Towards the Elimination of Centering Theory

Rodger Kibble
Department of Computing
Goldsmiths College
University of London
R.Kibble@gold.ac.uk

Abstract

The objective of this paper is to examine whether the rules and constraints of Centering Theory (CT) can be reduced to more general principles of informativity and processing efficiency. We argue that certain OT approaches to discourse anaphora share with CT an oversimplification of the costs incurred by speakers and hearers in collaborative reference resolution, and conclude by outlining a game-theoretic model which aims to take all relevant factors into account, and allows "classic" CT to be derived as a special case.

1 Introduction

Centering Theory (Grosz et al., 1995) is a theory of local discourse structure which has been motivated on the grounds that discourse transitions defined as "coherent" are easier for hearers to process. However, discussions of the cognitive aspects of CT have tended to overlook its implications for the processing load it assumes for the *speaker*: it turns out (Kibble and Power, 2000) the task of planning sequences of clauses and ordering arguments within clauses to maximise coherent transitions soon becomes intractable. And hearers only benefit from "coherent" transitions if they maintain a discourse model enabling them to predict the occurrence of preferred transition types, which itself imposes some processing load.

These considerations suggest that firstly, an adequate account of the phenomena addressed by CT requires both speaker and hearer processes to be modelled in parallel; and secondly, if CT's rules and principles are indeed congruent with reasonable assumptions about processing costs, it may be possible to reduce them to more general "cognitive" principles. This is not a novel idea: Hasida Nagao and Miyata (1995) proposed a game-theoretic account of communication which they claimed "reduces the specifics of anaphora to general properties, such as utility and probability" and recent work by Rob van Rooy (2003) in the same spirit aims to derive the key notions in David Beaver's (2003) OT-based reconstruction of CT from Blutner's (2000) theory of rational language use, bidirectional OT. Beaver (op. cit.) reformulates CT as COT in terms of constraints which are independently motivated in the linguistics literature, and (Buchwald et al., 2002) outline a CT-like model of discourse anaphora in an OT framework (ROT) where the goal is to dispense with notions like "topic" and "[allow] the whole system to emerge from the interactions between constraints that do not depend on special-status elements".

I shall begin with a brief critical overview of these proposals, which will lead me to question whether OT with its ordinally ranked constraints and strict domination is really an appropriate framework for modelling discourse phenomena. This will be followed by an updated exposition of the game-theoretic model first proposed in (Kibble, 2003), and conclude by considering some outstanding empirical problems.

2 Centering Theory

Centering theory (CT) models the interaction of *cohesion* and *salience* in the internal organisation of a text. The basic assumptions of the theory (Grosz et al., 1995) are[1]:

> 1. For each utterance in a discourse there is precisely one entity which is the centre of attention or *center* (**Constraint 1**). The center of utterance U_n is the least oblique argument of U_{n-1} which is mentioned in U_n (**Constraint 3**).
>
> 2. The center is the entity which is most likely to be pronominalised (**Rule 1**).
>
> 3. There is a preference for consecutive utterances within a discourse segment to keep the same entity as the center, and for the Subject to be interpreted as the center and predicted as the center of the following utterance (**Rule 2**).

These principles together account for the general intuitions that in the examples below, a proper name is needed in (c′) to force the subject to be understood as *Betty*, and the variants (c″, c‴) where the pronoun has to be understood as *Betty* are unnatural.

(1) a. Alice likes Betty.

b. She often visits her for tea.

c. Yesterday she baked a cake for her.

c′. Yesterday Betty baked a cake for her.

c″. ?#Yesterday Alice baked a cake for her.

c‴. ?#Yesterday she baked a cake for Alice.

Centering has been seen as a theory of anaphora resolution but it is more accurate to see it as a model of *predictive reference resolution*. In contrast to models where a search for antecedents is triggered by anaphoric expressions, CT assumes that the process of interpreting an utterance identifies some entity as the default referent for the most salient referring expression in the following utterance, thus enabling the hearer to begin constructing a semantic representation in advance. If this position is taken by an RE which is incompatible with this "default referent" there is a disruptive effect, as in (1c‴).

[1]This is only a cursory sketch; readers should turn to the cited papers for fuller accounts.

