Generation of Coherent Monologues

Jan Odijk
Institute for Perception Research (IPO)
P.O.Box 513 5600 MB Eindhoven, The Netherlands
e-mail: odijkje@prl.philips.nl

Abstract

In this paper a method for generating coherent texts is described. In this method only local conditions associated with sentences determine the appropriateness of a sentence at a certain point in the text. The method does not require any form of planning and it concentrates on maximizing the amount of variation of the texts generated.

1 Introduction

The topic of this paper is the generation of coherent texts. I will describe a method for generating coherent texts in which only local conditions associated with sentences determine the appropriateness of a sentence at a certain point in the text. The method does not require any form of planning and it concentrates on maximizing the amount of variation of the texts generated.

The contents of this paper will be as follows. First, I will introduce the general setting in which the text generator functions (section 2). An impression of the functionality of the system will be given by showing an example of some database information and of a text generated by the system (section 3). In section 4 I will point out some minimum requirements for the generation of coherent texts. I will describe the approach adopted to achieve maximum variation with minimum means in section 5. In section 6 the actual mechanisms to achieve coherency will be discussed. I concentrate on two aspects, viz. how to ensure that information presented in a text is grouped naturally (subsection 6.1) and how the relevant information can be presented in a natural order (subsection 6.2). I summarize the essential properties of this approach in section 7. Since this is a report on work in progress, I will point out some problems and undesirable aspects and I will explain how I propose to solve these problems (section 8). Finally, the major conclusions will be recapitulated in section 9.

2 Setting

The purpose is to generate correctly pronounced coherent texts which convey information from a database. An important requirement is that the texts generated are as varied as possible. Such a system may serve a useful purpose in all kinds of telephone services, e.g. tele-shopping (where a catalogue of products is the relevant database, and information on specific products must be conveyed); in audio or video on demand, where subscribers can acquaint themselves with the movies or songs available in the system. The texts generated may be an aid in choosing a movie or composition. Generally speaking, we have a user in mind who does not yet exactly know what he wants, and who wants to browse through a catalogue of available options, while the text generated presents relevant information in an auditive manner.

A more detailed description of this and similar settings in which such a text-generation system might be useful and a more detailed description of the generation system as a whole is given in Van Deemter et al. (1994). For certain aspects relating to the importance of ensuring correct pronunciation of the texts generated, I would like to refer to Van Deemter (1994). I will not discuss any aspects of the dialogue between the user and the system.

As already mentioned above, variation is important in the applications we envisage. The reason for this is that we expect that people will listen to several texts generated while browsing through the database. If these texts do not show sufficient variation, we expect that this will be very boring.[1] The texts must be sufficiently varied, in different ways, so that users at least will not regard them as an impediment to browsing through the database, and will preferably actually enjoy the browsing.

Variation can be achieved in many ways. Some of the ways in which we intend to obtain variation are:

- by varying the contents, the length and the degree of detail of the texts
- by grouping information in different ways in different texts
- by presenting information from a certain perspective, e.g. if the user has indicated specific interests.
- by taking into consideration information presented in earlier texts, referring to it and contrasting the current information with the earlier information
- by varying the form of the individual sentences
- by varying the form of the texts generated

I will discuss only the last of these below.

[1] In the near future we intend to design a system which presents all information in the same format each time to prove this.

3 Example

In this section I will present an example of the kind of input required for the generation system and a (real) example of a text that might be produced, to give an impression of the functionality of the kind of system I have in mind.

We have chosen the instrumental works by Mozart from the Philips Mozart Collection as a concrete domain for this text-generation system. The relevant information of the various compositions was encoded in a database. An example of some of the information in this database and the way in which it has been represented is shown here:

KV 309

DATE 10/1777 - 11/1777

SORT piano sonata

NUMBER 7

PERFORMER Mitsuko Uchida

PLACE London

VOLUME 17

CD 2

TRACK 4

From the full database entry of this composition the system can now generate texts, a real example of which is given below. This example still contains errors and infelicities which must be eliminated, but they will not be dealt with here.[2]

> The following composition is the first part of the seventh sonata. The composition is a sonata for piano in c. Influences of the Mannheimian orchestral techniques are discernible. The KV number of sonata Number seven is K. three zero nine. This work was composed for Rosine, the daughter of the court musician and composer Christian Cannabich in Mannheim. Mozart composed the middle part as a musical portrait of Rosine.
>
> The recording of K. three zero nine took place in London, England, in February nineteen eighty five. The quality of the recording is DDD.
>
> The seventh sonata consists of three parts: allegro con spirito, andante un poco adagio and rondo allegretto grazioso. The first part lasts five minutes thirty one seconds. The three parts are located on tracks four, five and six of the second CD of volume seventeen.

[2]Numbers in the text are written out in full since the text (actually, an enriched version of it) is input for a system which correctly pronounces the text.

> The piano is played by Mitsuko Uchida.
>
> K. three zero nine was written by the composer between October seventeen seventy seven and November seventeen seventy seven, in Mannheim.
>
> The following is a fragment of this allegro con spirito.

and now a fragment of the first part of this composition is played.

