Dynamic Optimisation of Information Enrichment in Dialogue

Stina Ericsson
Department of Linguistics
Göteborg University
Box 200
40530 Göteborg, Sweden
stinae@ling.gu.se

Abstract

Information enrichment is a process whereby explicitly realised information elements in a dialogue message make use of other information elements that are accessible through the context. I introduce information enriched constituents using four information structure primitives: *ground*, *focus*, *prominent element*, and *non-prominent material*. Five Optimality Theoretic (OT) constraints are then formulated for the generation of information enriched constituents in a dialogue system, and I show how dynamic constraint reranking is needed in a dialogue system. The usefulness of bidirectionality is also shown, and I end with a discussion of computational considerations.

1 Introduction

Information enrichment is the exploration of how some information elements in a dialogue message are explicitly realised as part of an utterance, and how others, the enriching elements, are accessible through the context.

Optimality Theory (OT), (Prince and Smolensky, 1993), has in recent years been applied to semantics and pragmatics, (e.g., (Hendriks and de Hoop, 2001; Buchwald et al., 2002; Zeevat, 2001; Beaver, 2004)). Below, OT is explored for information enrichment, and addresses the question: what light can OT shed on the generation of information enriched constituents in a dialogue system? The investigation leads to dynamic rerankings of a fixed set of constraints.

The paper focuses on the question of which information elements are to be realised and which ones not. The final realisation of an information enriched constituent includes considering morphosyntactic constraints, for instance, but these are not part of present considerations.

The following section elucidates the concept of information enrichment with the help of information structure. Section 3 presents the constraints, and section 4 the constraint rankings. 5 discusses the determination of which ranking to use, and 6 bidirectionality in the context of information enrichment. Finally, section 7 considers computational issues.

2 Information Structure and Information Enrichment

In what follows, utterances will be seen as connecting to the context along two dimensions. The first is illustrated by B' in the context of A in (1). '*Jones*' is the informative part that is meant to update the current information state, whereas '*my last name is*' is an anchor to what has already been established in the dialogue. I will call the former *focus* (F) and the latter *ground* (GR), following, e.g.,Vallduví (1992) and Ginzburg (1999).

(1) A: ok and what's your last name?[1]

[1] A and B are taken from the Amex Travel Agent Data, http://www.ai.sri.com/%7Ecommunic/amex/amex.html. B'

B: ah Jones

B′: my last name is Jones

The other dimension along which utterances connect to the context is illustrated by example (2).[2] Simplifying somewhat by leaving '*about*' out of the discussion, in $F3''$ the ground is something like '*I am ... from the left-hand-side of the page just now*', and the focus '*two inches*'. The element '*two*' in the focus is an alternative – in roughly the sense of alternative-evoking focus in (Rooth, 1996) – to '*four*' in $F1$, and will here be called a *prominent element* (P). Hence, prominence is here a semantic notion, used to mark constrastive or otherwise important material within the focus. The element '*inches*' in $F3''$ is non-prominent material (NON-P) within the focus.

For accounts that use the term focus in a way reminding of prominent element as introduced here, see (Pulman, 1997; Steedman, 2000).

Having introduced the relevant information elements, I define an *information enriched constituent*, or utterance, as one whose content in a shared context, *the contextual content*, is the result of embedding its *compositional content* in a larger semantic structure. For the purposes of the current study, the compositional content of an information enriched constituent consists of a single focus, or a single prominent element, or a prominent element together with ground, with the other information elements being supplied by the context. A non-information enriched constituent consists of a full ground and focus.

Examples of information enriched constituents in (1) and (2) – again ignoring '*about*' – are B (a focus), $F1$ (a focus), $G3$ (a focus consisting of a prominent element and some non-prominent material), $F3$ (a prominent element), and $F3'$ (a focus consisting of a prominent element and non-prominent material).