3 Game Theory and OT

Hasida Nagao and Miyata (1995) present a reversible decision matrix (Table 1) which allows S and H (as players in a communication game) to converge on a mapping between *contents* $\{c_1, c_2\}$ and *messages* $\{m_1, m_2\}$ in contexts where each message potentially refers to either content (e.g., one is a pronoun and one a definite description). Suppose P(c_1) > P(c_2) and U(m_1) > U(m_2) where P is probability and U is utility. Expected Utility is maximised if S chooses m_1, m_2 to convey c_1, c_2 respectively and R interprets m_1 as c_1 and m_2 as c_2:

$$P(c_1)U(m_1) + P(c_2)U(m_2) \; >$$

$$P(c_2)U(m_1) + P(c_1)U(m_2)$$

(2) a. Tom was late.

b. Bob scolded him$_{Tom}$.

c. The man$_{Tom}$ was angry with him$_{Bob}$.

In the analysis of the above example it is assumed that the pronoun *him* has greater utility than the proper name, as a "lighter" expression, and that *Bob* as Subject of (2b.) has greater probability of occurring in (2c.) The claim is that this account "reduces the specifics of anaphora to general properties, such as utility and probability…" But: the analysis seems to rest on specific stipulations about utility (of pronouns) and probability (assumed to correlate with grammatical salience), and assumes that utility of messages is the same for S and R.

Robert van Rooy (2003) follows Hasida et al's approach, aiming to derive the constraints and ranking in Beaver's (2003) OT account of CT from general principles of information theory. The essential components of an OT grammar are: a set of **Constraints** which are ranked, universal and violable; **Gen** which defines all possible outputs for

		Message	
		m_1	m_2
Content	c_1	P(c_1)U(m_1)	P(c_1)U(m_2)
	c_2	P(c_2)U(m_1)	P(c_2)U(m_2)

Table 1: Decision matrix for a communication game

a given input and **Eval** which selects the optimal candidate according to **strict domination**: candidate A is optimal if there is some constraint C_i such that A has fewer violations of C_i than any other candidate and there is no higher-ranked C_j for which A has more violations than some other candidate (McCarthy, 2002). Centering-related constraints in Beaver's COT include the following:

(3)

AGREE	Anaphors agree with their antecedents.
DISJOINT	Co-arguments of a predicate have disjoint reference.
PRO-TOP	The topic is pronominalised.
FAM-DEF	Each definite NP is familiar.
COHERE	The topic of S_n is the same as the topic of S_{n-1}
ALIGN	The topic occurs in subject position.
Rank:	AGREE > DISJOINT> PRO-TOP > FAM-DEF > COHERE > ALIGN

Beaver demonstrates that his model makes the same predictions as the classic BFP algorithm (Brennan et al., 1987), with the claimed advantage of offering a declarative, reversible set of constraints which determine both the optimal linguistic forms given a semantic input, and the optimal interpretation of a given discourse. The model is extended to evaluate the coherence of multi-sentence text, with violations of each constraint summed and optimality determined in the standard way following strict domination. Thus in example (4a-b-c) the interpretation of the (c) sentence is correctly predicted as *she* = *jane*, *the young woman* = *mary*.

(4) a. Jane$_i$ likes Mary$_j$.

b. She$_i$ often visits her$_j$ for tea.

c. She$_k$ chats with the young woman for ages.

This can be verified by reading Figure 1 from left to right and at each column, retaining only candidates that have no more marks against them than any other surviving candidate: the arrow → marks the winner (optimum) $k = i, l = j$.

However, the constraint ranking given in (3) turns out to be underdetermined by the examples discussed. For instance the same optimum is chosen for (4) if one reads the table from right to left. One example in (Beaver, 2003) where the constraint ranking affects the outcome is the slightly pathological (5) which is admitted to be "difficult to process":

(5) a. Mary$_i$ likes tennis.

b. She$_i$ plays Jim$_j$ quite often.

c. He$_k$ used to play doubles with Mary$_l$.

The problem here is that both BFP and COT interpret the two occurrences of *Mary* as referring to different people (see Figure 2). As an instance of the flexibility which is a hallmark of COT, Beaver derives the correct reading by re-ranking FAM-DEF above PRO-TOP giving the full ranking

AG > DIS > FAM > PRO > COH > AL

But hard cases make bad law, and in any case this is not the only re-ranking that does the job: as the reader may verify, the preferred reading $k = j,\ l = i$ is also selected by

AG > DIS > COH > AL > FAM > PRO

To conclude, Beaver succeeds in unifying CT's heterogenous rules and constraints as an elegant set of principles which conspire to identify optimal interpretations, but it remains to be proven that the additional OT apparatus of ordinal ranking and strict domination actually takes us much further.

