The Mental Salience Framework: Context-adequate generation of referring expressions

Christian Chiarcos

Abstract

This paper describes a general architecture for mechanisms of attention control in discourse. The proposal is based on a generalized notion of salience that abstracts from existing, theory-specific definitions of entity salience. I consider two dimensions of salience, hearer salience indicating the *status quo* or "givenness" of an entity, and speaker salience underlying the attempts of the speaker to manipulate this status. This framework is applied to three phenomena in the encoding of discourse referents, choice of referring expressions, word-order preferences, and assignment of grammatical roles. The adequacy of this proposal is illustrated by providing reconstructions of two theories, Givón's topicality approach and different instantiations of Centering. The proof for Centering-adequacy is sketched, and the framework is compared to related proposals, with particular consideration of the optimality-theoretic Centering reconstruction by Beaver (2004).

As a result, a parameterized architecture for modeling linguistic variability in discourse is presented. It provides a powerful, simple and intuitive mechanism to integrate cognitive-pragmatic aspects of coding preferences in the field of natural language generation (NLG).

1 Mechanisms of attention control

It had been noticed very early that the choice among syntactically well-formed expressions is by no means determined by semantic constraints alone. Consider the following classical text (Grosz et al. 1995, ex. (5), shortened) with possible truth-semantically equivalent variations as illustrated in (4'), (1') and (5'):

(1) Terry$_{te}$ really goofs sometimes.

(2) He$_{te}$ wanted Tony$_{to}$ to join him on a sailing expedition.

(3) He$_{te}$ called him$_{to}$ at 6AM.

(4) Tony$_{to}$ was sick and furious at being woken up so early.

(5) He$_{to}$ told Terry$_{te}$ to get lost and hung up.

(6) Of course, Terry$_{te}$ hadn't intended to upset Tony$_{to}$.

Here, I concentrate on referring expressions, in particular on three interacting levels of variability. Note, that for this exemple text, the textual function of the alternatives is roughly identical to that in the original examples.

choice of referring expressions (REF)

(4') **This guy**$_{to}$ was sick and furious at being woken up so early.

assignment of word-order preferences (WO) e.g. topicalization

(1') **Sometimes,** Terry$_{te}$ really goofs.

assignment of grammatical roles (GR) e.g. passive vs. active clauses

(5') Terry$_{te}$ **was told** to get lost and **Tony**$_{to}$ hung up.

Taking up popular assumptions among functional linguists, I regard these grammatical devices to serve (at least partly) as means a speaker can use to guide the hearer's flow of attention in discourse (Chafe, 1976; Tomlin, 1995; *inter alia*).

The flow of attention is a key mechanism controlling any kind of mental activity: The world surrounding us (and even our internal world) is far too rich to be realized, understood, or described as a whole. Rather, just relevant or especially significant elements are chosen to build up a finite symbolic representation describing the situation sufficiently but sparsely enough to be held in mind or to be communicated.

In this view, attention selects only a small subset of the information to be processed (Chafe 1976, "center of attention"), but shifts rapidly across the scene. Thus, complex representations arise not from the current center of attention alone, but from the sequence of attention shifts as well. Applied to text production, a speaker needs to make sure that the hearer's center (or "focus") of attention moves along the lines he had in mind. Otherwise, the hearer cannot obtain the mental representation the speaker wants him to construct. To prevent such a failure of communication, the speaker has to be aware of the hearer's state of mind and of the effects a given utterance might have on the hearer's model of discourse.

Adopting a functional perspective, iconic form-function mappings between mental states and grammatical devices are assumed as a basis of a general framework for mechanisms of attention control in discourse. However, these correlations have remained notoriously vague, which prohibits their practical application in the field of natural language generation (NLG). To overcome this problem, I introduce SALIENCE as a cover term of properties of mental states such as "discourse prominence" (Pustet 1997), "activation" (Chafe 1976), "topicality" (Givón 1983a), etc. that have defined in an abstract manner only.

I adopt a definition from the field of visual attention control (Koch and Itti 2000): Salience is a situation-bound, dynamic property of entities within a mental model. Opposed to this, attention is a binary property of a selected sub-set of entities, but it tends to be attracted by a high degree of salience. Thus, attention is an epiphenomenon of salience.[1] Depending on the salience-induced topological structure (ranking, order) over the entities within a (discourse) model, coding preferences are assigned.

2 Salience in discourse

2.1 A generalized conception of salience

In linguistics, psychology, artificial intelligence and neighboring fields, different (and partly contradictive) traditions using the notion of salience evolved during the last 30 years. Two extreme bounds in the usage of this term can be seen in the discussion of focal prosody (e.g. Davis and Hirschberg, 1988) and in the discussion of referential accessibility in discourse (e.g. Sgall et al, 1986).

> Pitch accents mark items as intonationally prominent and convey the relative 'newness' or 'salience' of items in the discourse. (Davis and Hirschberg 1988)

> ... salience, [i.e.] foregrounding, or relative **activation** (in the sense of being immediately 'given', i.e. accessible in memory). (Sgall et al. 1986, p.54f.)

I take these examples to be prototypes of two different dimensions of salience corresponding to the two most elementary perspectives on information intended to be uttered:

speaker salience (importance/newsworthiness) Speaker salient information is speaker-private and relevant, e.g. new for the hearer, not predictable or something the speaker wants to put special emphasis on.

hearer salience (accessibility/givenness) Hearer salient information is known and easily retrievable for the hearer.

[1] The original definition as used by Koch and Itti is based on visual fields of neurons, i.e. on *areas* within a scene, not entities. Though it seems that a similar dynamic notion of salience is appropriate on a higher level of abstraction as well, researchers on the interface of visual and linguistic salience modeled it in terms of size and absolute position (Kelleher and van Genabith 2004), denying the possibility of shifts of attention at all. However, it seems to be generally accepted that this static notion of salience is a heuristic approximation only, a generalization over a longer period of time describing the likelihood of an area to be salient.

Successful communication crucially depends on the availability of both perspectives to a speaker. The aspect of speaker salience or importance is a necessary pre-condition for any conversation, as it covers a speaker's motivations to produce an utterance. Besides this, if a speaker aims to produce text that is directed to the hearer, s/he must have some ideas about the hearer's current attentional state, i.e. hearer salience.

Assume there is an idealized scenario where the speaker has no special information about the hearer's state of mind. Then, we can characterize both dimensions of salience as follows:

attention control Hearer salience reflects the current *status quo*, e.g. the attentional state assigned to a discourse referent. Speaker salience arises from intentions to manipulate this state.

intentionality Speaker salience is induced by the intentions a speaker has. As no specific assumptions on the intentions of the hearers are available, hearer salience depends on contextual information available to both speaker and hearer, thus, it is a property of the common ground between them.

temporal scope Due to the lack of additional information, hearer salience must be approximated from situational factors, world knowledge and the previous discourse, thus, it is "backward-looking". As opposed to this, speaker salience or the underlying intentions affects the planning of the further discourse. So, it can be estimated heuristically from properties of the forthcoming discourse. In this respect, speaker salience is "forward-looking".

stimulus-dependence As speaker salience arises from intentional states, it can be independent of the current situational context, whereas hearer salience is stimulus-induced.

Previously, similar classifications have been proposed:

cognitive vs. surface-based Pattabhiraman (1992) introduced a distinction between "canonical salience" as a property of surface forms and "instantial salience" of cognitive concepts. He presented an algorithm for the assignment of grammatical devices such that the canonical salience of the resulting expression corresponds to the instantial salience of the underlying concept as much as possible. As a consequence, canonical salience of surface forms uttered before (i.e. *encoded* instantial salience) and instantial salience of cognitive concepts to be encoded in the forthcoming discourse can be distinguished where the former is available to both hearer and speaker but the latter is private to the speaker alone. However, he did not explore the implications of this distinction for communication in general but concentrated on the production perspective only.

perspective In their application of the Praguian model of salience onto dialogue, Hajičová et al. (1998) proposed a distinction between two different knowledge stocks (discourse models): The individual stock of dialogue knowledge (ISDK) and the shared stock of dialogue knowledge (SSDK). Accordingly, the activation or salience of entities in the SSDK is based on the common ground between the discourse participants, whereas the "activation degrees of entities in the ISDK depend on the participant's own attention, dialogue intentions, etc." (p.386). Thus, it is possible to account for different uses of referring expressions according to perspectives of different discourse participants.

salience indication vs. salience guidance Navaretta (2002) identified two components of salience affecting the interpretation of pronominal anaphors in Danish: givenness and explicit salience marking of antecedent (cf. the dichotomy of "inherent salience" and "imposed salience" by Mulkern (2003)). While givenness (or accessibility) derives from discourse factors such as frequency and previous mention, explicit salience marking can be used to boost the salience of a referent that is not sufficiently given. Accordingly, two functions of grammatical devices can be distinguished: *indication* (by using canonical constructions when referring to a referent) and *guidance* (e.g., foregrounding by means of marked constructions) of the inherent salience, i.e. the attentional state a referent has for the hearer.