The information enrichment approach covers partly the same dialogue phenomena as do approaches to what is variously called ellipsis, fragments, non-sententials, etc. (e.g, (Ginzburg, 1999; Schlangen, 2003)).

is a constructed utterance. 'Jones' in utterance B has been substituted for the anonymised name 'C' in the Amex transcipt to make the example more readable here.

[2] $G1$-$F3$ are from the HCRC Map Task corpus, http://www.hcrc.ed.ac.uk/dialogue/maptask.html. $F3'$ and $F3''$ are constructed utterances.

3 The constraints

Various considerations govern the amount of information – in terms of which information elements – a given utterance is to contain. In an OT setting, some of these considerations are encoded as constraints, and some of them in the ranking of these constraints. For the information elements introduced above and concerning information enrichment, five constraints are involved (not given in rank order):

FOCUS: Generate focus
***NON-PROM**: Avoid non-prominent material
GROUND: Generate ground
***GROUND**: Avoid ground material
PROM ELEM: Produce the prominent element

GROUND is a faithfulness constraint conveying that what is part of the input should also be part of the output. The mirror constraint *GROUND is instead a markedness constraint prescribing economy and simplicity.[3]

All five constraints reflect that the optimisation of utterances in terms of their desired degree of reliance on information enrichment, is a matter of balancing between markedness and faithfulness constraints, between dialogue economy and explicitness. The approach to discourse anaphora by Buchwald et al. (2002) involves similar considerations for noun phrases and the salience of referents.

4 Dynamic reranking of constraints

The generation component of a dialogue system can take a number of issues into account for determining the level of reliance on information enrichment in an utterance to be generated:[4]

• High speech recognition scores, rely on information enrichment; low speech recognition scores, rely less on information enrichment, or not at all

[3] Note that both of these constraints are needed – neither is sufficient on its own for all the rankings.

[4] The first three issues are also considered by Jokinen and Wilcock (2001) for NLG, but not in terms of OT.

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

• Polite/formal system, rely less on information enrichment; informal system, rely more on information enrichment
• Beginning of the dialogue, rely less on information enrichment; rest of the dialogue, rely more on information enrichment
• Naive users (or a system that is used seldom by the same person), rely less on information enrichment; expert users (or a system that is used often by the same person), rely more on information enrichment
• Adapt to the user's level of reliance on information enrichment, making the system appear more co-operative (cf. (Garrod, 1999))

Any one of these factors affects the ranking of the constraints introduced above, and their values will give rise to different rankings. For instance, a dialogue system that is designed to be very formal and correct, will make use of a constraint ranking where information enrichment is rare among the optimal candidates throughout the dialogue. The opposite is true for a more informal system.

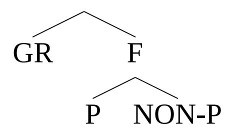
What's more, a dialogue system can be designed to *rerank the constraints* depending on conditions that change during the dialogue. For example, a system making use of recognition scores, will use one type of ranking if the score was high, and will need to rerank the constraints if the score was lower. Another example is the reranking of constraints to give an optimal candidate that relies on information enrichment to the same extent that a preceding user utterance does. In human-human dialogue the latter can be seen in the frequent occurrence of information enriched question-answer pairs (see, e.g., $G3 - F3$ in (2)).

Concretely, in a dialogue like (2) above, the question is whether to generate $F3$, $F3'$, or $F3''$, and how to rank the constraints to give precisely the desired optimal candidate.

I will now go through the different rankings that can be selected dynamically. First, a note on input and candidates. Input is here the contextual content of an utterance to be generated, where the content is marked up for information structure. For the tableaux below, a contextual content with all of ground, focus, prominent element and non-prominent material is used. For explanatory purposes, a prominent element may correspond to 'two', a focus (prominent element together with non-prominent material) to 'two inches', a ground to 'the distance is', and a ground-focus to 'The distance is two inches', all in the context of, say, the question 'What is the distance?'.

