Being As Informative As the User Wants: The generation of information enriched utterances

Stina Ericsson
Göteborg University
Department of Linguistics
Box 200, 405 30 Göteborg
Sweden
stinae@ling.gu.se

Abstract

The work in this paper is based on the claim that collaborative dialogue involves taking into account how much information an utterance conveys, in the sense that it thereby shows how it is connected to the linguistic and extra-linguistic context shared by speaker and hearer, by system and human user. I introduce a notion of *information enrichment* to handle this phenomenon, and discuss information enriched utterances in human-human dialogue and in the generation component of a dialogue system.

Introduction

Human communication often succeeds in providing just the amount of information needed. That is, it neither contains too much nor too little information, where too much information would involve unnecessary repetition of something already known and make for clumsy dialogue (and may even cause confusion as to precisely what has already been established by the dialogue participants), and too little information would increase the risk of misunderstanding and other communication failures.

As an example, in a post office domain, some of the possibilities for realising the message that sending a letter to Italy costs 53 cents are:

(1) B: *A letter to Italy costs fifty-three cents*
B′: *Fifty-three cents*
B″: *Fifty-three*

Which realisation is appropriate, hence what the right amount of information is, depends, in general, on a number of factors such as the communicative goal of the speaker, the speaker's view of what has been established so far in the dialogue, and what (other) information (the speaker thinks) is accessible to the hearer at a given point in the dialogue.

The generation of utterances containing the right amount of information in this sense, involves an investigation of what I term *information enriched constituents*.

This paper is concerned with the generation of information enriched constituents in a dialogue system. Information enrichment in human-computer interaction in part concerns different issues than does information-enriched interaction between humans, but many issues are also the same. In fact, for information enrichment, it is important to remember that one of the dialogue participants, the user, is still subject to e.g. memory constraints, although the system is not. This means that the system's utterances must be adpated to the user's constraints. Hence, it is important to work out general models of how users view aspects of context and their accessibility for information enrichment.

Modelling the user, qua interpreter of information enrichment, also implicates solving some specific HCI issues. All-embracing issues involve creating a system that appears co-operative and friendly, that finds a balance between efficiency and security (in the form of transparency and feedback), and that avoids misunderstanding and confusion. I will show that information enrichment plays a role in all of these.

In particular, I discern the following key issues: (1) how the information enrichment approach can be used to create a natural, co-operative, and friendly system, and (2) how the approach can be used to avoid misunderstanding. The second issue involves the sub-issues of adjusting to the user's level of expertise, making use of speech recognition scores, and (semantically) monitoring the progress of the dialogue.

In the next section I describe in more detail just what I mean by information enrichment. Then, in the section *Modelling information enrichment using human-human dialogue*, I use corpora to analyse what is needed for modelling information enriched communication. The section *Generating information enrichment in a dialogue system* then describes the generation component of a dialogue system producing information enriched utterances. And the section *Adapting to different users* discusses the user models that make information enrichment a solution to the issues identified above. Finally, I go through a sample interaction with the system.

Information enriched constituents

I define an *information enriched constituent* as a constituent whose content in a shared context, *the contextual content*, is the result of embedding its *compositional content* in a larger semantic structure. The object of study is spontaneous spoken dialogue between humans and between humans and ma-

chines.[1]

As an example, take B' in (1) above, in the context of a question:

(2) A: *How much is a letter to Italy?*
B: *Fifty-three cents*

In the context of A, the information enriched B does not merely give a price – which is the compositional context of B – but conveys that 53 cents is the price for a letter to Italy, which is the contextual content. The embedding structure, then, is something like '*the price of a letter to Italy is* _' for this example, and it embeds the (compositional) content of B's utterance to give the contextual content. A more precise characterisation of the larger embedding structure will be given in the next section.

The information enrichment approach embodies a view of human communication that is at least partly Gricean. One central assumption is that utterances are adjusted to the context in that they contain enough compositional content for the hearer to create the contextual content (when these two differ), but typically not more (*the maxim of quantity*). A second assumption is that speakers typically use information enrichment whenever they can (cf. 'Be brief' as part of *the maxim of manner*). See (Dale & Reiter 1995) for a related interpretation of Grice's maxims for the generation of referring expressions.