Comparable distinctions between different types of salience can be found in psychology, too.

In her classical work on emotional focus, Nissenbaum (1985) identified two dimensions of salience (of emotions),

markedness	−	+
REF	pronoun	full description
WO	left-peripheral	right-peripheral
GR	subject	non-subject
salience	+	−

Figure 1: Simplified markedness hierarchies and salience.

1. salience$_1$ induced by long-term experiences beyond the retrievable situational context ("active salience"), and
2. salience$_2$ arising as a reaction on environmental factors ("passive salience") .

Salience$_1$ is stimulus-independent and spontaneous, imagine a person thinking about about a person (s)he loves – (s)he does so without necessarily seeing the object of her/his desire. As opposed to this, salience$_2$ is stimulus-induced, e.g. a burglar being afraid of a dog snapping at him. Whereas the occurrence of thoughts or utterances expressing the respective emotions is predictable from situational context for an observer in the second case, it is not in the first.

From the perspective of an emoter, thoughts/utterances arising from salience$_1$ are private, while thoughts/utterances arising from salience$_2$ arise from the common ground between the emoter and other individuals partaking in the same situation.

Following a general psychological conception, I consider salience here to be not necessarily a property of linguistic expressions and the perceived environment, but to be a general cognitive conception, thus a matter of the perception of situational factors, of the interpretation of linguistic cues, and of the mental representation of intentional and emotional states. Especially, salience is a necessary condition for shifts of attention as described above.

2.2 Phenomenology

Here, I focus on three phenomena: Choice of referring expressions (REF), assignment of grammatical roles (GR) and word order effects (WO).

Defining salience in terms of givenness or accessibility, it has been frequently remarked that the more salient a discourse referent is, the less complex, semantically rich and emphasized referring expressions are expected to encode it (Ariel 1990). Especially, pronouns are expected to denote more salient referents than full descriptions. Similarly, salience rankings of grammatical roles (Grosz et al. 1995; Givón 2001) follow a conventional hierarchy of markedness with subject (nominative case) being more frequent and less phonologically complex than direct object, etc. Following Givón (1995, p.25-69), markedness is defined in terms of complexity and non-conventionality (inverse frequency). So, for REF and GR, an iconic mapping can be assumed correlating surface or empirical measures of markedness with the underlying degree of relative salience:

The more salient a discourse referent is, the less marked it is expected to be encoded.

The third dimension under consideration is word order. Both assignment of grammatical roles and choice of referring expressions are entangled with word order preferences. Generally, less complex forms (e.g. pronouns) tend to precede more complex forms (Hawkins 1992), and subjects tend to precede other grammatical roles (Greenberg 1963). To integrate these tendencies into the iconicity principle, a gradient increase of markedness (and a decrease of underlying salience) is assumed along with the sequential order of elements within a clause from left to right (Sgall et al. 1986). With this hypothesis, a unified model for the planning of the choice of referring expressions, grammatical roles and word-order preferences can be developed relying on iconic mappings as illustrated in Fig. 1.

The notion of salience as defined here involves two different temporal or participant perspectives, and indeed, similar claims on multi-dimensionality have been made for the levels of referring expressions

(cf. deviations from iconic mapping (Ariel 1990, p.191ff.)) and grammatical roles (cf. the aspects of "discourse prominence" (Pustet 1997): indicating a horizon and fore-/backgrounding). With respect to word-order preferences, multiple factors have been considered related to both dimensions of salience. On the one hand, it is claimed that given elements tend to precede new elements (Sgall et al. 1986), on the other one, it has been shown that – at least for German – this tendency is not absolute (Weber and Müller 2004). Instead, Strube and Hahn (1999) suggested that relative ordering has an effect on the accessibility in forthcoming discourse, thus it is a device of attention guiding similar to grammatical roles .

2.3 A general framework

I suggest that from the interaction of two dimensions of salience, coding preferences can be deduced. A generalized framework is sketched that allows for the reconstruction of two major theories of referential coherence in discourse: Givón's 1983a; 2001 topicality approach, and Centering Theory (Grosz et al. 1995). Both approaches distinguish two perspectives on discourse, a backward-looking/anaphoric aspect on the one hand, and a forward-looking/cataphoric aspect on the other hand, that can be related to hearer salience and speaker salience respectively.

Generalizing over the observations made in the last two sections, I propose the following characteristics for an operationalizable framework of attention control and referential coherence in discourse:

- Salience induces a ranking over entities within a mental models, e.g., a discourse model. Here, I distinguish
  - *hearer salience*, i.e., the degree of attention/prominence a speaker assumes that a hearer assigns to a given discourse entity, and
  - *speaker salience*, i.e., the degree of attention/prominence/emphasis a speaker puts on an entity.
- The prototypical function of hearer salience is the indication of the (assumed) degree of attention a referent is assigned according to situation and previous discourse.
- The prototypical function of speaker salience is to announce shifts of attention, thus it is sensitive to speaker-private knowledge and properties of the *forthcoming* discourse.
- Hearer salience and speaker salience interact and are mapped iconically onto grammatical devices according to underlying markedness hierarchies, cf. Pattabhiraman's (1992) mapping from instantial onto canonical salience.
- The parameters, hearer salience, speaker salience, and deduced coding preferences can be represented by numerical scores, with salience scores and coding preferences defined as normalization of the weighted sum (linear combination) of parameter values respectively of hearer and speaker salience. Then, different configurations can be implemented by assignment of weights.
- Both grammar-dependent parameter values and preference deduction are based on markedness hierarchies.

For presentational purposes, salience scores are modeled as real numbers from the scale [0:1] with 0 encoding the lowest degree of salience, and 1 the highest degree. As general form for salience scores, the following standard representation is proposed:

$$sal(r) = \frac{1}{1 + \sum_i w_i x_i(r)} \tag{1}$$

In eq. (1), x_i denotes the value of the i-th salience factor for the referring expression r in the actual utterance and w_i the corresponding weighting.

This set of assumptions constitutes the Mental Salience Framework.

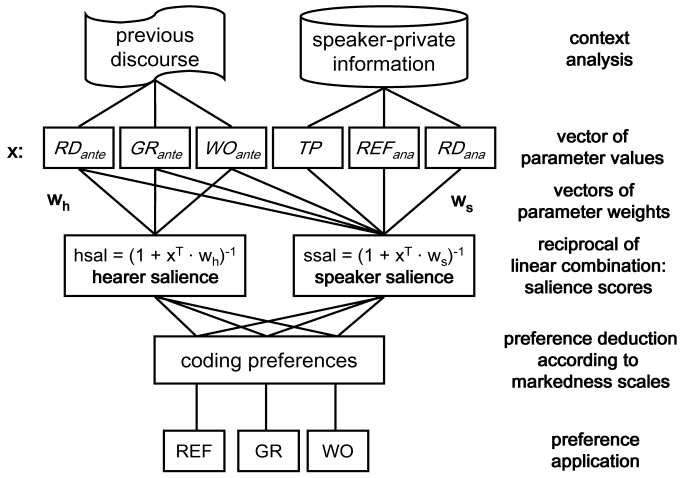

Figure 2: A minimal parameterized framework, schematically.

3 Adequacy

The proposal sketched in 2.3 is a generalization over a number of existing models of salience Thus, I do not claim the framework to be *cognitively* valid, but just to be adequate with respect to the underlying theories. This adequacy claim is justified by the reconstruction of two frameworks of referential coherence, Centering and Topicality.

3.1 Reconstructing Centering Theory

Centering Theory is a model of local discourse coherence that defines relationships between centers ($\sim$ referring expressions) in subsequent utterances, often applied with special emphasis on its effect on pronominalization and anaphora resolution.

In Canonical Centering Theory (CCT) (Grosz et al. 1995), the entities of an utterance U_n constitute the set of *forward-looking centers* $C_f(U_n)$ that subsumes possible antecedents for anaphoric references in the forthcoming discourse. The forward-looking centers are ordered according to their relative salience as encoded by their grammatical roles (SBJ $>$ DIR-OBJ $>$ INDIR-OBJ $>$ OTHER). The *backward-looking center* $C_b(U_{n+1})$ of the following utterance U_{n+1} is then defined as the highest-ranked entity from $C_f(U_n)$. Centering Theory posits a weak constraint on the usability of pronouns: If a pronoun occurs in U_{n+1}, then $C_b(U_n + 1)$ must be pronominalized, too ("Rule 1"). Further, it states a preference for transitions between utterances based upon salience characteristics in U_n and U_{n+1} ("Rule 2"): keep the same entity as backward-looking center in both U_n and U_{n+1}, interpret the subject of U_n as the backward-looking center of U_n, and keep the subject ("preferred center") of U_n as the backward-looking center of the following utterance U_{n+1} (Kibble 2003).

Consequently, two tracks of salience are distinguished: Salience of potential backward-looking centers resulting from the assignment of grammatical roles in the *preceding* utterance, and salience of forward-looking centers as expressed by the assignment of grammatical roles in the *actual* utterance. The dichotomy of two types of "centers" follows a similar criterion of temporal scope as the distinction between hearer salience and speaker salience as introduced above.