The candidates created from the input are the 'power set' of the information elements in the contextual content, with the reservation that the mark up is hierarchical (see the figure in example (3) below): the presence of a prominent element and non-prominent material implies the presence of a (full) focus. In the tables below this is indicated using the notation 'P_NON-P/F'.

(3)



As is usual, a dotted line between two constraints indicates that the ranking of these two constraints in relation to each other is unknown, that is, the outcome is independent of the order of these two particular constraints in relation to each other. In addition, I will use a double line to indicate indeterminacy between several constraints, in the sense that it demarcates partial rankings.

Note that my use of partial rankings is not to be confused with the partial orderings of Anttila (1997). He uses partial rankings between constraints to explain examples of variation in Finnish morphology; different total orders (different tableaux) created from the partial one (the grammar) give different winners. In my approach, constraints are partially ordered because there is no conflict between them; the various total rankings that can be created from the partial ranking all give *the same* winner.[5]

4.1 Maximal reliance on information enrichment

Maximal reliance on information enrichment is the generation of just the prominent element when such can be determined. It involves the following partial rankings: *NON-PR, PROM ELEM >> FOCUS, and *GROUND >> GROUND. The tableau is given in figure 1.

In this tableau, the left-most column lists all the output candidates, as described above. The ranking between PROM ELEM and *NON-PR is not known (or, equivalently, their ranking in relation to each other does not affect the outcome), as indicated by the dotted line, but both of them are ranked higher than FOCUS. *GROUND is ranked higher than GROUND, but the ranking of these two constraints in relation to the other three does not change the result, which is the meaning of the double line.

Each star indicates a violation of a constraint by a candidate, and the optimal candidate is determined in the usual way, which can be described as: the optimal candidate is the one with the fewest violations of the highest constraint on which the two candidates differ.

Thus, in figure 1, *P* ('Two') is the optimal candidate. Informally, and intuitively, what this tableau says, is that for maximal reliance on information enrichment, avoiding ground is more important than producing ground, and producing the prominent element and avoiding non-prominent material are both more important than producing a full focus.

[5]The assumption in OT is that theoretically there *is* a complete ranking. My partial rankings are then to be interpreted as that given the current constraints there is no way of finding this ranking.

4.2 Minimal reliance on information enrichment

Minimal reliance on information enrichment means producing a full ground-focus utterance. The partial rankings are FOCUS >> *NON-PR, and GROUND >> *GROUND, and the tableau is given in figure 2. The optimal candidate is *GR P_NON-P/F* ('The distance is two inches').

4.3 Intermediate reliance on information enrichment

Intermediate reliance on information enrichment occurs in two cases. In one case, the optimal candidate is *P_NON-P/F* ('Two inches'), and the partial rankings involved are FOCUS >> *NON-PR, and *GROUND >> GROUND. This is depicted in figure 3.

In the other case, figure 4, the partial rankings are *NON-PR >> FOCUS, and GROUND >> *GROUND, to make *GR P* ('The distance is two') the optimal candidate. For this second case, the presence of PROM ELEM is required to separate *GR P* from the candidate involving just *GR*, and it can be ranked anywhere among the constraints.[6]

4.4 Focus-ground and all-focus utterances

The discussion and tableaux above assumed an utterance whose contextual content could be partitioned for all of ground, focus, prominent element, and non-prominent material. Now, many utterances have a contextual content consisting of only a focus and a ground, or just a focus. These can also be handled by the rankings and constraints given so far.

For focus-ground contents, there will be four candidates: *GR F*, *GR*, *F*, and $\emptyset$. The constraints that play a role in determining the optimal candidate are FOCUS, GROUND, and *GROUND, the other two (PROM ELEM and *NON-PR) being violated or vacuously satisfied by all candidates.

[6]The effect of PROM ELEM in the rankings for maximal and intermediate reliance on information enrichment is to ensure that the null candidate, the empty utterance, does not end up as the optimal candidate. The same effect could be achieved, in perhaps a more transparent way, through a constraint stating that a candidate should have semantic content.