Rob van Rooy seeks to motivate Beaver's constraints and ranking on information-theoretic grounds: "light" expressions such as pronouns appropriately realise high-probability referents,

Example (4c)	Ag	Dis	Pro	Fam	Coh	Al
→ k = i, l = j				*		
k = j, l = i			*	*		*
k = i, l = i		*		*		

Figure 1: Optimality tableau for example (4c)

Example (5c)	Ag	Dis	Pro	Fam	Coh	Al
k = j, l = i			*			*
→ k = j, l ∉ {i, j}				*	*	
k ∉ {i, j}, l = i			*	*		*
k, l ∉ {i, j}			*	**	*	*

Figure 2: Optimality tableau for example (5c)

while sequences conforming to COHERE and ALIGN minimise violations of PRO-TOP by maximising candidates for pronominalisation and so reducing speaker's effort. However it is questionable whether the facts about extended discourse can be adequately captured in a standard OT framework, which is after all supposed to be a *grammatical* formalism. Specifying constraints in terms of quantifiable "effort" is difficult to reconcile with **strict domination**. Suppose a unit of effort is one *zipf* or ζ. Suppose the cost of violating C_i is $m\zeta$ and the cost of violating the lower-ranked C_j is $n\zeta, m > n$. Clearly, for some x, y, $xm < yn$: the effort incurred by violations of C_j would outweigh the effort saved by observing C_i.

Buchwald et al (2002) propose a bidirectional, speaker-oriented account of discourse anaphora ROT: S computes the optimal output PF for a given LF, and PFs are screened to ensure the LF is "recoverable" for H. This account can be seen to follow mainstream OT in essentially modelling a conflict between *markedness* and *faithfulness*: simpler NP forms are unmarked, according to *STRUCTURE constraints, while "faithfulness" constraints favour fuller NPs supporting feature-checking with PF over the Salience list (RECOVERABILITY), and sequences where grammatical salience in PF_i is aligned with LF_{i-1} (LINEARITY). The "ultimate preference" of the LINEARITY constraints is "for the orders of entities in SaL_{i-1} and Sal_i to align with each other" (op. cit.:41), where SaL_{i-1} and Sal_i represent S's model of the common ground before and after utterance U_i. It is not clear to me how ROT would handle example (2) above, since both LINEARITY and RECOVERABILITY favour interpreting *The man* as *Bob*, which would leave *him* to be interpreted as *Tom*, contrary to the desired reading. Space forbids a fuller discussion of this account in this short paper, but we may note that in common with other approaches we have discussed it only models benefits for H, disregards planning costs for S and the cost to H of maintaining a discourse model, and assumes a common, grammatical salience ranking is available to S and H.

4 A Framework for Centering Games

Kibble (2003) characterises reference resolution as a (timed) game where both players benefit in terms of efficient communication if H correctly guesses which item will be referred to next, or the intended reference of attenuated expressions (anaphors). So it is in S's interest to produce sequences which are as *predictable* as possible and it is in H's interest to learn what patterns S is producing *through repeated plays of the game*. According to Shannon's information theory, applied to NL semantics by e.g. Dretske (1981), predictability/high probability correlate inversely with informativity; so A's task is to *minimise* the information generated by each referring event. In a random sequence each item has equal probability. The more organised the sequence, the more items have better than chance predictability. Increasing levels of optimisation require more planning effort and memory allocation for S, while H can only take advantage of this if a record is kept of previous utterances (e.g., the previous clause or the entire discourse).

We now sketch a framework for collaborative reference resolution as a non-cooperative game of incomplete information[2], taking account of the considerations outlined above. *Non-cooperative* need not imply competition or conflict, rather that players do not overtly coordinate their strategies, and players have *incomplete information* as they do not know what strategies other players follow, perhaps even after the event. Players seek to maximise utility and minimise costs, choosing their strategies on the basis of expected utility and expectations about other players' strategies. To capture the fact that not only do S and H have conflicting preferences, but that each player needs to balance competing costs and utilities, we model dialogue participants as *multi-agent systems*, and dialogue as a multi-player game involving speaker agents text-planner **P**, RE generator **REG** and hearer agents Reference Resolver **RR** and Discourse Modeller **DM**. An optimal outcome will then be a "Nash equilibrium" (Heap and Varoufakis, 1995) where no player can improve their payoff by unilaterally changing strategy, but all references are successfully resolved.