Hearer salience $hsal_{CCT}(r)$ of a referring expression r in utterance U_n (i.e. salience of r as a potential backward-looking center) can be modeled as the relative grade of the grammatical role of the antecedent of

r (GR_{ante}, cf. Fig. 3) if it occurred in the directly preceding utterance U_{n-1}. Accordingly, speaker salience $ssal_{CCT}(r)$ should be predictable from pronominal references to r in the directly following utterance U_{n+1} (REF_{ana}).

The restriction of the canonical model to relations between directly neighboring sentences seems to be unnatural, so it was suggested that more distant utterances contribute to the salience of a discourse referent, but to a lower degree than the last utterance. This assumption yields Left-Right-Centering (LRCT) (Tetreault 1999).

To model hearer salience and speaker salience respectively, a measurement of distance must be integrated explicitly. Referential distance is defined as the number of clauses between an antecedent and an anaphor (Givón 1983a), RD_{ante} is the referential distance of an anaphoric link with the anaphor in U_n, RD_{ana} is the distance of a link whose antecedent is in U_n.

Using this definition, hearer salience and speaker salience for LRCT can be modeled as follows:[2]

$$hsal_{LRCT}(r) = \frac{1}{1 + RD_{ante}(r) + (1 - GR_{ante}(r))} \tag{2}$$

$$ssal_{LRCT}(r) = \frac{1}{1 + RD_{ana}(r) + (1 - REF_{ana}(r))} \tag{3}$$

As demanded in the preceding section, the outcome is normalized with 1 denoting the highest possible salience score. If no antecedent (anaphor) exists, RD_{ante} (RD_{ana}) is infinite, thus $hsal_{LRCT}$ ($ssal_{LRCT}$) converges against 0. For Canonical Centering, the locality constraint can be implemented by replacing RD_{ante} with $1/\lfloor 1/RD_{ante} \rfloor$.

As an alternative to the canonical salience ranking, in Functional Centering (Strube and Hahn 1996, 1999), the ordering of potential backward-looking centers is replaced by a ranking based on information status, embedding depth and relative word-order. Following Rambow's (1993) account on Centering and word order in German, I concentrate on word order as a determinant of the ranking of forward-looking centers (WO_{ante}).[3]

$$hsal_{WO}(r) = \frac{1}{1 + RD_{ante}(r) + WO_{ante}(r)} \tag{4}$$

Again, speaker salience (i.e. salience of forward-looking centers) can be approximated by the choice of referring expressions for an anaphor in the following utterance (REF_{ana}). Then, grammatical roles and word order are predicted from the relative ranking of discourse referents according to their speaker salience, whereas referring expressions are predicted from hearer salience directly.

However, as the pronominalization rule (Rule 1) of Centering is underspecified with respect to coding decisions, a stronger formulation is needed for practical application in NLG (Kibble and Power 2000). As referential distance lower than 1 is a necessary (but no sufficient) condition for the use of pronouns in CCT, a pronominalization threshold of 0.5 is suggested as a first approximation, i.e. if $hsal(r) > 0.5$, use a pronoun, unless this is prohibited by ambiguity of reference or a higher-ranked referent has been encoded as a full description already, otherwise, use a full description.

3.2 Reconstructing Topicality

Following Givón (2001), topicality is a cognitive dimension that has to do with attention control mechanisms and discourse prominence. The two functional dimensions underlying topicality are *anaphoric* ($\sim$ givenness) and *cataphoric* topicality ("importance"). Heuristically, anaphoric topicality is approximated by referential distance (RD), whereas cataphoric topicality can be approximated by topic persistence (TP), i.e. the number of mentions of the referent in the subsequent (up to) 20 clauses.

[2] In this formalization, referential distance is the most influential factor on salience (step-width is 1), with GR_{ante} and REF_{ana} providing minor distinctions among cases with equal distance ($(1 - GR_{ante}) < 1$, $(1 - REF_{ana}) < 1$).

[3] Note that this covers only one of Strube and Hahn's (1999) original salience ranking determinants. However, it is generally assumed that word order in German (with the possible exception of the *vorfeld*), also reflects information status (Kruijff et al. 2001), which is consistent with the simplification of Functional Centering elaborated here. For the sake of clarity, however, the abbreviation concerns the reconstruction, but not Strube and Hahn's (1999) original proposal.

Consider an utterance U_n and a referring expression r with antecedent q in a preceding utterance U_k and anaphor s in a subsequent utterance U_l ($k < n < l$).

properties of antecedent

$$RD_{ante}(r) = \begin{cases} \infty & \text{iff. } r \text{ has no antecedent} \\ n-k-1 & \text{else} \end{cases}$$

$$GR_{ante}(r) = \begin{cases} 0 & \text{iff. } r \text{ has no antecedent} \\ 1 & \text{iff. } q \text{ is subject} \\ 0.9 & \text{iff. } q \text{ is direct object} \\ 0.8 & \text{iff. } q \text{ is indirect object} \\ 0.7 & \text{else} \end{cases}$$

$$WO_{ante}(r) = \begin{cases} 0 & \text{iff. no antecedent} \\ \frac{\#\text{words in } U_k \text{ before } q}{\#\text{words in } U_k - \#\text{words in } q} & \text{else} \end{cases}$$

properties of anaphor(s)

$$RD_{ana}(r) = \begin{cases} \infty & \text{iff. no anaphor to } r \text{ exists} \\ l-n-1 & \text{else} \end{cases}$$

$$REF_{ana}(r) = \begin{cases} 0 & \text{iff. } r \text{ has no anaphor} \\ 1.0 & \text{iff. } s \text{ is pronominal} \\ 0.5 & \text{else (i.e. } s \text{ is a full description)} \end{cases}$$

$$TP(r) = \frac{\#\text{mentions of } r \text{ within the next 20 utterances}}{20}$$

Figure 3: Parameters considered

parameters	weights for hearer salience			weights for speaker salience	
	LRCT	WO	TOP	LRCT/WO	TOP
RD_{ante}	1	1	1		
$1 - GR_{ante}$	1	0	0		
WO_{ante}	0	1	0		
RD_{ana}				1	0
$1 - REF_{ana}$				1	0
TP				0	1

Figure 4: Reconstructing Left-Right-Centering (LRCT), Centering with word-order-based salience ranking of forward-looking centers (WO), and Topicality (TOP).

In the reconstruction, hearer salience is equated with anaphoric topicality, with referential distance as its only factor, whereas speaker salience is equated with cataphoric topicality, with topic persistence as its only factor. Topic persistence is normalized.

$$hsal_{TOP}(r) = \frac{1}{1 + RD_{ante}(r)} \quad (5)$$

$$ssal_{TOP}(r) = \frac{1}{1 + \big(1 - TP(r)\big)} \quad (6)$$

The interaction of both dimensions on the preference deduction layer seems to be complex but is not explicitly described. Instead, Givón argues that both dimensions of topicality form one single and homogeneous dimension of topicality, illustrating effects by revealing correlations between grammatical devices and topicality *measures* directly. Here, hearer salience is taken as the main determinant of REF (a strong correlation of pronominalization with referential distance has been proven), speaker salience is taken as the only determinant of WO (Givón claims that the impact of cataphoric topicality is greater than the impact of anaphoric topicality), but GR preferences are calculated by the interaction of both dimensions (according to Givón (2001), both factors contribute). For this combination, addition of hearer salience and speaker salience is suggested.

As pronominalization threshold, we assume 0.5, again.

3.3 A minimal instantiation

For a minimal instantiation of the framework described above, the set of parameters as shown in Fig. 3 is considered. Topicality and the instantiations of Centering Theory can be represented by the choice of weights using these factors or parameters. Then, hearer and speaker salience are calculated as reciprocal of the weighted sum of parameter values. As result values, GR and WO preferences are assigned according to relative differences in salience scores, whereas derived REF preferences depend on absolute values (and ambiguity interference) as described above.