A note on terminology. I sometimes talk about information enriched *constituents* and sometimes information enriched *utterances*. The information enriched constituent B in (2) above was also an utterance, but in the general case an information enriched constituent need not make up a whole utterance (an utterance can consist of one or more additional sentences or 'sentence-like' parts), hence the term constituent instead of utterance. In this paper I will only look at information enriched constituents that are also utterances, and the two terms will be used interchangeably.

Note, finally, that information enrichment is not the same as non-sentencehood. B in the following is an example of a sentential information enriched constituent:

(3) A: Why are you in such a hurry to get to the post office?
B: It closes at five

Modelling information enrichment using human-human dialogue

With a view to equipping a dialogue system with the ability to generate information enriched utterances in an appropriate way, there are three main issues to be dealt with:

(i) *Information units in the utterance*: how much, or little, information in an information enriched utterance?
(ii) *Informational components in the context*: how is context structured for information enrichment?
(iii) *The rules*: what rules govern accessibility of contextual components to the utterance?

[1]Information enrichment is related to but also differ from accounts of ellipsis, short answers and fragments, (Ericsson ms).

Information units in the utterance

The A and B turns in (4) are quite a common type of question-answer exchange in spontaneous spoken dialogue, with B giving no more and no less than the information asked for by A.

(4) A: at what time does it get to Toronto?[2]
B: ah three thirty five p.m.
B′: ah it gets to Toronto at three thirty five p.m.

Compare this to B' where the utterance also contains material already present in the preceding question (viz. *'it gets to Toronto at'*). A useful distinction can then be made on information structure grounds beteween, on the one hand, that part of the utterance which reflects the context, the information that is assumed to be shared by the speakers (*'it gets to Toronto at'* in B'), and, on the other hand, the part that is informative, that is to update the shared information (*'three thirty five p.m.'* in B'). This division is similar to the focus-ground distinction made in linguistics – see e.g. (Vallduví 1992) and (Ginzburg 1999) for some (fairly) recent accounts – and I will therefore use this terminology here, calling the contextual part *ground* and the informative part *focus*. Thus, the answer in B contains no ground – it consists of only a focus.

In addition to focus-ground, I make a further distinction within the focus, based on examples like (5):

(5) G1: Where are you in relation to the top of the page just now?[3]
F1: Uh, about four inches.
G2: Four inches?
F2: Yeah.
G3: Where are you from the left-hand side?
F3: About two.

A full-focus answer in $F3$ would have been '*(about) two inches*'. The element '*two*' in $F3$, leaving '*about*' out of the discussion, I classify using a notion of prominence, which is here taken as a semantic notion. A *prominent element* is typically used to mark contrastive or otherwise important material within the focus. My notion of prominent element is related to focus within a rheme as used by (Steedman 2000).

Summing up, as an initial formulation for the generation of information enrichment, a non-information-enriched utterance consists of a full ground and a full focus, whereas an information enriched utterance consists of either a full focus or a prominent element.[4]

[2]The A and B turns are taken from the AmEx Travel corpus: http://www.ai.sri.com/%7Ecommunic/amex/amex.html. B' is a constructed utterance.

[3]Excerpt from the HCRC Map Task corpus: http://www.hcrc.ed.ac.uk/dialogue/maptask.html

[4]Note that this is not the complete picture. An information enriched utterance may, for instance, consist of a focus together with some ground material less than the full ground.

Informational components in the context

How does context need to be structured for the generation of information enriched utterances, or, to put it differently, what contextual components contain information that can enrich an utterance?

We have already come across one such component: a question, or an issue, under discussion. This is straightforwardly exemplified in (2), (3), and (4). The answer in each of these addresses what is currently under discussion as introduced by the immediately preceding question. The answers do not contain a ground as part of the compositional content, but instead use material in the question for the ground part of their contextual content.

Not only a question currently being discussed, but also past questions are available as informational components in the context. One example is $G1$ in (5), which enriches $G3$ with '*of the page (just now)*'. Note that parallelism between $G3$ and $G1$ are important for the establishment of information enrichment here.

Another contextual component is that which has already been agreed to in the dialogue. An example is $F1$ (or $F1$ together with $G1$), which enriches $F3$ with the element '*inches*' ($F3$ of course also makes use of the current question under discussion, $G3$). Again, parallelism between utterances ($F3$ and $F1$, and also $G3$ and $G1$) plays an important role here.