4 Minimum Requirements for Generating Coherent Text

The question is: how can we generate coherent text from database information in the form indicated by the given example.

First, there are a number of minimum requirements. The basic ingredients of texts are sentences, so we need a mechanism for generating sentences. In addition, anaphoric devices must be used appropriately within these sentences and in the sequence of sentences. I will not deal with these issues in this paper, but simply assume that sentences are available and that the anaphoric devices are used appropriately. Actually, the sentences generated are not just strings, but strings enriched with syntactic structures and various other annotations required for adequate handling of anaphoric devices and certain other aspects. For a discussion of these issues I would like to refer to the aforementioned Van Deemter et al. (1994).

Various approaches can be adopted for generating a coherent text given the relevant sentences with appropriate anaphoric devices. In one approach, with which we briefly experimented in an earlier phase, on a different domain, one could write an explicit grammar which states where each sentence may occur. A different approach could make use of a form of planning, i.e. grouping fragments of information to be conveyed before their linguistic realization in such a way that a coherent text results. Many other approaches are also conceivable.

In the approach we adopted we concentrated on the requirement that the texts must show maximum variation.

5 Variation

Since, as indicated above, variation is of the utmost importance, we adopted an approach with which variation can be maximized. We do not encode all the possible variations in a text grammar, but assume, as a starting point, that in principle each sentence can occur anywhere. Conditions now have to be imposed to prevent sentences from occurring in inappropriate positions.

One could perhaps compare this strategy with the strategy followed in transformational grammar: in the first stages of the development of this theory, various specific rules where written, and the application order of these rules was explicitly encoded. Later, the application order of the rules was left unspecified (each rule can in principle be applied anywhere), and general principles are to prevent wrong applications or application orders of rules. Whereas in an explicit grammar it is specified explicitly where each sentence can occur, in the current approach it is assumed that each sentence can occur anywhere, in principle, but conditions will prevent its occurrence in certain cases.

This is a simple strategy for maximizing the amount of variation without having to explicitly specify all the possible kinds of variations.

6 Text Coherency

We must now define the conditions which determine whether a sentence is appropriate.

First, I will assume that a different system determines what is to be said. This system will determine this in cooperation with the user, who specifies his/her interests. The information to be conveyed is stored in a variable called *WHATTOTELL*. In addition, we associate each sentence with an attribute which states what the sentence conveys. This attribute is called *TELLSABOUT*. This attribute need not be stipulated, but can be computed during the generation of the sentence. I will not go into this here.

Secondly, I will assume that two factors determine the coherency of a text, viz. (1) the information must be presented in a natural order, and (2) the information must be presented in natural groupings. There may be other factors contributing to coherency, but they will not be considered here.

6.1 Natural Grouping

To start with the latter factor, each sentence is associated with one or more *topics*. These topics give a more general characterization of what the sentence is about than the attribute *TELLSABOUT*. Examples of such topics are: *tells-about-recording, tells-about-performers, etc.*

Each possible topic is, in turn, made the *current topic*. Each topic corresponds to a paragraph. A prerequisite for a sentence to be uttered is that the current topic is a member of the topics of the sentence. This will ensure that sentences with the same topic occur together within one paragraph, and in that way a natural grouping of information is achieved. The order of sentences within a paragraph, and the question whether a sentence may or may not occur, even if it includes the current topic among its topics, is determined by other conditions, which will be specified below.

The grouping of information is clearly visible in the example text. The sentences of the second paragraph are all about the recording and the sentences of the third paragraph are all about the parts of this composition. If these sentences were not grouped, but scattered throughout the whole text, a much less natural text would result.

This accounts for a natural grouping of information. We must now still ensures a natural order.

6.2 Natural Order

To get sentences in a natural order, I will first assume the existence of a *knowledge state*. This knowledge state keeps track of which information has been presented before and which information has not yet been conveyed. In addition, it keeps track of the way in which this information has been conveyed, in particular: has it been conveyed explicitly or implicitly? For instance, if Mozart wrote a composition in March 1766, then we can convey this date by an explicit expression such as *in March 1766*, or by a more implicit expression such as *when he was only ten years old*. Finally, the knowledge state keeps track of when the relevant information was presented (e.g. how many sentences, paragraphs or texts ago).

Next, I will assume that each sentence is associated with conditions formulated in terms of this knowledge state. For instance, a sentence such as *The following composition is a piano sonata* can only be used if the composition and its sort have not been introduced earlier.

Finally, each sentence is associated with a number of actions to be performed on the knowledge base after it has been uttered.

In that way, one can view sentences as functions which map a knowledge state which satisfies certain conditions into a different knowledge state, as specified by the actions associated with the sentence.

A sentence can be used if the following conditions are satisfied: First, the value of *TELLSABOUT* must be a subset of the value of *WHATTOTELL*. Secondly, as we have seen above, the current topic must be a member of the topics of the sentence to ensure the naturalness of the grouping of information. And finally, the conditions of the sentence on the knowledge state must evaluate to true.

If more than one sentence may occur now (which will often be the case), one is chosen arbitrarily.

After a sentence has been uttered, the associated actions update the knowledge state, and a new sentence can be generated.