	parameter values						reconstruction							orig	
	backward-looking			forward-looking				LRCT		WO		TOP			
		Tony	Terry		Tony	Terry		Tony	Terry	Tony	Terry	Tony	Terry	Tony	Terry
1	RD_ante	$\rightarrow\infty$	$\rightarrow\infty$	RD_ana	0.00	0.00	hsal	$\rightarrow 0$	$\rightarrow 0$	$\rightarrow 0$	$\rightarrow 0$	$\rightarrow 0$	$\rightarrow 0$		
	GR_ante	0.00	0.00	REF_ana	0.50	1.00	ssal	0.67	1.00	0.67	1.00	0.57	0.56		
	WO_ante	0.00	0.00	TP	0.25	0.20	WO				(single)		(single)		(single)
							GR		sbj				**non-sbj**		sbj
							REF		full		full		full		full
2	RD_ante	$\rightarrow\infty$	0.00	RD_ana	0.00	0.00	hsal	$\rightarrow 0$	1.00	$\rightarrow 0$	1.00	$\rightarrow 0$	1.00		
	GR_ante	0.00	1.00	REF_ana	1.00	1.00	ssal	1.00	1.00	1.00	1.00	0.56	0.54		
	WO_ante	0.00	0.00	TP	0.20	0.15	WO			first*	first*	**first**	**non-first**	non-first	first
							GR	sbj*	sbj*			non-sbj	sbj	non-sbj	sbj
							REF	full	prn	full	prn	full	prn	full	prn
3	RD_ante	0.00	0.00	RD_ana	0.00	1.00	hsal	0.91	1.00	0.82	1.00	1.00	1.00		
	GR_ante	0.90	1.00	REF_ana	0.50	0.50	ssal	0.67	0.40	0.67	0.40	0.54	0.53		
	WO_ante	0.22	0.00	TP	0.15	0.10	WO			non-first	first	**first**	**non-first**	non-first	first
							GR	non-sbj	sbj			**sbj**	**non-sbj**	non-sbj	sbj
							REF	full*	prn	full*	prn	full*	full*	prn	prn
4	RD_ante	0.00	0.00	RD_ana	0.00	0.00	hsal	0.91	1.00	0.67	1.00	1.00	1.00		
	GR_ante	0.90	1.00	REF_ana	1.00	0.50	ssal	1.00	0.67	1.00	0.67	0.53	0.53		
	WO_ante	0.50	0.00	TP	0.10	0.10	WO			(single)		(single)		(single)	
							GR	sbj				sbj*		sbj	
							REF	full*	prn	full*	prn	full*	full*	full	
5	RD_ante	0.00	1.00	RD_ana	0.00	0.00	hsal	1.00	0.50	1.00	0.50	1.00	0.50		
	GR_ante	1.00	1.00	REF_ana	0.50	0.50	ssal	0.67	0.67	0.67	0.67	0.51	0.51		
	WO_ante	0.00	0.00	TP	0.05	0.05	WO			first*	first*	first*	first*	first	non-first
							GR	sbj*	sbj*			sbj	non-sbj	sbj	non-sbj
							REF	prn	full	prn	full	prn	full	prn	full
6	RD_ante	0.00	0.00	RD_ana	$\rightarrow\infty$	$\rightarrow\infty$	hsal	1.00	0.83	1.00	0.80	1.00	1.00		
	GR_ante	1.00	0.80	REF_ana	0.00	0.00	ssal	$\rightarrow 0$	$\rightarrow 0$	$\rightarrow 0$	$\rightarrow 0$	0.50	0.50		
	WO_ante	0.00	0.25	TP	0.00	0.00	WO			first*	first*	first*	first*	non-first	first
							GR	sbj*	sbj*			sbj*	sbj*	non-sbj	sbj
							REF	**prn**	full*	**prn**	full*	full*	full*	full	full

Figure 5: Example: Parameter values, Hearer salience ($hsal$), speaker salience ($ssal$) and coding preferences.

For the purpose of illustration, consider example sentence (5). Using the parameter weights as summarized in Fig. 4, hearer and speaker salience are calculated respectively, and preferences can be derived.

Considering Terry (te), we find that $RD_{ante}(te) = 1$ (his last mention was in sentence (3)) and $GR_{ante}(te) = 1.0$ (subject in (3)). Inserting these values in equation (1) using the parameter weights for hearer salience ($hsal$) in LRCT reconstruction as summarized in Fig. 4, we achieve a formula identical to equation (2). Thus, $hsal_{LRCT}(te)$ can be calculated as 0.5. As the proposed pronominalization threshold is not met, we predict reference with a full description. Accordingly, $hsal_{CCT}(te)$ converges against 0, thus the coding preferences in Canonical Centering would be identical. The antecedent of Terry is sentence-initial in (3), so $hsal_{WO}(te)$ is 0.5, too. In the topicality reconstruction, where referential distance is the only parameter of hearer salience, the same prediction is calculated, too.

The corresponding parameter values for Tony (to) in (5) are: $RD_{ante}(to) = 0$ (last mention in (4)), $GR_{ante}(to) = 1.0$ (subject) and $WO_{ante}(to) = 0$ (sentence-initial). So, hearer salience of Tony is calculated as 1 for LRCT, CCT, WO and TOP equally. This exceeds the respective pronominalization threshold. As the only possible interfering referent Terry has a sufficiently lower degree of hearer salience, no restrictions arise from ambiguity avoidance strategies. So, we can safely refer to Tony with a pronoun, just as taken in Grosz et al.'s original example.

For speaker salience ($ssal$), we find anaphoric references to Terry and Tony are in the directly following utterance ($RD_{ana} = 0$), both with full descriptions ($REF_{ana} = 0.5$), but only once in the forthcoming discourse ($TP = 1/20$). Thus, speaker salience is identical for both Terry and Tony, in 1 in Centering reconstructions 1 and $\frac{20}{39}$ in topicality reconstruction.

As grammatical roles resp. word order preferences are determined in Centering reconstruction by relative differences between speaker salience scores, no preferences for GR or WO can be deduced here.

The same holds for WO preferences in the topicality reconstruction. However, GR preferences in the topicality reconstruction are calculated from the interaction (e.g. addition) of hearer and speaker salience scores, but not from speaker salience alone. Therefore, Tony's score ($hsal_{TOP}(to) = 1$) exceeds Terry's ($hsal_{TOP}(te) = 0.5$), and we predict Tony to be preferred subject and Terry to be non-subject. In fact, the opposite decision was taken in Grosz et al.'s constructed example. However, this is very likely to be due to constraints from verbal semantics, as a more agentive realization of Tony in a sentence semantically roughly equivalent to ex. (6) would be rather odd (cf. ex. (6')).

(6') #Of course, Tony$_{to}$ has not been intended to get upset by Terry$_{te}$.

In Fig. 5 the deduction of coding preferences for the whole text is summarized. Besides the effects of a heuristic ambiguity rule[4] and partial indistinguishability, few crucial deviations from the original coding decision have been found. Here, (3) seems to be the most critical instance, where actual word order and grammatical role assignment deviate from both Centering and topicality preferences. However, the interpretation of the pronouns in (3) depends on parallelism with the previous utterance. A sentence like *He$_{to}$ was called (by him$_{te}$) at 6 AM.* is nearly incomprehensible. It would be necessary to use a full description such as the name at least for Tony or both referents (as suggested by the theories).

This short example already showed up some limitations of approaches of this kind. First, pragmatic preferences on word order, the assignment of grammatical roles, and possibly the choice of referring expressions, too, are by no means unrivaled. Rather, their application is most likely in cases where no other constraints arising from syntax (e.g. binding restrictions), semantics (e.g. valency frames of applicable verbs) or higher communicative goals (e.g. to add further hearer-new information about a referent within a noun phrase) interfere. Second, the theories and the corresponding reconstructions rely on surface-oriented heuristics that are often too coarse-grained to generate clear distinctions as shown for word-order preferences in ex. (5). Third, other factors might contribute to salience, too, such as parallelism effects and others.

4 The Mental Salience Framework: A summary

4.1 General characteristics

The Mental Salience Framework described in this paper consists of essentially three components:

- differentiation between hearer salience (reference to attentional states of the hearer) and speaker salience (manipulation of attentional states of the hearer),
- hearer salience and speaker salience are modeled as normalized linear combination of different contextual factors, and
- coding preferences are traced back to the linear combination of different contextual factors.

4.2 Adequacy with respect to existing theories

As the linear combination of contextual factors and the interaction between hearer salience and speaker salience involves different parameters, different parameter configurations can be considered, and as argued above, different variants of Centering and Givón's approach can be reconstructed by the choice of parameter values. The idea of an adequacy proof for these reconstructions (from a full proof I will refrain here for reasons of space) is as follows:

- Provide a definition of adequacy, i.e. all predictions of the original formulation of the theory are predicted by the reconstruction (completeness), and no prediction of the reconstruction is incompatible

[4] In Fig. 5, `full*` means to use a non-pronominal form to avoid ambiguity if the absolute salience score is sufficient for pronominalization, `subject*` and `unspec` indicate that Tony and Terry are ranked equally, deviations from original are marked **bold**.

with the predictions of the original formulation of the theory (compatibility).[5]

- Prove completeness of the reconstruction: Identify the set of empirically verifiable assumptions and predictions made in the original theory, and prove that these are predicted by the reconstruction as well. For Centering Theory, we have to show
  - any difference in the ordering of two potential backward-looking centers entails a difference between the hearer salience scores of the corresponding referents (for Canonical Centering by definition, for Left-Right Centering by induction),
  - if one element is pronominalized, then the backward-looking center is pronominalized (proof by contradiction, assume that the backward-looking center, i.e. the most hearer-salient referent, is not pronominalized. As it is more salient than the other pronominalized element, it must exceed the pronominalization threshold. As it is the most salient element, there can be no interference from ambiguity[6]), and
  - if one element in the following utterance is pronominalized, its grammatical role has preferably been higher in the current utterance than those of non-pronominalized elements appearing in both clauses (proof by contradiction: in the reconstruction, grammatical roles are assigned depending on the speaker salience scores. The only factor of speaker salience in the Centering reconstruction is pronominalization in the following utterance. In the reconstruction, a violation of this preference is possible only if the semantics in the current utterance do not permit this relative ranking of grammatical roles. This can be easily contradicted by enumerating the band-width of grammatical devices which allow the pragmatically adequate generation of grammatical roles.)
- Prove compatibility of the reconstruction with the original formulation: For Centering Theory, we have to show
  - if two elements differ in their hearer salience scores, then the lower-ranked one must not have been more salient according to Centering Theory. (proof by contradiction),
  - if a non-pronominal description is predicted by the reconstruction, then Centering Theory must not predict pronominalization (proof by contradiction, analogous to completeness proof above), and