The processing modules are schematically described below, with very selective lists of possible strategies listed in ascending (partial) order of processing effort and/or memory costs.

Speaker perspective:

1. Planner/Content Determination (P): responsible for organising input propositions into a text structure, which may already be partially ordered according to coherence relations; plan sentences by e.g. choosing verb forms to realise preferred order of arguments. Increased planning effort will increase objective *predictability* of referring events.

Random: Do nothing, resulting in a random order of propositions and arguments within clauses.

Salient-Arg: Promote arguments within a clause according to perceptual salience.

Align: Plan consecutive clauses to align salience rankings.

Global: Plan sequence of clauses to maximise referential continuity, in addition to salience alignment.

2. Realiser (REG): generates appropriate referring expression to denote arguments of predicates.

Null: Do (almost) nothing: $\emptyset$ or personal pronoun

Short-NP: Reduced definite using a basic-level predicate, such as *the dog*

Full-NP: Uniquely identifying definite *the small white poodle in the kennel*

Hearer perspective:

1. Discourse Modeller: maintains a record of entities mentioned in the discourse which will be candidates for anaphora resolution:

$\emptyset$: no record of discourse referents. Only explicit definites will be resolvable.

1-Cl Keep a one-clause buffer of focal referents

Center Model: retain salience-ranked list from previous utterance U_{n-1} marking which was most salient in U_{n-2}; propose default referent for Subj(U_{n+1})

Full DM: fully-structured discourse model including salience-ranked history list, segmented to reflect discourse structure.

2. Reference Resolver: responsible for identifying referent of REs with an entity in the domain.

Default: Resolve to referent predicted by **DM** (if available)

Search-DM: Search the discourse model constructed by **DM**

Search-Dom: Search the domain if no candidate found in the discourse model

Repair: Recovery strategies: reread the text; internally "replay" speech or query the speaker.

It should be clear from the preceding discussion that the more effort **P** expends in constructing coherent plans, the more opportunities there will be for **REG** to use reduced REs, and that **RR**'s success in interpreting these with least effort will depend on the work done by **DM**. We can visualise these modules as a set of interlocking cogwheels: as **P** and **DM** turn in one direction, **REG** and **RR** will turn in the other.

[2] A more detailed account with worked examples can be found in (Kibble, 2003; Kibble, fthc).

Anaphora in the Centering Game

Following Shannon, the quantity of *information generated* by an event (or *surprisal*) is a function of its *probability* in context: $I(e) = \log_2 1/p(e)$. So for instance if all events $e \in E$ are equally likely then for any e, $I(e) = \log_2 1/p(e) = \log_2 \mid E \mid$, or the number of binary decisions needed to reduce E to $\{e\}$. Consider the schematic example *A B C; A C; ? D E*, where A, B, etc are distinct referents:

$$\text{Assume: } p(? = A) > p(? = C) > p(? = B) >$$

$$p(? = D) = p(? = E);$$

$$\text{then } I(A) < I(C) < I(B) \ldots$$

So equating ? to A will have the smallest "inferential cost" for H. S saves effort on Referring Expression Generation (**REG**) by planning sequences to *minimise* information generated by referential events, allowing for use of non-informative or attenuated expressions (pronouns, null string). Resolving anaphors requires **Hearer** to perform non-monotonic inference. The preferred case is where this involves the smallest possible quantity of information. The optimal state is where a referent d is realised as an anaphor in clause C only if there is no other d' s.t. $I(d') \leq I(d)$ and d' is not also realised as an anaphor in C: in other words the "lightest" RE must pick out the most predictable referent. This is quite similar to Centering Theory's Rule 1. And CT's Rule 2 can be reconstructed as follows: assuming that S and H converge on a scheme of text organisation which maximises predictability of reference, we may expect a grammatically-based salience ordering to emerge along with a preference for continuity of topics. It would follow that the rules and constraints stipulated by CT can in principle be eliminated in favour of general principles of rational communication, as we surmised earlier, though we may retain the terminology of CT for descriptive convenience.