These three components are all part of the linguistic context. Another type of component is domain context. For instance, in a house-buying domain, an utterance giving a price as '*two fifty*' means '*two hundred and fifty thousand*', that is, it translates not as 250 but 250,000.

Another domain example can be seen in (5) above. That $F1$ conveys that F is four inches *below* the top of the page – as opposed to *above* or anywhere else – comes from domain knowledge: in the Map Task domain you cannot be outside of the page, so the only position to be in relation to the top of the page is below.

From this brief discussion, we can see that we need to keep track of at least the following informational components in the context: the current question under discussion, previously asked questions, propositions (assumed) jointly agreed to, and domain knowledge that govern information enriched usage.

The rules governing informational accessibility

Having defined the amount of information in an utterance in terms of ground, focus and prominent element, and having identified some of the contextual components containing information used for enrichment, what are the rules governing the accessibility of contextual information, which in turn determines the extent to which a given utterance can be enriched by other utterances?

First of all it is important to note that not all information is available all the time. As a rule of thumb, more recent material tends to be more accessible, so something like *distance* seems to place a constraint on information accessibility for information enrichment.

In the Map Task dialogues I have looked at, there are, for instance, very few examples of the word '*inches*' supplied through information enrichment when a measure is given. One notable exception is $F3$ in (5) above. There, '*inches*' has been active throughout a number of preceding turns, that is, the distance is very short. Similarly, a question like $G1$ in (5) is not available throughout the dialogue, but only within a short distance.

Next, imagine $G2$ and $F2$ in (5) being replaced by the following $G2$ and $F2$ ($G1$, $F1$, and $G3$ remaining the same):

(6) G1: Where are you in relation to the top of the page just now?
F1: Uh, about four inches.
G2: Where are your gorillas in relation to the mountain?
F2: Just below
G3: Where are you from the left-hand side?

When the contextual content for $G3$ is to be determined in this modified context, it is not clear whether it is enriched by $G1$. Indeed, it is quite difficult to say whether the left-hand side of the page or of the mountain is meant. Note that $G1$ is at exactly the same distance – in number of turns – from $G3$ in both (5) and (6). (6) gives an example of what I call *interference*, borrowing a term from (Givón 1983) and adapting it to information enrichment: interference is the presence of semantically compatible contextual material that can give rise to ambiguity.

Distance and interference are then two kinds of constraint that govern the production of information enrichment. Another rule is that information enriched constituents often follow upon other information enriched constituents in adjacency pair exchanges. In (7), an information enriched question is followed by an information enriched answer:

(7) A: going where?[5]
B: Orange County

There are also language-specific morphosyntactic constraints that govern the final form of an utterance in terms of its degree of information enrichment.

Generating information enrichment in a dialogue system

Information enriched constituents have been implemented in the GoDiS dialogue system. In this section I describe the overall system architecture, and in particular the generation component for information enrichment.

The system

GoDiS (Gothenburg Dialogue System) is an information state based dialogue system built using the Trindikit, (Larsson & Ericsson 2001), (Larsson *et al.* 2000).

GoDiS uses a blackboard architecture, see figure 1. The information state acts as the blackboard, and represents information that a dialogue participant has at any given point during the dialogue. It is separate for each dialogue participant, and is updated through dialogue moves made by the

[5] Excerpt from an AmEx dialogue.

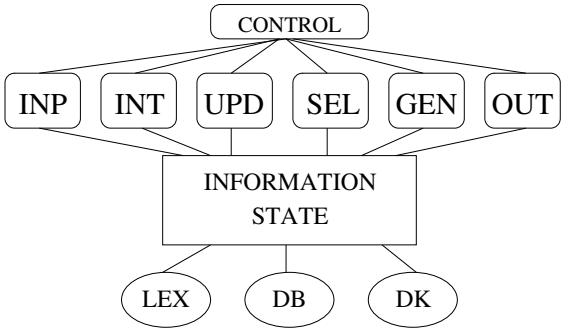

Figure 1: The system architecture

speakers. Examples of moves are asking a question and giving an answer.