		PR. ELEM :	*NON-PR	FOCUS	*GROUND	GROUND
	a. GR P_NON-P/F	:	*		*	
	b. GR P	:		*	*	
	c. GR NON-P	* :	*	*	*	
	d. GR	* :		*	*	
	e. P_NON-P/F	:	*			*
▷	f. P	:		*		*
	g. NON-P	* :	*	*		*
	h. ∅	* :		*		*

Figure 1: Tableau for maximal reliance on information enrichment

		PR. ELEM	FOCUS	*NON-PR	GROUND	*GROUND
▷	a. GR P_NON-P/F			*		*
	b. GR P		*			*
	c. GR NON-P	*	*	*		*
	d. GR	*	*			*
	e. P_NON-P/F			*	*	
	f. P		*		*	
	g. NON-P	*	*	*	*	
	h. ∅	*	*		*	

Figure 2: Tableau for minimal reliance on information enrichment

		PR. ELEM	FOCUS	*NON-PR	*GROUND	GROUND
	a. GR P_NON-P/F			*	*	
	b. GR P		*		*	
	c. GR NON-P	*	*	*	*	
	d. GR	*	*		*	
▷	e. P_NON-P/F			*		*
	f. P		*			*
	g. NON-P	*	*	*		*
	h. ∅	*	*			*

Figure 3: Tableau for intermediate reliance on information enrichment (focus)

		PR. ELEM	*NON-PR	FOCUS	GROUND	*GROUND
	a. GR P_NON-P/F		*			*
▷	b. GR P			*		*
	c. GR NON-P	*	*	*		*
	d. GR	*		*		*
	e. P_NON-P/F		*		*	
	f. P			*	*	
	g. NON-P	*	*	*	*	
	h. ∅	*		*	*	

Figure 4: Tableau for intermediate reliance on information enrichment (prom. element and ground)

For all-focus contents there are two candidates, F and $\emptyset$, and one constraint, FOCUS, will determine the winner.

5 Determining which ranking to use

Just how is one to determine which ranking is to be used by a particular system or for a given utterance in some context? It is important to note that this will be something *outside of* the ranking. As with the design of any dialogue system, the solution lies in the answers to questions asked about various features of a dialogue system; when, where, and how is the system to be used, who is going to use it, what kind of system behaviour is desired, etc.

For information enrichment, as for many other aspects of dialogue system development, one needs studies of human-human dialogue, or experimental setups such as Wizard-of-Oz, or evaluations involving real users.

Looking in more detail at the factors identified at the beginning of section 4, for speech recognition scores, various levels need to be tried when the system is being developed, and possibly also during evaluation with real users.

Regarding the choice between a polite/formal system and an informal one, questions such as the following may need to be answered: What is best suited to the system? What do users think? Maybe what is needed is a system that comes with a choice regarding degree of formality?

When it comes to the distinction between the beginning and the rest of a dialogue, it needs to be determined, as for all factors, whether it is a useful distinction in the system, and a measure is needed for what counts as the beginning of a dialogue.

For distinguishing between naive and expert users, some form of user modelling is needed.

An example of a system used often by the same user may be a personalised system in the home used several times daily, and one used seldom a flight information system utilised by a number of different people and less frequently by each one. Determining which category the dialogue system belongs to, also determines information enriched behaviour.

Finally, adaptation to the user's level of reliance on information enrichment can be determined from linguistic studies and experiments, and through system evaluation.

All of these factors will interact with the various rankings in different ways. For instance, a situation involving intermediate level of reliance on information enrichment, say the production of a full focus although a prominent element has been determined, can be a high recognition score in conjunction with a system that is not completely informal.

6 Interference and bidirectionality

When producing utterances that rely on information enrichment, speakers need to take into account hearers' ability to construct an appropriate embedding structure, hence an appropriate contextual content, given a compositional content and the context. This becomes evident in examples like (4). Suppose that Edith's extension number is 1439, and this is what B is going to tell A. To what extent can B, in $B2$, rely on information enrichment (assuming B wants to exploit information enrichment maximally)?