[5] This definition of adequacy is inspired by the formal definition of equivalence. However, equivalence differs from adequacy in that it requires *soundness* rather than *compatibility*. A reconstruction is sound if "all predictions of the reconstruction are predictions of the original formulation as well". However, it has been recognized before that neither Centering Theory nor Givón's approach are fully specified models of discourse processing:

As such, Centering does not provide a model describing the cognitive underpinnings of the assignment of grammatical roles or other grammatical devices which indicate the ranking of forward-looking centers, and thus, it explains only the effect of these grammatical devices, but not their assignment in discourse. Nevertheless, the preference to keep the backward-looking center over a sequence of utterances (cf. the notion of "preferred center" (Strube and Hahn 1999)) can be exploited to predict the assignment of grammatical roles (Kibble and Power 2004). It should be noted, however, that these preference are deducted from preferences of transitions between utterances *only within the same discourse segment* (Grosz and Sidner 1986), and that it is not clear to what degree these preferences extend towards the discourse as a whole.

Similarly, Givón's approach involves a differentiation between anaphoric and cataphoric aspects of topicality, but he does not describe the interaction between these dimensions in order to derive concrete coding decisions.

Therefore, as any practical application of a theoretical construct, a reconstruction within a formal framework relies on an interpretation which is maximally predictive in order to achieve *concrete* predictions, and thus, researchers are usually not interested in equivalent reconstructions, but in reconstructions which involve a gain in predictive power. However, such a reconstruction cannot be *equivalent* as it systematically violates the soundness criterion.

As an example, Beaver's (2004) equivalence proof between Centering (as formulated by Brennan et al. (1987)) and his reconstruction of Centering in Optimality Theory represents in fact a proof of adequacy, as he claims that the reconstruction Centering entails additional predictions that were not entailed from the original formulation: "This declarativity means that COT is equally suited for generation or interpretation. In contrast, the BFP algorithm is suited for interpretation only. It could not be used to generate texts directly ...". As an alternative to equivalence proofs for subsets of the data considered, I suggest to distangle equivalence and adequacy and focus on the adequacy between original formulation and the reconstruction rather than on equivalence.

[6] However, this argument is dependent on the concrete definition of ambiguity. If ambiguity is defined on morphological agreement only, then, violations of Centering predictions are possible. However, ambiguity is often resolved from verbal semantics, and thus, also these factors have to be considered.

– if the reconstruction predicts the highest possible grammatical role, Centering Theory does predict preferred center (proof by contradiction: assume, Centering Theory unambiguously predicts another element to be preferred center, then, it must have been the only pronoun in the following utterance, but then, it must have been the backward-looking center of the following utterance, then, it must preferably have been the preferred center of the last utterance.)

For reasons of brevity, I restrict myself to this short sketch of the ideas behind the proof. Based on these considerations, however, I conclude that the reconstructions described above are *adequate* with respect to Centering Theory. A similar proof for Givón's approach can be made,[7] thus proving that the respective reconstructions within the Mental Salience Framework are adequate, and thus, the framework is capable to allow the reconstruction of two classical approaches.

4.3 Fields of application

Here, only a minimal set of parameters was considered, capable to reconstruct two classical approaches. The investigation of other factors is subject to later research. Finally, empirical measurements of speaker salience are intended to approximate intentions a speaker has about a referent, such as his wish to emphasize the role of a referent for the forthcoming discourse. However, it cannot be expected that such intentions can always be recovered from frequency measures such as topic persistence.

Still, important results can be achieved. Whereas speaker salience and hearer salience can be plausibly extrapolated from the original formulation of the theories, the derivation of concrete coding preferences is underspecified (especially for Centering), the interaction of hearer and speaker salience is not fully clear (TOP) or controversial (derivation of word-order preferences in WO and TOP), and the set of factors considered is incompatible. Within the Mental Salience Framework, such divergencies can be represented, and by means of salience metrics, they can be studied (and possibly resolved) on an empirical basis. An integrated framework as suggested here can be used

- to perform a comparative empirical evaluation of different theories resp. their reconstructions,
- to identify elementary factors considered in different theories and investigate their respective effect on salience scores,
- to evaluate hybrid or modified models by introduction or re-weighting of parameter values,
- to provide further insights in the interaction between speaker salience and hearer salience based on empirical results, and finally
- for practical application in natural language generation (NLG).

With respect to the last point, it seems reasonable to implement speaker salience in NLG systems as an external parameter providing an interface to integrate external "importance" assignments. Such importance assignments can be used by a system designer to guide the attention of a user in a goal-directed way. Besides this, hearer salience provides a mechanism for cohesive coding decisions based on text-oriented measures. One of the most important results to be achieved in empirical research is the clarification of the interaction of hearer and speaker salience and their respective influence for the choice of different grammatical devices.

4.4 Extensions and challenges

The original motivation underlying the Mental Salience Framework was the insight that a model of the attentional states of the hearer does not sufficiently constrain the choice of referring expressions, but that at least one additional dimension interfering with "givenness" must be considered as well. This observation has been made before, though, however, different candidates for this alternative, interfering dimension

[7] For Givón, only compatibility can be proven, as Givón's model is concerned with the analysis of empirical preferences, without specifying concrete predictions.

affecting the use of referring expressions have been proposed, e.g. contrastiveness, emphasis, importance (Givón 1983a; Levelt 1989; Chafe 1994), etc.

While the differentiation between common ground (as reflected in hearer salience) and speaker-private knowledge (from which speaker salience is built) is well justified and probably uncontroversial, the question remains whether these aspects of salience, hearer salience and speaker salience, form by themselves uniform dimensions of attentional states. Instead of defending this specific hypothesis, I motivate this assumption from methodological considerations, i.e. theoretical minimalism. The postulation of another distinction between, say, two kinds of speaker salience, must be justified from empirical findings which cannot be covered by the existing model. However, additional dimensions of salience arising from other modalities, e.g. visual salience, or other domains, e.g. property salience, are certainly independent from hearer salience and speaker salience which are solely concerned with the degree of attention *a discourse referent* is assigned at a given point in discourse.

A challenging question, however, is whether the grammatical devices of one type, say referring expressions, can be characterized by only one linear combination of salience scores or whether hearer salience and speaker salience differ in their impact on different grammatical devices. In fact, it has been suggested for demonstratives as compared to personal pronouns, that the condition licensing the use of demonstratives are more specific than those of personal pronouns. In Finnish, Kaiser and Trueswell (to appear) found a preference for personal pronouns to co-refer with the subject of the preceding utterance, whereas demonstratives tended to take the last mentioned possible antecedent. They explained this difference with different interoperating dimensions, i.e. linguistic structure (as indicated by grammatical roles) and information structure (as indicated by word order in Finnish), that differ in their relevance to the choice of pronouns as compared to the choice of a demonstrative. In the Mental Salience Framework, this configuration can be modeled by defining (a) hearer salience in terms of grammatical roles, and (b) speaker salience in terms of word order (if indicating non-salience, i.e. non-givenness). Then, the observed pattern can be achieved by defining that personal pronouns depend on hearer salience alone, whereas demonstrative pronouns are sensitive to speaker salience besides hearer salience.

This specific model of demonstratives, however, requires that not *one* cumulated salience score for the generation of referring expressions is generated, but that for certain smaller classes of referring expressions, individual scores are calculated and then, interpreted as the probability to use a specific kind of referring expression. Thus, the association between salience score and a certain grammatical device is no longer a direct one, say, a mapping from a certain score on a scale to a preference for a certain form, but it is a mapping from a two-dimensional space onto the preference for a given form, guided by the proximity between the canonical salience of that form and the scores currently achieved. The Mental Salience Framework permits this kind of extension, though it is currently concentrating on the most elementary classes of grammatical devices, abstracting from more fine-grained differentiations such as the differentiation between pronouns and demonstratives.

Another possible extension is the application of the Mental Salience Framework in learning algorithms. As factors, salience scores and coding preferences are specified by numerical scores, which are retrieved from linear combinations, this network can be interpreted as a multi-layer perceptron whose weights (parameters) can be set by backpropagation.

As a result, the Mental Salience Framework allows not only for the comparative representation and evaluation of different theories, but also for data-driven parameter weighting.