The six boxes above the information state in figure 1 are modules for INPUT, INTERPRETATION, UPDATE, SELECT, GENERATE, and OUTPUT, respectively. These all write to and read from the information state. The three circles beneath the information state are passive resources such as lexicon/grammar, database, and domain knowledge. Passive resources cannot read from and write to the information state, but they are attached to the information state through resource interfaces. At the top, a control algorithm wires together the modules.

In non-information-enriched versions of the GoDiS system, the mechanisms of generation are divided between the SELECT and GENERATE modules. Based on what is in the information state, SELECT chooses the next move to be made, including an appropriate GoDiS-style move content in the form of, roughly, a quantifier-free predicate logic formula. GENERATE then maps this move and content to a corresponding string. The OUTPUT module, finally, is responsible for producing output in the form of text or speech.

Generating information enriched utterances

The generation of information enriched utterances in GoDiS presupposes formalisations of the three issues identified in the previous section. I will start by the second of these issues – informational components in the context – as the formalisation of the first issue (information units in the utterance) makes use of the informational components in the context.

Contextual components How are the contextual informational components represented in the GoDiS system? Domain knowledge pertaining to information enrichment, once determined for a particular domain, is simply encoded in the domain resource. The other three components that were identified are all represented in the information state proper.

The question under discussion is represented by a variant of QUD as defined in (Ginzburg 1996):

(8) a set that specifies the currently discussable questions, partially ordered by $\prec$ ('takes conversational precedence'). If q is maximal in QUD, it is permissible to provide any information specific to q using (optionally) a short answer.

In GoDiS, QUD is a stack of questions.

Propositions that (the speaker assumes) are jointly agreed to, are represented in the form of a set of propositions, called commitments. Questions not currently under discussion but recently raised, together with recent utterances not in the form of questions, are recorded in a dialogue history – a queue of utterances – of a length that is determined empirically.

Information units Let us now turn to the first of our questions in the previous section, concerning information units in the utterance, and discuss just how these – focus, ground, and prominent element – are determined by the system.[6]

The SELECT module for information enrichment operates in two steps (these two steps could also be separated into two different modules). During the first step, the system determines what move is to be made, and what the semantic content of this move is going to be.

The second step then tries to partition the content of the move into information structural units, using the content of the information state. Two main rules are currently responsible for this:

- The rule `qudFG` of the rule class `select_fg`. Determines focus-ground from the question on QUD.

  - Preconditions: The next move to be made is an answer move, and there is a question topmost on QUD such that the answer move answers the question and the answer and the question are parallel.
  - Effects: Focus is assigned to the appropriate element(s) in the answer move (ground is left unmarked).

- The rule `dialhistFProm` of the rule class `select_fprom`. Identifies prominent elements within the focus from the dialogue history.

  - Preconditions: The next move to be made is an answer move, and among the utterances in the dialogue history there is another answer move such that the two answer moves are parallel.
  - Effects: Prominence is assigned to the appropriate element(s) in the focus of the answer move (the appropriate elements are the *alternatives*, see below).

The check for parallelism used by the two rules is encoded as follows. An information unit is a proposition, question, ground, focus, or basic element, where a basic element is a simple predicate or argument. Two information units, $a = a1 \circ a2$ and $b = b1 \circ b2$ ($\circ$ means composition), are parallel when $a1$ is parallel with $b1$ and $a2$ is parallel with $b2$. Two basic terms are parallel when they are either identical or alternatives (alternatives belong to the same semantic sort as defined in the domain, but are non-identical).

The rules Turning to the third and final question from the previous section, rules governing accessibility of information, distance is encoded in the dialogue history. One way

[6]The work on information structure determination for information enrichment builds on previous work on determining information structure for prosodic realisation, (Kruijff-Korbayová *et al.* 2003).

of doing this is in the form of a dialogue history of a fixed length, which is what is used in the current version of the system. There, dialogue history contains the preceding three utterances (note that an utterance may consist of several dialogue moves, for instance a greet and an ask move, or several answer moves), as this is all that is needed for the question-answer dialogues currently handled in terms of information enrichment.

A more complex alternative is to allow the dialogue history to have a flexible length, based on the content of the utterances and the semantic relationships between them, that is, in a way that enables a more elaborate way of calculating constraints based on distance and interference. The first approach sees distance as a matter of number of utterances, whereas the second approach regards distance in terms of questions under discussion: the distance is short between utterances concerning similar or identical questions under discussion.