(4) A1: What is your extension number?
B1: One eight three nine
A2: And what is Edith's extension number?

The contextual content of $B2 - to - be - generated$ can be paraphrased as: '*Edith's extension number is _*' as the ground, '*1_39*' the non-prominent part of the focus, and '*4*' the prominent element. Now, although it is possible for the speaker to utter only 'Four', that is, just the prominent element, this gives the hearer no chance of unambiguously recovering the contextual content. This is an example of what I call *interference*. The term is borrowed from Givón (1983), and adapted to information enrichment it involves the presence of semantically compatible contextual material that can give rise to ambiguity.

The solution to interference that can be constructed within OT is one that lends itself naturally to a dialogue context: bidirectionality (e.g., (Blutner, 2000; Jäger, 2002; Buchwald et al., 2002)). Blutner and Jäger formalise two communicative principles, one minimising hearer effort, and the other minimising speaker effort, and show the interaction of these principles using bidirectionality.

These two principles are clearly in play in dialogue involving information enrichment. For example (4) above, the optimisation of the (speaker's) output needs to be followed by an optimisation of the (hearer's) interpretation. That is, given the input form *4*, what is its optimal interpretation in the given context? Clearly, no such interpretation can be found. The optimal candidate using bidirectionality will instead be one that includes the full number, *1439*. That is, in the given context, the maximal reliance on information enrichment is not using a prominent element, but a full focus.[7]

I omit full details here, but the analysis involved will need to create all the candidate interpretations for the input *4*. Next, these candidates are evaluated with regard to the constraints, and candidates involving *4839*, *1439*, *1849*, and *1834* will be equally optimal – there is no constraint separating them.[8]

Once this has been determined, the system needs to 'back off' to a lesser degree of reliance on information enrichment. If this is intermediate reliance using a full focus – 'One four three nine' – candidate contents will be created for this. The winner from the interpretation perspective will be *1439* as a focus relying on information enrichment for its ground.

In comparison to (Blutner, 2000), my approach needs to handle the existence of several different tableaux, for instance through the 'backing off' to a lesser degree of reliance on information enrichment as just described. An alternative is to do bidirectional optimisation for all the tableaux, which gives that both 'One four three nine' and 'Edith's extension number is one four three nine' give optimal candidates. Then, other factors are used to determine which of these two candidates is to be selected.

[7]An alternative is possibly using the prominent element together with only part of the non-prominent material, as in the utterance 'One FOUR', where capitals indicate nuclear stress.

[8]These candidates are based on the assumption that in this particular context, an extension number always consists of four digits. Without this piece of information, the number of candidates will of course be even larger.

Note that the step optimising interpretation involves the utterance's being marked up for information structure. The generation step in section 4 similarly assumes that this has already been done. I presume that it is possible to determine OT contraints and rankings also for this.

7 A computational note

The OT analysis presented here is intended for a dialogue system, that is, it is intended to be implemented, so a few remarks on OT in a computational setting are in order.

Several approaches to the implementation of OT constraint checking make use of finite-state techniques. Karttunen (1998) uses an example from phonology and shows how the generation of candidates and constraint application can be composed into a single transducer, a single network. Jäger (2002) reformulates Blutner (2000) and also discusses some formal properties of bidirectional OT as outlined by the latter. Notably, Jäger discusses bidirectionality in a finite-state setting.

Now, both Karttunen and Jäger acknowledge the limitations of the OT models that can be formulated as finite-state transducers. Jäger mentions that finite-state techniques are in general too simple to handle syntax, semantics and pragmatics, so the implementation of such analyses in OT seems to be an open research question.

The OT analysis that I presented in section 4 involves a small and finite set of candidates, the constraints all involve checking whether a particular informational element is part of the candidate or not, and constraints only have at most one violation. This may mean that a finite-state implementation is possible. However, the step presupposed in section 4, the assignment of information structure (a step that is also involved in the bidirectional analysis), involves reasoning using a fairly complex information state, which is probably less likely to lend itself to a finite-state analysis.

