of the LF remains the same in the replanned version. The resulting rephrased sentence is “Make sure that the plus on the battery lines up with the plus on the battery compartment.”. Since the object of the sentence is an embedded sentence, COMET can use either the verb “check” or the collocation “make sure” as the

[3]Note that another alternate method for checking polarity is to make sure that the bump on the battery is on the top after inserting it in the radio

verb realizing the mental process concept **c-check**. The fact that “make sure” cannot be used in the original sentence is based on syntactic constraints in the lexical chooser itself.

```
;; Instance definitions for polarity

(tellm (polarity polarity-1)
       (polarity polarity-2))

;; More detail is represented for one instance

;; polarity is represented as two plusses
       (:about polarity-2
               (identified plus-1)
               (identifier plus-2))
;; one is located on the battery
       (:about plus-1 (on-loc battery-1))
;; one is located on the battery compartment
       (:about plus-2 (on-loc bc-1))))
```

Figure 4: Knowledge base tokens for polarity

In the full paper, we will also provide details on another class of examples that use this strategy: those where the action can be defined by describing its substeps. In this case, the unknown term is left in the explanation and several explanatory sentences defining it are added. This is used, for example, to define how to do a self-test following the sentence “Run the self-test.” when the user does not know the term “self-test.”

5.2 Generating New Referential Descriptions

If the user does not know the name of an object, the content planner must be reinvoked to generate object attributes that can be used in a referential description. This aspect of our work is similar to others [Dale 88, Appelt 85, Reiter 90], but our selection algorithm is not as sophisticated because we do not use a detailed model of user beliefs. More importantly, we account for interaction between the new description and other parts of the original sentence and rephrase where necessary. Basically, there are two types of object attributes that can be used in a referring expression in COMET: object subpart (and part-of) relations and spatial relations to other objects. While other physical attributes are represented (e.g., size, color), if they are to be used it is better to use a cross

```
;; Initial LF:

((TOKEN-TYPE COMPOUND-EVENT)
 (DIRECTIVE-ACT SUBSTEPS)
 (SUBSTEPS
  ((DISTINCT
    ~(((PROCESS-TYPE MENTAL)
       (CONCEPT C-CHECK)
       (TOKEN T-CHECK-POLARITY-1)
       (TOKEN-TYPE ATOMIC-EVENT)
       (SPEECH-ACT DIRECTIVE)
       (ROLES
        ((PROCESSOR  ((CONCEPT C-USER)
                      (TOKEN T-USER-1)
                      (TOKEN-TYPE PHYSOBJ)
                      (REF-MODE NAME)))
         (PHENOMENON ((CONCEPT C-POLARITY)
                      (TOKEN T-POLARITY-1)
                      (TOKEN-TYPE ABSTRACT)
                      (QUANTIFICATION ((DEFINITE YES)))
                      (REF-MODE DESCRIPTION)))))))))))
```

Figure 5: Logical Form for Original Sentence

reference to an accompanying illustration that conveys the attributes since a picture can convey size and color much better than can words. COMET selects that attribute that uniquely identifies the object.

For example, consider the sentence “Disconnect the COMSEC cable from the KY57 audio connector.” If the user does not know the term “COMSEC,” COMET selects a spatial relation, **connects** that exists between the radio (called RT) and the KY57, since this cable is the only one that connects the radio and the KY57. If the user does not know the name “KY57 audio connector” COMET uses the part-of relation between KY57 and audio-connector to simply omit “audio connector” as there is only connector on the KY57 into which a cable could be plugged. Since it knows no other name for “KY57” it relies on an accompanying illustration to disambiguate this reference. Selecting these two attributes would result in the generated sentence “Disconnect the cable that connects the RT to the KY57 from the KY57.” However, COMET notes the redundancy in descriptions (two side by side references to the KY57) and determines that it can remove one from the description of the cable by selecting the verb “runs to” instead which only requires one role in the generated sentence. This results in the sentence “Disconnect the cable that runs to the RT from the KY57.” Finally, to remove ambiguity, the from location is moved to result in “Disconnect from the KY57 the cable that runs to the RT.” Details of the implementation of this ex-

```
((TOKEN-TYPE COMPOUND-EVENT)
 (DIRECTIVE-ACT SUBSTEPS)
 (SUBSTEPS
  ((DISTINCT
    ~(((PROCESS-TYPE MENTAL)
       (CONCEPT C-CHECK)
       (PROCESS-TYPE ATOMIC-EVENT) (TOKEN T-CHECK-POLARITY-1)
       (SPEECH-ACT DIRECTIVE)
       (ROLES
        ((PROCESSOR  ((CONCEPT C-USER) (TOKEN T-USER-1)
                      (TOKEN-TYPE PHYSOBJ) (REF-MODE NAME)))
         (PHENOMENON
         ((PROCESS-TYPE CIRCUMSTANTIAL) (CONCEPT C-POLARITY)
          (TOKEN-TYPE ATOMIC-RELATION) (TOKEN T-POLARITY-2)
          (RELATION-MODE EQUATIVE) (SPEECH-ACT ASSERTIVE)
          (ROLES ((IDENTIFIED ((CONCEPT C-PLUS)
                               (TOKEN-TYPE PHYSOBJ)
                               (TOKEN T-PLUS-1)
                               (REF-MODE DESCRIPTION)
                               (ROLES
                                ((ON-LOC ((CONCEPT C-BATTERY)
                                          (TOKEN T-BATTERY-1)
                                          (TOKEN-TYPE PHYSOBJ)
                                          (QUANTIFICATION
                                           ((DEFINITE YES)))
                                          (REF-MODE DESCRIPTION))
                                 )))))
                  (IDENTIFIER
                   ((CONCEPT C-PLUS)
                    (TOKEN-TYPE PHYSOBJ) (TOKEN T-PLUS-2)
                    (REF-MODE DESCRIPTION)
                    (ROLES
                     ((ON-LOC ((CONCEPT C-BC)
                               (TOKEN-TYPE PHYSOBJ)
                               (TOKEN T-BC-1)
                               (QUANTIFICATION ((DEFINITE YES)))
                               (REF-MODE DESCRIPTION)))))))
                  ))))))))))))
```