Interference is handled by parallelism based on semantic sorts and by the dialogue history. As an example, if the utterance '*Three*' was used to refer to a price, and is part of a given dialogue history, this utterance, although accessible with regard to distance, cannot subsequently enrich an utterance of the kind produced by B in (9):

(9) A: By the way, what time is it?
B: ??Ten past

That is, times and prices are seen as belonging to different semantic sorts, and '*Ten past*' cannot use the previously uttered price '*Three*' in this context to convey that the time is ten past three.

Adapting to different users

We have seen how a corpus study of information enrichment can be used to identify important aspects of this phenomenon, and how these can be incorporated in the generation component of a dialogue system to determine foci and prominent elements for enriched utterances.

What remains to be discussed is the final step in the generation process, the work carried out in the GENERATE module: the determination of just which information units are to be realised in the actual utterance, that is, whether an utterance is going to be information enriched or not.

This ties in with the two overarching HCI issues identified in the introduction. The first of these is the issue of how the information enrichment approach can be used to create a natural, co-operative, and friendly system. Firstly, we already have part of the key to this: producing information enriched utterances conforms to the maxim of quantity and part of the maxim of manner. Given that these are part of what a human dialogue participant would consider co-operative and natural, generating information enrichment gives us the effect that we want.

Secondly, a possibility for user models is suggested by the corpus data. As we noted above, an information enriched constituent is often followed by another in an adjacency pair. This gives the proposal that the system could simply adapt to the user's level of information enrichment (at the same time taking care to show the user that information enrichment *is* handled by the system). This is similar to the process of *information co-ordination* studied in (Garrod 1999), and is likely to make the system seem more natural and co-operative.

The second major issue put forward in the introduction is how the information enrichment approach can be used to avoid misunderstanding. This issue has several dimensions. One aspect also ties in with the user's experience of the system as regards reliability: users not used to a dialogue system are often uncertain as to its capabilities, and greatly benefit from ample feedback from the system. Experienced users, on the other hand, are more concerned with the efficiency of the interaction.

A distinction between naive and experienced users can thus be made, and, in terms of information enrichment, naive users should be met by many ground-focus utterances, whereas experienced users benefit more from utterances consisting of only a focus or a prominent element. At the application level, an analogous distinction can be made between systems that are used seldom by a particular user (such a system should generate less information enrichment) and systems that are used more often by the same user and are maybe personalised (those systems should generate more information enrichment).

Of course, giving feedback to the user and verifying system understanding involves more than producing utterances consisting of both a focus and a ground; ground information, on its own, needs to be manipulated in different ways. While implicit verification can be achieved through utterances consisting of both ground and focus, explicit verification is achieved through asking the user to comfirm what the system has understood as ground. In the information state approach, this issue of *grounding* has been discussed at length in (Larsson 2002).

The production of ground or otherwise known information in dialogue has several functions. Three such functions are identified and studied in (Walker 1996). *Attitude* is part of what I would call grounding. *Attention* involves making dialogue participants focus on certain entities, and *consequence* means making inferences explicit. All of these functions overlap with information enrichment, but they have not been the focus of the present study.

Another way of avoiding misunderstanding, besides using ground information in various ways, is to use recognition scores from the speech recogniser. If the scores are high, it is safe to rely on information enrichment. If the scores are low, information enrichment should be avoided. The same idea is used in (Jokinen & Wilcock 2001), where the notions *NewInfo* and *Topic* are used in a way that resembles *focus* and *ground* as used here.

A third way of avoiding misunderstanding, or further misunderstanding, is a careful monitoring of error rates in the sense of misunderstandings and confusion in the dialogue, and the adjustment of the level of information enrichment to this. If the dialogue goes smoothly, information enrichment is to be used. This issue is clearly connected to the use of implicit and explicit information. It may also be a more adaptive way of handling the difference between naive and

experienced users.

In GoDiS, systems used often vs. seldom by the same user can be 'hardwired' as maximisation and minimisation, respectively, of information enrichment. Adapting to the user's level of information enrichment can be done with the help of the information state as a user model, recording whether the user's utterances were information enriched or not. Speech recognition scores are supplied by the system and can be straightforwardly used. Errors in the sense of misunderstanding and confusion can be handled using information in the information state: the moves, with contents, that the user makes in response to the system's implicit and explicit verifications.