We can make this claim more precise: the interaction of the modules listed above defines a 4-dimensional solution space, which is partially mapped out in Figure 3. Each point marked $\mathbf{O}_n$ identifies a partial equilibrium (solution) where

Planning/ Modelling	Modeller	1 ∅	2 1-Cl	3 CM
Planner	1. Random	$\mathbf{O}_1$		
	2. Align		$\mathbf{O}_2$	
	3. Global			$\mathbf{O}_3$

Planning/ RE Gen	Realiser	1 Null	2 Short-NP	3 Full NP
Planner	1. Random			$\mathbf{O}_4$
	2. Align		$\mathbf{O}_5$	
	3. Global	$\mathbf{O}_6$		

Modelling/ Resolution	Resolver	1 Def	2 S-DM	3 S-Dom
Modeller	1 ∅			$\mathbf{O}_7$
	2 1-Cl		$\mathbf{O}_8$	
	3 CM	$\mathbf{O}_9$		

RE Gen Resolution	Resolver	1 Def	2 S-DM	3 S-Dom
Realiser	1 Null	$\mathbf{O}_{10}$		
	2 Short-NP		$\mathbf{O}_{11}$	
	3 Full-NP			$\mathbf{O}_{12}$

Figure 3: Aspects of a Referring Game

neither player can unilaterally increase their utility by moving along the relevant axis. In terms of coordinates $\langle P,\ REG,\ DM,\ RR \rangle$, Centering Theory in its classic form assumes that discourse operates at a point close to $\langle 3, 1, 3, 1 \rangle$: speakers and hearers are assumed to be ideally rational cooperative agents who put the maximum effort into high-level tasks of planning and modelling so they they can each minimise their efforts on the low-level tasks of realisation and reference resolution. A dialogue involving uncooperative speakers, who do not plan ahead or align the order of nominal arguments with salience in the common ground, is represented by $\langle 1, 3, 1, 3 \rangle$. An intermediate state where speakers and hearers are prepared to be cooperative but have conflicting claims on their time and attention (*bounded rationality*) might fall somewhere near $\langle 2, 2, 2, 2 \rangle$.

There are two ways of looking at this model.

One way is to see it as a hypothesis about how two rational agents might co-evolve a set of strategies which enable them to communicate efficiently, over repeated plays of a communication game. For instance Agent A might start out by using grammatical salience to minimise the randomness of his referring actions and Agent B would try different salience rankings until she finds one that provides the most reliable predictions about what A will refer to next. Agent B may notice that A is using pronouns and reduced definite NPs, so it is worth her while to keep a record of what has been referred to in order to resolve anaphors. And so on. On the other hand we could see the various solutions as *conventions* in the sense of (Lewis, 1969). A strategy which has successfully solved a coordination problem gives rise to expectations that this strategy will be followed in future and thus becomes a precedent; A and B will follow the strategies they have co-evolved when talking to C rather than beginning a new process of mutual adaptation.

5 Some Empirical Issues

The above reconstruction of CT rests on the assumption that S and H will converge on a salience ranking corresponding to grammatical obliqueness; however this assumption is challenged by results discussed by Oberlander (1998) suggesting that S and H have distinct preferences, with S following a grammatical hierarchy when choosing the topic of consecutive utterances while H seems influenced by thematic roles in forming expectations of upcoming referring events.

(6) a. John gave the book to Bill so he$_{John}$ ...

b. John gave the book to Bill so John ...

c. John gave the book to Bill so he$_{Bill}$ ...

d. John gave the book to Bill so Bill ...

In continuation experiments, speakers prefer type (a) when talking about John and type (d) when talking about Bill, confirming CT's predictions. But in reading-time experiments, type (c) sentences were processed faster than type (a), indicating an effect of *thematic roles*. Since it has so far proved difficult to reliably identify thematic roles in naturally occurring text, the type of evidence that can be brought to bear on this question is rather limited and so I will regard "Oberlander's paradox" as a puzzle for CT rather than a stumbling block.

Assuming the mismatch is real, it points to a deeper assumption: that S and H will both arrive at the same subjective probability assessment of the distribution of referents in a discourse and that both will agree with the *objective* probability which could be determined by corpus analysis. Reasons why S and H may differ in their assessments are firstly that owing to resource constraints, each may only consider a small part of the discourse history and secondly, they may operate at different levels of representation. For instance: if P plans the sequence of referents in U_i according to their salience in the previously generated utterance U_{i-1}, the *syntactic* structure of U_{i-1} is likely to be more accessible to P than its semantic representation. However when DM calculates the expected referents in U_i according to salience in the previously interpreted utterance U_{i-1} it is more likely that the *semantic* representation of U_{i-1} will be retained and salience for H will depend on thematic rather than grammatical roles. We assume therefore that it is more costly for S to retrieve the underlying semantics of the immediately preceding utterance, and for H to reconstruct the syntactic structure. These operations, though costly, are available for repair strategies in case of failure of anaphora resolution. This hypothesis is consistent with the differing preferences of S and H noted by Oberlander (1998), while its formulation relies on the game-theory framework we propose. The question remains however, why H would continue to operate with a salience ranking which seems orthogonal to the ranking actually used by S. What follows is a conjecture: it may be reasonable to assume that H learns a thematically-based salience ranking as long as there is a significant positive correlation between the distributions of grammatical and thematic roles, so that predictions based on thematic roles have a better than chance probability of being correct. Again, evaluation of this conjecture will require progress in the reliable annotation of thematic roles in text.