5 Related research

The Mental Salience Framework represents a mathematical model of a certain insight on the nature of attention control in discourse, that is, the differentiation between different dimensions of the salience of discourse referents, associated with different functions in the flow of discourse: hearer salience which is part of the speaker's hearer model and is exploited by him to generate expressions in a way that a hearer can relate them to elements introduced in the discourse before, and speaker salience which correlates to the intention a speaker has to focus the hearer's attention on certain referents, e.g. their role for the further development of the discourse.

5.1 Multidimensional Models of Salience in the Generation of Referring Expressions

While similar proposals have been proposed before (Givón 1983a; Clamons et al. 1993; Mulkern 2003), these often remained merely theoretical, and, to my knowledge, have not been formalized within a model for the prediction of the choice of referring expressions, the assignment of grammatical roles, and the deduction of word order preferences. The differentiation between different types of salience in NLG contexts as proposed by Pattabhiraman (1992) concerns another distinction, that is, the relationship of the degree of (instantial) salience a cognitive representation has, and the degree of (canonical) salience a grammatical device, or a given lexeme, is capable to express. In his terminology, hearer salience and speaker salience are both different aspects of instantial salience, whereas canonical salience is concerned with the mapping between grammatical devices and salience scores. In fact, Pattabhiraman's model of salience in NLG can be used as an alternative to the deductive linear combination approach presented here.

Pattabhiraman's canonical salience is related to the notion of salience as developed in the field of semantics of comparisons and metaphors. In her investigation of metaphorical and literal readings of potentially metaphorically interpretable expressions, Giora (1999) introduced the notion of salience as an assessment for the likelihood of a semantic meaning a given sequence of words can be assigned. Similarly,in his classical work on comparisons, Tversky (1977) postulated that semantic features differ in their relative salience for different elements, and that these differences have an effect on the ordering of elements in a comparison. In a later extension of Tversky's work, Ortony (1979) found that feature salience has an effect on the well-formedness of metaphoric expressions. Horacek's algorithm for the generation of referential descriptions (Horacek 1997) broadened Tversky's and Ortony's understanding of salience by identifying the role of *property salience* in the generation of referring expressions in general, that is, to account for the observation that referring expressions often involve attributions that are not primarily motivated by their capability to distinguish the given referent from a set of semantically compatible distractors, but from independent considerations. Though these approaches are differing in their assumptions about the exact nature of salience, they share the idea that salience operates on the interface between meaning and description, while in the study of reference, salience is seen as a property of discourse referents describing their availability to the hearer or changes in this degree of accessibility.

Thus, property salience and object salience are independent from each other, and, as suggested by van der Sluis and Krahmer (2001), it can be assumed that both dimensions co-operate with other dimensions of salience in the production of the form of referring expressions. The third dimension of salience considered by van der Sluis and Krahmer comes from environmental factors, especially the visual surrounding. Effects of the situational context on the choice of referring expressions have been observed frequently before. Similar to Bühler's (1934) interpretation of deixis as an extension of anaphora, Prince (1981) considers "situationally evoked" and "textually evoked entities" to form a homogeneous group of highly activated (evoked) referents. Also in the context of multi-modal generation of referring expressions, the interaction between visual salience and linguistic salience has been investigated (cf. the contribution by John Kelleher, in this volume).

With respect to other existing multi-dimensional models of salience, we may conclude that besides hearer salience and speaker salience, additional dimensions of salience can be assumed besides hearer and speaker salience which differ from salience in the context of intra-textual reference in their domain (canonical salience/feature salience/property salience) or their modality (visual salience), and are thus independent from the dimensions of salience discussed here, which are more strictly concerned with the flow of discourse. Due to this independence, however, these are compatible with the differentiation between hearer salience and speaker salience and thus, they can be regarded as potential augmentations of the Mental Salience Framework.

The differentiation between hearer salience and speaker salience, however, is theoretically well justified (Clamons et al. 1993; Mulkern 2003), but has not been formalized before, and accordingly, also the Mental Salience Framework can be regarded to provide a more precise model of linguistic salience as compared to older mono-dimensional accounts of linguistic salience as currently employed by existing models for the generation of referring expressions, also in multi-modal contexts, which concentrate on hearer salience, e.g., van der Sluis and Krahmer (2001); Kelleher and van Genabith (2004).

5.2 Centering in Optimality Theory (COT)

It has been shown above that the Mental Salience Framework is capable to allow for an adequate representation of existing theories such as classical variants of Centering Theory and related theories such as Givón's bi-dimensional account of topicality. Similar attempts for the integration of previously independent lines of research within one framework have been proposed before,[8] but the Mental Salience Framework differs from these in that its theoretical implications are fairly minimal, that is, essentially only that speaker-private intentions and beliefs have to be separated from the assumptions of the speaker about attentional states of the hearer.

In this theoretical minimalism, the Mental Salience Framework shares a certain resemblance with Optimality Theory, which can also be viewed as a formal apparatus within which existing theories such as Markedness Theory (Aissen 2001), or Centering Theory (Beaver 2004) can be reconstructed.[9]

Optimality Theory relies on the observation that grammars contain constraints on the well-formedness of linguistic structures, and often, these constraints are heavily in conflict. The rapid and systematic resolution of such conflicts implies that constraints are not equal in their violability, but that they are ranked. According to OT, constraints are elements of the universal grammar, and language-specific grammars are instantiations of the UG in that they represent different possible rankings of universal constraints.

Formally, constraints in OT are conditions on the relationship between an underlying form, or input, and a set of possible surface candidates, i.e. possible output. For the generation of referring expressions, the input is an underspecified logical form of an utterance, the output is a candidate utterance. The optimal candidate output is selected based on the ranking of violated constraints. Given two candidate forms A and B, A is more optimal than B if the highest-ranking constraint which is violated by B is not violated by A, and no violations of higher-ranked constraints occur for A.

Beaver proposes a set of constraints which capture the main ideas of Centering as formulated by Brennan et al. (1987), involving the following constraints:

- PRO-TOP The topic is pronominalized. (Rule 1)
- COHERE The topic of the current sentence is the topic of the previous one. (dis-preference of shifts, Rule 2)
- ALIGN The topic is in subject position. (dis-preference of shifts, Rule 2)

Further, Beaver provides a constraint-based definition of the backward-looking center ("topic"):

- ONE-SENTENCE-WINDOW Only discourse entities mentioned in the previous sentence are salient. (salience definition)
- ARG-SALIENCE One discourse entity is more salient than another if the first was referred to in a less oblique argument position than the second in the same sentence. (salience definition)
- UNIQUE-TOPIC With respect to any sentence, there is exactly one discourse entity which is the topic of that sentence. (definition backward-looking center)
- SALIENT-TOPIC The topic of a sentence is the most salient discourse entity referred to in that sentence, and undefined if no previously salient entities are referred to. (definition backward-looking center)

The minimal version of COT also involves further constraints which are not directly motivated from Centering Theory:

- FAM-DEF Each definite NP is familiar. This means both that the referent is familiar, and that no new information about the referent is provided by the definite.

[8] Previous proposals include the attempts of Hajičová and Kruijff-Korbayová (1997), Krahmer and Theune (2002) and Navaretta (2002) to bring together the Praguian notion of salience developed by Hajičová and Vrbova (1982) and Centering Theory (Grosz et al. 1995).

[9] However, the concrete claims by Optimality Theory are more rigid, but only concern the *nature* of constraints as a component of Universal Grammar.

Using this reconstruction of Centering, Beaver shows the equivalence between COT and Brennan et al.'s original account with respect to pronominalization. However, it should be noted, that like the reconstruction of Centering within the Mental Salience Framework, the predictions made by COT are more specific and more elaborate than the predictions of Brennan et al. (1987). The constraint FAM-DEF, though a reasonable assumption, is not motivated from Centering Theory, and PRO-TOP differs from Rule 1 in that it is indistinctive between two critical cases, i.e., (a) no pronominalization in the output, and (b) pronominalization of non-backward-looking center in the output, but not of backward-looking center.

For his equivalence proof, Beaver concentrated on proper names and pronouns only (excluding FAM-DEF), and proves equivalence between COT and Centering with respect to three critical cases:

- Purely anaphoric resolutions breaking syntactic constraints are never COT optimal, and never correspond to preferred BFP transitions.
- Fully anaphoric resolutions which violate Rule 1 are never COT optimal, and never correspond to preferred BFP transitions.
- Suppose two fully anaphoric resolutions A and B of a sentence satisfy syntactic constraints and Rule 1. If COT ranks candidate A above candidate B then BFP ranks candidate A above candidate B and vice versa.

For the third case, however, Beaver's proof relies on the assumption that " Since Rule 1 is satisfied by A and B and there are pronouns, PRO-TOP is also satisfied by A and B." Earlier, he described the motivation of PRO-TOP: "PRO-TOP has essentially the effect of Centering s Rule 1. ... If there are pronouns, then PRO-TOP will function comparably to Rule 1, providing a preference for interpretations that make the topic (i.e., CB) into a pronoun." However, the formulation "if there are pronouns" involves a great abstraction, in that is assumes that pronominalization is triggered only by salience (and agreement filters). As noted in the sketch of the adequacy proof above, this assumption predicts the same results as the original Centering rule only if the definition of agreement filters may extend beyond strict morpho-syntactic congruency.