Figure 6: New Phenomenon Role for Rephrased Sentence

ample are not provided here due to lack of space, but will be provided in the full paper.

6 Using Past Discourse

For subsequent reference, the presence of a discursive context allows for a wider variety of strategies to get around gaps in the user's vocabulary. COMET takes advantage of this fact by maintaining a discourse history. Before sending a concept description for realization to the lexical chooser the content planner records the description into the discourse history. It creates one record for the description as a whole and a separate record for each of its roles. The entry has four fields:

- The name of the concept.
- The description used in the reference.
- The described action in which the referring description plays a role.
- The list of roles that the description fills in that action request.

For each subsequent reference, the concept name is used as the access key and the three other fields are updated; they thus always contain the information on the *last* reference. By looking up information in the discourse history, the content planner is able to construct object descriptions in terms of the last action it was involved in. This provide an additional alternative strategy for describing objects that can be used to cope with gaps in the user vocabulary.

Consider the following paragraph generated by COMET when the user has no vocabulary gaps:

1. "Disconnect the COMSEC cable from the KY57 audio-connector."
2. "Plug in the handset to the KY57 audio-connector."
3. "Test the COMSEC cable."

When the user is unfamiliar with the term "COMSEC" sentence (3) will be rephrased using past discourse as "Test the cable *that you disconnected.*" Implementation of this class of examples is in final stages and we expect it to be completed by the time the reviewers read the paper.

7 Related Work

COMET performs several lexical choice tasks. When generating an action request, it chooses a level of detail in the action description appropriate to the user. When generating both initial and subsequent referring expression, it selects a set of distinguishing properties of the referent and chooses words to express the selected properties. It can substitute words or phrases for any part of speech.

For each of these subtasks, COMET takes into account a wide variety of constraints:

- *In generating action requests*: the user's vocabulary
- *In generating initial references*: the user's vocabulary, the illustration accompanying the text, taxonomic relations, part-of relations and locative relations.
- *In generating subsequent references*: in addition to those for initial references, participation of the referent in actions requested in the previous discourse, location of the referent mentioned in the pre-

vious discourse.

COMET is the first attempt to deal with these issues in a single system, although other generation systems also perform some of these tasks and take into account some of these constraints. In addition, because COMET is a multimedia system, it can use the accompanying illustrations advantageously for disambiguation.[4] Both EPICURE [Dale 88] and KAMP [Appelt 85] address the issue of choosing the level of detail at which to describe an action, but they do not constrain this choice upon the user's *lexical* knowledge. EPICURE uses the user's *domain* knowledge and KAMP mutual beliefs about the domain, notions which are related to but distinct from a model of user's vocabulary. These two systems also address the issue of reference, but only the selection of appropriate properties to distinguish an object. They do not choose between alternative wordings for the selected properties. In contrast, this type of choice is made by FN [Reiter 90], which, like COMET, uses an explicit model of the user's vocabulary. However, FN does not deal with rewordings of actions and does not use past discourse to avoid terms the user does not know.

Bateman and Paris [Bateman and Paris 89] do address the problem of rephrasing a text based on user class. While their approach is quite interesting, it is substantially different from ours. They develop the notion of *registers* for different classes of users, allowing the system to basically use different dialects depending on which class of users it is addressing. Different registers define different mappings from semantic concepts to phrases for each user class (currently they have 3: programmer, end users who want to following reasonings, and end users who want no explanations). The difference in the generated text appears to be primarily in the amount of detail included (e.g., the most abstract text just specifies that "inputs are fine" while more detailed texts give values for the inputs) and modification of objects (e.g., whether and how quantification is used). While one of our rephrasing techniques (see section 5.1) does vary detail, it provides definitional information for terms and this does not appear to be an issue that Bateman and Paris address. Note that the difference in our approach is that while Bateman and Paris predefine a language for <different user classes through the use of register mappings, in COMET different terms can be mixed and matched depending on the individual user model. Each approach is appropriate for different circumstances.