A sample interaction

The current application is a small system in the post office domain, capable of answering questions about postages for letters and parcels to various countries around the world. It can also answer questions about what time it is. It is a system that assumes that the user always wants to be met by a co-operative, friendly and efficient system, and it therefore produces a maximally information enriched utterance whenever it can. This behaviour can naturally be modified in the ways indicated above. Let us walk through a dialogue example.

(10) U1: How much is this parcel to Italy?
S1: Two euros
U2: And (how much is) this (parcel) to Japan?
S2: Three
U3: Do you know what time it is?
S3: A quarter past three

When the system is about to produce $S1$, `qudFG`, but nothing else, fires because of $U1$'s being topmost on the QUD. Thus, only the focus is produced, given the system's goal to maximally use information enrichment as encoded in the GENERATE module. For $S2$, both `qudFG` and `dialhistFProm` fire, since now there is also a parallel proposition in the dialogue history. The alternative in the focus, the element *3*, is marked as a prominent element.

In the final utterance, $S3$, `qudFG` fires (leaving out '*The time is*'). However, `dialhistFProm` does not fire (if it had, $S3$ would only have consisted of '*A quarter past*'), since even though the element *3* is in the dialogue history, times and prices belong to different semantic sorts, and parallelism between $S3$ and $S2$ fails.

Conclusions

In my work on the generation component of a dialogue system, I have used the idea that dialogue participants use the amount of information in their utterances to reflect the context and to provide new information.[7] To implement this, it is necessary to provide a model of (what the system believes to be) the user's view of the dialogue at a given point, including what information is accessible.

[7] Another way of doing this is through intonation, (Kruijff-Korbayová *et al.* 2003).

I have provided such a model through analysing information enrichment in human-human dialogue, and I have discussed the incorporation of the findings from this analysis in the generation part of a dialogue system, shown a sample interaction with this system, and also discussed some further ideas for user modelling in relation to information enrichment.

Naturally, a number of issues remain. Some of them are to do with extending the coverage of the current generation component: include information enriched dialogue moves other than answers, include richer versions of the constraints on distance and intereference, and include other constraints. Another issue is working out just how utterances consisting of only ground information interact with information enrichment.

References

Dale, R., and Reiter, E. 1995. Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive Science* 19.

Ericsson, S. ms. Information rich constituents in dialogue. unpubl ms.

Garrod, S. 1999. The challenge of dialogue for theories of language processing. In Garrod, S., and Pickering, M., eds., *Language Processing*. Psychology Press. 389–415.

Ginzburg, J. 1996. Interrogatives: Questions, facts and dialogue. In *The Handbook of Contemporary Semantic Theory*. Oxford: Blackwell.

Ginzburg, J. 1999. Semantically-based ellipsis resolution with syntactic presuppositions. In Bunt, H., and Muskens, R., eds., *Computing Meaning*, volume 1. Kluwer. 255–279.

Givón, T. 1983. Topic continuity in discourse: An introduction. In Givón, T., ed., *Topic Continuity in Discourse: A quantitative cross-language study*. John Benjamins.

Jokinen, K., and Wilcock, G. 2001. Confidence-based adaptivity in response generation for a spoken dialogue system. In *Proceedings of the 2nd SIGdial workshop*.

Kruijff-Korbayová, I.; Ericsson, S.; Rodríguez, K. J.; and Karagjosova, E. 2003. Producing contextually appropriate intonation in an information-state based dialogue system. In *Proceedings of EACL'03*.

Larsson, S., and Ericsson, S. 2001. Godis – qud-based dialogue management in a multi-domain dialogue system. Technical abstract of demo, NAACL 2001.

Larsson, S.; Berman, A.; Bos, J.; Grönqvist, L.; Ljunglöf, P.; and Traum, D. 2000. Manual for trindikit 2.0. Deliverable D5.1, TRINDI.

Larsson, S. 2002. *Issue-based Dialogue Management*. Ph.D. Dissertation, Göteborg University.

Steedman, M. 2000. *The Syntactic Process*. MIT Press/Bradford Books.

Vallduví, E. 1992. *The Informational Component*. Garland.

Walker, M. A. 1996. The effect of resource limits and task complexity on collaborative planning in dialogue. *Artificial Intelligence Journal* 85(1-2).