Further, aside from the critical cases identified above, we can construct an example in which PRO-TOP and Centering make different predictions about pronominalization:

Mary$_m$ watched Sue$_s$ crossing the street over to Harry's$_h$ house. (Mary, Sue $>$ Harry)

7. She$_{m/s}$ wondered about the low traffic today.
8. He/Harry$_h$ did not realize her$_{m/s}$.
9. He$_h$ did not realize Mary$_m$/Sue$_s$.
10. Harry$_h$ did not realize Mary$_m$/Sue$_s$.

The examples 7 to 10 are possible continuations of the first sentence. The well-formedness of example 7 for both interpretations illustrates that Mary and Sue are equally possible antecedents of a pronoun in the subsequent sentence, thus, a feminine pronoun would be ambiguous between Mary and Sue. Therefore, ex. 8 is fully ambiguous between both readings, and thus from cooperativity considerations, we may conclude that it is not a feasible candidate output. As a consequence, only exx. 9 and 10 are to be considered by COT respectively Centering. However, Harry is clearly more oblique than Mary and Sue in the first utterance, and thus, it cannot be the backward-looking center. Therefore, ex. 9 violates both Rule 1 and PRO-TOP. However, ex. 10 violates PRO-TOP, but does not violate Rule 1. Therefore, if ex. 8 is excluded for independent reasons, the Centering-optimal output is ex. 10, whereas COT is indistinctive between ex. 9 and ex. 10.

At this point, I would like to emphasize that because of the unconditional formulation of PRO-TOP in COT and the existence of the FAM-DEF constraint, COT makes predictions beyond the original Centering Theory, and thus, must be deemed *adequate* with respect to Centering, rather than *equivalent*. The

equivalence proof provided by Beaver is concerned with a subset of critical cases and only with the differentiation between pronouns and the use of proper names. It is possible to construct a critical example in which Centering and COT make different pronominalization predictions.[10]

Thus, Centering in OT and the reconstruction of Centering within the Mental Salience Framework are comparable with respect to their adequacy ("equivalence").

However, the theoretical implications of an OT modeling cannot be underestimated. Essentially, all possible constraints must be part of the universal grammar. Postulating a constraint like FAM-DEF entails the assumption that definite NPs form a *universal* syntactic category, which is clearly contradicted by the existence of languages which have no explicit definiteness markers. Further, the OT reconstruction of Centering, like the original formulation of Centering, are inherently symbolic, categorial accounts, which are capable to predict a finite and fixed set of possible categories of referring expressions. One of the central criticisms of *categorial* accounts of givenness brought forward by Mira Ariel (1990; 1994; 2001) states that the number of grammatical devices distinguished in a specific language, is theoretically unlimited, and if all relevant distinctions among referring expressions are to be captured in an extension of COT similar to the familiarity criterion of definite NPs by the postulation of the corresponding constraints, the formulation of these categories and their salience characterization in OT also entails that these categories are also present in universal grammar, which is probably misleading. As opposed to this, in the Mental Salience Framework, the number of possible referring expressions is not *a priori* limited, but can be justified in terms of their salience characterization.

In the OT and in the anaphor resolution communities, further instantiations of Centering in OT have been developed (Buchwald et al. 2002; Bouma 2003; Byron and Gegg-Harrison 2004; Hardt 2004). From these, the conceptual motivations underlying the Recoverability Optimality Theory (ROT) model (Buchwald et al. 2002) are very closely related to underlying insights of the Mental Salience Framework. Both share a production perspective which heavily depends on the availability of two discourse models, the model of the speaker's private intentions and the discourse model, the salience list or "common ground". Both share the assumption that cues from the subsequent discourse must be considered to model the generation of referring expressions properly. And, as well as the Mental Salience Framework, ROT is a parameterized framework in the sense that the set of constraints considered is subject to possible extensions. Indeed, the Mental Salience Framework could be applied for the ranking of the current and the following salience list, and thus serve as a complement to ROT with respect to the concrete model of salience which is left unspecified so far.

Nevertheless, the Mental Salience framework is less constrained in its theoretical implications and in its adaptive character. Especially, it supports language-specific categories of referring expressions whose treatment in Optimality Theory is uncertain. In the best case, the integration of additional categories of referring expressions only requires to associate them with certain salience scores. Accordingly, the Mental Salience Framework is more oriented towards a broad-scale practical application.

5.3 Centering as a parametric theory

Besides approaches dealing with the reconstruction of different theories within a more general framework, also the variation of parameters within one theory has been considered.

By its impressive acceptance across different disciplines of linguistics, Centering Theory has become widely adapted throughout a great community. However, as a necessary consequence of this wide spread, the theory was modified in certain contexts. As one example, OT approaches abstract from the formulation of transitions between utterances (Beaver 2004), and even from the concept of backward-looking center (Hardt 2004), thus leaving essentially nothing of the original theory but the metaphor that attention has to be "centered" during discourse processing.

[10] Of course, this can be compared by a formulation of PRO-TOP which is closer to the original Rule 1 and thus, it provides no counter-evidence for the reliability of a reconstriction of Centering in OT in general.
Note, that also for the Mental Salience Framework, Centering-conformant behaviour can be achieved only by the use of specialized ambiguity filters. For this example, the Mental Salience reconstruction of Centering predicts 9, if ambiguity is determined by morphological agreement only. However, the Centering-optimal prediction can be achieved if ambiguity is defined without any morphological restrictions, which ultimatively leads to the following Centering-conformant, but not very natural, strategy: pronominalize nothing but the backward-looking center (Kibble and Power 2004).

But also in more conservative formulations of Centering Theory, parameters such as the definition of *utterance*, the definition of possible forward-looking and backward-looking centers, the criterion of forward-looking centers to be *realized within an utterance*, and different salience rankings are varying throughout the literature. Some of these parameters have been empirically investigated by Poesio et al. (2000; 2004) who considered empirical effects of variation in the definition of utterance (sentence, finite clauses, all clauses with a verb, ...), realization (indirect realization: consider not only anaphoric, but also bridging relations between forward-looking centers and potential backward-looking centers; considering non-third person pronouns as forward-looking centers), and different salience rankings.

While the empirical evaluation of different parameters of Centering Theory is a worthful and important achievement, it opens the question what concrete claims of the original theory really remain. From their study, Poesio et al. (2004) motivate a re-formulation of certain aspects of Centering Theory, which, however, is not compatible with radical approaches such as Hardt's Dynamic Centering (2004). One of the most important results of the study is, however, that Centering Theory cannot be evaluated without considering concrete instantiations of the different parameters it involves. As long as these parameters remain not fully specified, it is unclear to what degree Centering Theory can be falsified at all. Therefore, the central criticism of Centering Theory is not in any of its specific claims, but only in its theoretical status. That is, essentially it must be regarded as a framework which proposes a certain terminology and formalism, but not as a theory in the strict sense.

On the other hand, the achievements of Centering cannot be denied. For the first time, a common terminology on several discourse phenomena has been established across different disciplines of linguistics. This proposal, however, differs from Centering Theory in that it does not claim that it represents a *theory*, but merely a formalism, or a *framework*. The crucial difference is that a theory must be falsible, whereas a "parametric theory" as long as it cannot be evaluated independently from its parameters, is nothing but a metaphor.

However, the main difference between Centering Theory and the Mental Salience Framework is that within the Mental Salience Framework, a numerical account of salience is provided and explicitly modeled with respect to the choice of referring expressions, grammatical roles and word order preferences, whereas in Centering, pronominalization is seen as a bye-product of entity coherence with only very weak consequences on the choice of referring expressions at all. In the Mental Salience Framework, however, this relationship is formulated in a very explicit way, in that numerical scores are mapped onto specific coding preferences. Further, it applies beyond the scope of pronominalization as opposed to the choice of full nominal NPs, in that it is compatible with the fine-grained specification of an arbitrary number of different grammatical devices in terms of the salience conditions their appropriate use depends on.

Further, Centering does not provide a model for the assignment of grammatical roles, but only for their effect on local coherence. In functional linguistics, this function is identified as "foregrounding"; speaker salience can thus be described as the need of the speaker to place entities in the foreground of a scene, e.g. in order to process the subsequent discourse. Centering relies on surface-oriented factors that indicate foregrounding, it implicitly takes an interpretation perspective on the discourse it is applied to. As opposed to this, the Mental Salience Framework clearly takes a production perspective in that it includes an explicit model of attentional states of the speaker, and thus, it is more specialized for the needs of Natural Language Generation.

5.4 Centering Games

As an extension of the identification of the parameters of Centering (Poesio et al. 2000) and the existence of different reconstruction of instantiations of Centering Theory in Optimality Theory, Kibble (2003) proposed a game-theoretic reconstruction of Centering Theory as a framework for collaborative reference resolution as a non-cooperative game of incomplete information. With our approach, the game-theoretic reconstruction of Centering shares the assumption that two perspectives, hearer perspective and speaker perspective, have to be distinguished.