There is also an assumption underlying CT that S and H share a model of the common ground: S will only make grammatically salient reference to entities which are known to be salient to H. This assumption is undermined by the findings of (Branigan et al., 2003) who conducted experiments using task-based dialogues between pairs of subjects, where the knowledge states of the speaker and addressee were manipulated independently. They found that speakers tended to reproduce a Given-New ordering in their utterances, but only with respect to the speaker's knowledge: items which were New to the hearer would be referenced more saliently than hearer-Given items if they were Given for the speaker. Branigan et al. are cautious about drawing more general lessons from these experiments owing to the relatively simple nature of the hearer's task. However there are two points worth considering in relation to the model outlined above: firstly "salience" has to take account of the visual field and other perceptual sources as well as the discourse history; secondly, this is further evidence that speakers and hearers can have misaligned internal models, and yet communication proceeds. The framework I have outlined allows these mismatches to be modelled, and the study of these kinds of asymmetry may turn out to be more revealing in the long run in the study of human communication than adherence to theories which assume perfect alignment.

6 Acknowledgements

This work has been presented, in various stages of development, at the ITRI Seminar, University of Brighton; the 5th International Workshop on Computational Semantics, University of Tilburg; and the ICCS Seminar, University of Edinburgh. I'm grateful to the participants in these events for much useful feedback.

References

David Beaver. 2003. The optimization of discourse anaphora. *Linguistics and Philosophy*. To appear.

R. Blutner. 2000. Some Aspects of Optimality in Natural Language Interpretation. *Journal of Semantics*, 17:189–216.

Holly Branigan, Janet McLean, and Hannah Reeve. 2003. Something Old, Something New: Addressee Knowledge and the Given-New Contract. In *Proceedings of CogSci 2003*, Boston.

Susan Brennan, Marilyn Walker Friedman, and Carl Pollard. 1987. A centering approach to pronouns. In *Proceedings of 25th ACL.*

Adam Buchwald, Oren Schwartz, Amanda Seidl, and Paul Smolensky. 2002. Recoverability Optimality Theory: Discourse Anaphora in a Bi-directional Framework. In *EDILOG 2002: Proceedings of the sixth workshop on the semantics and pragmatics of dialogue*.

Fred Dretske. 1981. *Knowledge and the Flow of Information*. MIT Press. Reissued by CSLI Publications, Stanford, 1999.

Barbara Grosz, Aravind Joshi, and Scott Weinstein. 1995. Centering: a framework for modelling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

K. Hasida, K. Nagao, and T. Miyata. 1995. A game-theoretic account of collaboration on communication. In *Proceedings of the First International Conference on Multi-Agent Systems*.

Shaun P. Hargreaves Heap and Yanis Varoufakis. 1995. *Game Theory: A Critical Introduction*. Routledge, London and New York.

Rodger Kibble and Richard Power. 2000. An integrated framework for text planning and pronominalisation. In *Proceedings 1st International Natural Language Generation Conference*, Mitzpe Ramon, Israel.

Rodger Kibble. 2003. Both sides now: Predictive reference resolution in generation and interpretation. In *Proceedings of the 5th International Workshop on Computational Semantics*, Tilburg.

Rodger Kibble. fthc. Centering Games. MS, in preparation. Department of Computing, Goldsmiths College, University of London.

David Lewis. 1969. *Convention: A Philosophical Study*. Blackwell, Oxford.

John J. McCarthy. 2002. *A Thematic Guide to Optimality Theory*. Cambridge.

John Oberlander. 1998. Do the right thing. *Computational Linguistics*, 24(3):501–507.

Robert van Rooy. 2003. Relevance and Bidirectional OT. In R. Blutner and H. Zeevat, editors, *Pragmatics in Optimality Theory*.