[4] WIP [Wahlster et al. 89] can also generate cross references though not based on a user model.

8 Conclusions and Future Work

In the previous sections we have seen the different strategies that COMET uses to cope with a gap in the user's vocabulary. We have illustrated each strategy with an example. However, COMET is also able to simultaneously use several strategies to cope with multiple vocabulary gaps. For example, COMET will use both an accompanying illustration and either word substitution or rephrasing when nether alone results in an understandable example. When faced with a choice of strategies, any one of which is as good as the other, COMET uses the following ordering:

1. multimedia disambiguation
2. word substitution
3. previous discourse
4. rephrasing

We have shown here how COMET integrates a wider variety of strategies for rewording than other generation systems and how it handles cases requiring substantial rephrasing.

Figure 7: Accompanying Picture Clarifies Referent

Figure 8: The Use of Cross References:
Remove the holding battery, shown in the cutaway view

References

[Appelt 85] Appelt, D.E.
Planning English Sentences.
Cambridge University Press, Cambridge, England, 1985.

[Bateman and Paris 89]
Bateman, J.A. and Paris, C.L.
Phrasing a text in terms the user can understand.
In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 1511-1517. Detroit, MI, 1989.

[Dale 88] Dale, R.
Generating referring expressions in a domain of objects and processes.
PhD thesis, University of Edinburgh, 1988.

[Elhadad 88] Elhadad, M.
The FUF Functional Unifier: User's Manual.
Technical Report, Columbia University, June, 1988.

[Elhadad 90] Elhadad, M.
Types in Functional Unification Grammars.
In *Proceedings of the 28th meeting of the Association for Computational Linguistics.* Pittsburgh, Pa, June, 1990.

[Elhadad and Robin 92]
Elhadad, M. and Robin, J.
Control Strategies for Functional Unification.
In *submitted to 6th International Workshop on Language Generation.* Trento, Italy, April, 1992.

[Elhadad et al. 89] Elhadad, M., Seligmann, D., Feiner, S., and McKeown, K.
A Common Intention Description Language for Interactive Multi-media Systems.
In *A New Generation of Intelligent Interfaces: Proceedings of IJCAI89 Workshop on Intelligent Interfaces*, pages 46-52. Detroit, MI, August 22, 1989.

[Feiner and McKeown 90a]
Feiner, S. and K.R. McKeown.
Generating Coordinated Multimedia Explanations.
In *Proceedings of the IEEE Conference on AI Applications.* Santa Barbara, CA., March, 1990.

[Feiner and McKeown 90b]
Feiner, S. and K.R. McKeown.
Coordinating Text and Graphics in Explanation Generation.
In *Proceedings of the National Conference on Artificial Intelligence.* Boston, Mass., August, 1990.

[Feiner and McKeown 91]
Feiner, S. and McKeown, K.R.
Automating the Generation of Coordinated Multimedia Explanations.
IEEE Computer 24(10):33-41, October, 1991.

[Kay 79] Kay, M.
Functional Grammar.
In *Proceedings of the 5th meeting of the Berkeley Linguistics Society.* Berkeley Linguistics Society, 1979.

[Mac Gregor & Brill 89]
Mac Gregor, Robert and David Brill.
LOOM Reference Manual.
Technical Report, USC-ISI, Marina del Rey, CA, 1989.

[McKeown 85] McKeown, K.R.
Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text.
Cambridge University Press, Cambridge, England, 1985.

[McKeown et al 92]McKeown, K. R., Feiner, S.K., Robin, J., Seligmann, D., and Tanenblatt, M.
Generating Cross References for Multimedia Explanations.
In *submitted toProceedings of AAAI-92.* AAAI, July, 1992.

[McKeown et. al. 90]
McKeown, K.R., Elhadad, M., Fukumoto, Y., Lim, J., Lombardi, C., Robin, J., and Smadja, F.
Language Generation in COMET.
In Mellish, C., Dale, R., and Zock,M. (editors), *Current Research in Language Generation.* Academic Press, London, 1990.

[Reiter 90] Reiter, E.B.
Generating appropriate natural language object description.
PhD thesis, Center for research in computing technology, Harvard University, 1990.

[Seligmann and Feiner 89]
Seligmann, D.D., and Feiner, S.
Specifying Composite Illustrations with Communicative Goals.
In *Proc. ACM Symposium on User Interface Software and Technology*, pages 1-9. Williamsburg, VA, November 13-15, 1989.

[Swartout 83] Swartout, W.R.
XPLAIN: a system for creating and explaining expert consulting systems.
Artificial Intelligence 21(3):285-325, 1983.

[Wahlster et al. 89]Wahlster, W., Andre, E., Hecking, M., and T. Rist.
WIP: Knowledge-based Presentation of Information.
Technical Report WIP-1, German Research Center for Artificial Intelligence, Saarbruecken, FRG, May, 1989.

Table of Contents