Instead, I think that in a context such as this – the computation of information structure and the generation of information enrichment in a practical dialogue system – work could usefully be spent on making the system avoid having to create all the candidates. One possible solution is to incorporate the effect of the constraints and their ranking in the

GEN component, making GEN different for the different degrees of reliance on information enrichment, and only producing the optimal candidate in each case. Similarly, when bidirectionality is considered, the system should not produce all of, say, *4839*, *1439*, *1849*, and *1834*, but be able to determine, *a priori*, that these would be equally (un)optimal.

8 Conclusion

Five OT constraints have been introduced to handle the kinds of information enrichment discussed above. Various rankings of the constraints are needed to give different optimal candidates, and the notion of dynamic constraint reranking in the generation component of a dialogue system was introduced to model the flexibility of information enrichment. The different rankings show that the degree of reliance on information enrichment arises from, on the one hand, a conflict between generating a full focus and avoiding non-prominent material, and, on the other, a conflict in whether to generate ground or not. The need for bidirectionality in a dialogue system generating information enriched constituents was also discussed, and some computational considerations were presented. Given the theory, a subsequent step is the precise formulation of bidirectionality for information enrichment, and the implementation of the constraints and the dynamic rerankings in a dialogue system.

Acknowledgements. I would like to thank Benjamin Lyngfelt and the two anonymous reviewers for helpful comments and suggestions.

References

Arto Anttila. 1997. Deriving variation from grammar: A study of finnish genitives. In Benjamins, editor, *Variation, Change and Phonological Theory*, pages 35–68. F. Hinskens and R. van Hout and L. Wetzels.

David Beaver. 2004. The optimization of discourse anaphora. *Linguistics and Philosophy*, 27(1):3–56.

Reinhard Blutner. 2000. Some aspects of optimality in natural language interpretation. *Journal of Semantics*, 17(3):189–216.

Adam Buchwald, Oren Schwartz, Amanda Seidl, and Paul Smolensky. 2002. Recoverability optimality theory: Discourse anaphora in a bidirectional framework. In *Proceedings of Edilog 2002*.

Simon Garrod. 1999. The challenge of dialogue for theories of language processing. In Simon Garrod and Martin Pickering, editors, *Language Processing*, pages 389–415. Psychology Press.

Jonathan Ginzburg. 1999. Semantically-based ellipsis resolution with syntactic presuppositions. In Harry Bunt and Reinhard Muskens, editors, *Computing Meaning*, volume 1, pages 255–279. Kluwer.

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

Petra Hendriks and Helen de Hoop. 2001. Optimality theoretic semantics. *Linguistics and Philosophy*, 24:1–32.

Gerhard Jäger. 2002. Some notes on the formal properties of bidirectional optimality theory. *Journal of Logic, Language and Information*, 11:427–451.

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

Lauri Karttunen. 1998. The proper treatment of optimality in computational phonology. In *Proocedings of FSMNLP'98*, pages 1–12. International workshop on Finite-State Methods in Natural Language Processing, Bilkent University, Ankara, Turkey.

Alan Prince and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Technical report, Rutgers University. RuCCS Technical Report 2.

Stephen Pulman. 1997. Higher order unification and the interpretation of focus. *Linguistics and Philosophy*, 20:73–115.

Mats Rooth. 1996. Focus. In Shalom Lappin, editor, *The Handbook of Contemporary Semantic Theory*. Blackwell.

David Schlangen. 2003. *A Coherence-Based Approach to the Interpretation of Non-Sentential Utterances in Dialogue*. Ph.D. thesis, School of Informatics, University of Edinburgh.

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

Enric Vallduví. 1992. *The Informational Component*. Garland.

Henk Zeevat. 2001. The asymmetry of optimality theoretic syntax and semantics. *Journal of Semantics*, 17(3):243–262.