In that way, the information is naturally grouped and presented in a natural order.

7 Essential Properties

I will now summarize the main properties of this approach. First, only local conditions of a sentence on the knowledge state, and its topic(s), determine the possibility of occurring at a certain point in a text. The conditions are 'local' in the sense that they have no access to what other possible sentences might convey, and they are only sensitive to the knowledge state at the point at which the sentence is to be uttered. This approach ensures maximization of the amount of possible variation. No planning, i.e. grouping of information before its linguistic realization, is required. No backtracking is allowed, i.e. once a sentence can be uttered and has been uttered, one can no longer retract the sentence and try a different sentence instead. No global properties of the text are determined by a grammar or a schema or set of schemata. The system can be extended very simply, namely by adding a sentence, its topics and the conditions and the actions on the knowledge state. Once they have been specified, the sentence can function fully in the text generator.

8 Problems and Further Research

As already indicated above, the work presented here is still in progress. I would like to point out a number of undesirable properties and problems that the system currently faces and indicate some ways in which these problems might be solved. Basically, there are four such problems.

First, in the current system, *topics* of sentences are simply stipulated. But it is clear that there is at least some overlap with the attribute *TELLSABOUT*. It would be desirable to be able to compute topics from this attribute. In the near future, I want to investigate whether this is feasible.

Secondly, the conditions on the knowledge state are stipulated for each individual sentence (or actually for each object from which a sentence is generated). It would again be desirable to derive such individual conditions from more general considerations. For instance, it may be possible to derive a part of the conditions from the structure of the database. A reasonable condition associated with sentences is frequently that if a sentence provides information about specific properties of an entity from the database, then the relevant entity must have already been introduced earlier. This could be a more general principle which would obviate the need to stipulate specific instantiations of this principle with each individual sentence. Secondly, certain conditions appear to encode whether information in the sentence is presented (by linguistic means) as *new* or as *given* information. Such conditions could be computed automatically from the sentence by formulating general rules for how *new* and *given* information is linguistically encoded in sentences.

Thirdly, the grouping mechanism appears to function correctly, but in certain cases it leads to paragraphs which are too small. This is illustrated by the fourth and fifth paragraphs of the example text, which each consist of a single sentence. It may hence be necessary to reconsider the relation between a topic and a paragraph.

Fourthly, since there is no explicit planning, there is no guarantee that all the information which must be conveyed (as represented in *WHATTOTELL*) is actually conveyed, even if all the sentences required are in principle available.

This can be simply illustrated. Suppose that there are three sentences, A, B and C, and two topics: *topic1* and *topic2*. A and B are associated with *topic1*, and C with *topic2*, and the conditions are such that B can only occur after C.

Now first choose *topic1*. B cannot be uttered, since C has not been uttered yet. But A can be uttered, and it will form the only sentence of the paragraph having *topic1* as its topic. Next, we choose *topic2*, and sentence C can be uttered. But then we have no more topics, and there is no way of getting sentence B uttered.

This situation may arise, of course, due to the fact that the topics can be chosen independently of the conditions and of TELLSABOUT. One way of solving this problem is by computing the topic from other properties in such a way that this situation cannot arise. We intend to investigate this option in the near future.

Another possibility is to turn this defect into a virtue. It is possible to get sentence B uttered by introducing an appropriate connecting sentence, e.g. *By the way, what we forgot to mention...*, or *We have to add...*, etc, and then starting up all topics again. If the situation does not arise too often, this will actually make the texts more lively and more natural, since it appears to mimic the behavior of human beings when they speak spontaneously.

Currently, the situation occurs very rarely, which suggests that it may be possible to compute the topics so as to avoid the problem altogether, but there is no guarantee that it will not occur, so we would certainly like to find a principled solution to solve this problem.

9 Concluding Remarks

I have presented a method for generating coherent texts from information formally represented in a database. A coherent text is obtained by ensuring that the relevant information is presented in a natural order and in natural groupings. I have pointed out some problems of the current system and some undesirable properties, and indicated ways of overcoming these problems.

I have concentrated on the fact that the texts should be as varied as possible, and have adopted a strategy for maximizing variation without having to specify all the possible variations explicitly. The system can be extended in a very simple manner.

A coherent text is obtained by minimum means. The information is presented in a natural order by formulating conditions on a knowledge state, which, in my opinion, any text generator will require in some form anyway. The information is naturally grouped by specifying topics, which we hope we will be able to compute automatically from independent properties in the near future. Other means for achieving coherency have not been considered, but may prove necessary.

It is still too early to be able to fully evaluate the system. The measures for ensuring coherency appear to be sufficient for the texts generated in the current domain. But it is yet to be investigated whether this will remain the case when other domains, or different kinds of texts, are considered. Other mechanisms may be required for different kinds of texts in addition to or instead of the ones currently employed. I will leave this to future investigations.

References

Van Deemter, K. (1994). Contrastive stress, contrariety and focus. paper presented at CLIN V, Twente University.

Van Deemter, K., Landsbergen, J., Leermakers, R., and Odijk, J. (1994). Generation of spoken monologues by means of templates. In *Proceedings of TWLT 8*. Twente University. Twente.