The relevant processing modules of the hearer perspective include:

discourse modeler maintains a record of entities mentioned in the discourse which will be candidates for anaphora resolution. Possible discourse models include (a) a centering model, (b) the list of focal

referents from the previous clause, or (c) a fully specified discourse model.

reference resolver identifies the referent of a referring expression with an entity in the discourse model.

The relevant processing modules of the speaker perspective include:

planner/content determination organizes input propositions into a text structure; plan sentences by e.g. choosing verb forms to realize preferred order of arguments. Possible strategies include to (a) promote arguments within a clause according to their perceptual salience, (b) plan consecutive clauses to align salience rankings, or (c) plan sequence of clauses to maximize referential continuity, in addition to salience alignment.

realizer generates appropriate referring expression to denote arguments of predicates.

Some details of Kibble's approach remain abstract, and the adequacy of this approach has not been proven so far. However, with the exception of the reference resolver which has no direct parallel in the Mental Salience Framework, its concepts can be interpreted in terms of Kibble's Game-theoretic framework. Hearer salience is clearly a part of the discourse modeler, though a fully specified discourse model involves additional aspects beyond the modeling of attentional states of the hearer. The strategies enumerated in the planner/content determination module are partly concerned with the assignment of grammatical roles. Strategy (a) is concerned with perceptual salience only, but is roughly parallel to the word order and grammatical role-strategies specified for speaker salience in TOP. Strategies (b) and (c) involve an "alignment of salience rankings" with utterances from the subsequent discourse, and seemingly, this corresponds to the extrapolation of speaker salience from coding decisions in the subsequent discourse according to the Centering reconstructions. Finally, the realizer covers the determination of coding preferences (from the linear combination of salience scores) and their application.

Hence, the conceptions of the Mental Salience Framework seem to be closely related to Kibble's Game-theoretic reconstruction ("elimination") of Centering, and it might be regarded a more concrete framework for formulation of the strategies suggested by Kibble.

6 Summary and outlook

A generalized parameterized framework was sketched providing an architecture for mechanisms of attention control by the salience-based assignment of coding preferences for referring expressions in discourse.

Relying on the previously noticed multi-dimensionality of salience, the distinction of two dimensions of salience was suggested which is consistent with different terminological traditions relating the notion of salience to accessibility/givenness and importance/newsworthiness respectively. As an illustration of theoretical adequacy, a minimal instantiation has been proposed capable to represent Givón's topicality approach and two instantiations of Centering Theory. Further, a proof to the adequacy of these reconstruction was sketched.

Hence, the Mental Salience Framework provides a proper basis for the comparative evaluation of these and related theories. Beyond this, the numerical character of the parameters allows for the application of learning algorithms, e.g. based upon an interpretation of the architecture as illustrated in Fig. 2 as a neural network. Thus, a supervised learning algorithm can be applied to assign parameter weights according to empirical data.

As a result, an integrated architecture for cognitive-pragmatic aspects of attention control in discourse has been suggested. Due to its appealing simplicity and intuitivity, the implementation for NLG systems becomes likely and is the perspective aim of this research. In this domain, it provides key mechanisms for both optimizing coherence/cohesion of automatically generated texts (by coding preferences due to hearer salience) and the assignment of judgments of emphasis, relevance or importance (speaker salience, if interpreted as relevance, provides an interface to guide the hearer's attention onto certain aspects or entities according to external parameters).

7 Appendix: Proving the adequacy of Centering reconstruction

This paper deals with a reconstruction of several theories within a certain framework, so far, basically on an intuitive line of argumentation. However, in order to satisfy the conclusion presented above, that the framework is *adequate*with respect to the theories, that is, it allows for a reconstruction of a given theory within the framework, equivalence between the reconstruction and the original theory has to be formally proven. For the examples of Canonical and Left-Right Centering, a short proof will be sketched in this section.

Def. 1 (Equivalence) *The formal reconstruction of a theoretical model is equivalent to the model itself iff.*

(a) any prediction of the model is predicted by its reconstruction (completeness), and

(b) any prediction of the reconstruction is predicted by the model as well (soundness).

The equivalent re-formulation of a theory within a given framework does not represent a theoretical achievement by itself, but rather, it means that the formalisms applied within the framework are flexible and powerful enough to allow for a reconstruction.

Therefore, researchers are usually not intended to develop an *equivalent* reconstruction, but they aim to prove that their reconstructions provide additional insights or allow for further generalizations beyond the scope of the original theory.

As an example, Beaver's reconstruction of Centering Theory in Optimality Theory is proven to be "equivalent, but greater in its explanatory power". In fact, this understanding of 'equivalence' is incompatible with a more traditional definition as formulated in Def. [ABOVE]. Essentially, this statement entails that the completeness condition (a) of equivalence is fulfilled, but that the soundness condition (b) is violated in a systematic way.

In order to capture this loosened understanding of equivalence, I suggest a reformulation of condition (b) as follows:

> (b') no prediction of the reconstruction is contradictory with predictions of the original model. (compatibility)

The crucial difference between compatibility and soundness is that (b') allows for a reconstruction which produces more detailed predictions than the original theory, i.e. it represents a stronger hypothesis than the original theory. However, such strong formulations are inavoidable for making the approach practically applicable (Kibble 2000), that is, they are part of any attempt to operationalize theoretical considerations.

The difference between (b) and (b') concerns theories which claim that their predictions are underspecified. For Centering Theory, as one example, it has always emphasized that different factors contribute to salience besides grammatical roles, but as it could be shown that subjecthood has an especially robust effect on pronominalization, the salience ranking of Canonical Centering was defined in terms of grammatical roles. In fact, recent results support the multi-factorial view on salience.

However, in order to avoid confusion between the general understanding of equivalence in mathematics and Beaver's usage of the term, I introduce the term "adequacy".

Def. 2 (Adequacy) *The formal reconstruction of a theoretical model is adequate to the model itself iff*

(a) any prediction of the model is predicted by its reconstruction (completeness), and

(b) no prediction of the reconstruction is contradictory with predictions of the original model (compatibility).

As argued above, what reseachers are really interested in, are reconstructions that are adequate, but not equivalent, i.e. those formalisations of theories that are consistent with the theories, but allow for more precise and more detailled instantiations of the theories.

In the remainder of this section, I will prove the adequacy of the reconstruction of Left-Right Centering within the framework with respect to pronominalization.

First, I prove that the salience ranking according to Canonical Centering and Left-Right-Centering is the same as the one in the reconstruction.

7.1 Adequacy of hearer salience scores

Sentence 1 (Completeness of hearer salience scores) *The reconstruction of the salience ranking by hearer salience scores is complete.*

Proof proof by induction

- induction anchor first sentence in the text, no salience ranking. Therefore, no prediction.
- induction step

  Canonical Centering: Referring expressions from the preceding utterance are ranked according to their grammatical role.

  Left-Right-Centering: Referring expressions from the preceding utterance are ranked higher than those from the utterance before, etc. Within any utterance, referring expressions are ranked according to their grammatical role.

  Reconstruction: only the previous discourse is considered ...

Sentence 2 (Compatibility of hearer salience scores) *The reconstruction of the salience ranking by hearer salience scores is compatible.*

(Canonical Centering)

Proof proof by induction

induction anchor first sentence in the text. No prediction from Centering Theory, thus, no violation of these predictions.

induction step given two referents r_1 and r_2 that have certain salience characteristics at Un-2

(a) both are mentioned at Un-1 salience score is newly assigned depending on grammatical role (by definition identical to Canonical Centering)

(b) r_1 is mentioned at Un_1, but r_2 not; salience score of r_1 is higher than the score of r_2 (Canonical Centering: r_2 has no salience score, thus lower than r_1 LRCT: search in U_1 for a potential antecedent, thus r_1 will be retrieved earlier than r_2, which can be modeled by a partial ordering with entities from U_{n-1} ranked higher than entities from $U_n - 2$

(c) neither r_1 nor r_2 are mentioned in $Un-1$ reconstruction: (salienzscores umformulieren, so dass auf step-salienz zurückrechenbar) Canonical Centering: both are excluded from the salience ranking, thus no contradiction with the predictions of Canonical Centering LRCT: if iteration of the search cyclus is permitted, then the ranking of r_1 and r_2 is preserved from the step mit scores beweisen

From completeness and compatibility, the adequacy of the hearer salience scores can be concluded.

Corollary 1 (Adequacy of hearer salience scores) *The reconstruction of the salience ranking by hearer salience scores is adequate.*

7.2 Adequacy of pronominalization

Based upon an adequate reconstruction of salience within the framework, the adequacy of pronominalization

Sentence 3 (Completeness of pronominalization predictions) *Any pronominalization prediction made by Left-Right Centering and Canonical Centering is predicted by the reconstruction as well.*

Proof proof by induction

- induction anchor: The first utterance in the text, no backward-looking center. No prediction on pronominalization by Left-Right Centering. Trivially fulfilled by the reconstruction.

- step:
  (a) No element from the preceding utterance Un-1 is mentioned in the actual utterance Un . No prediction by Canonical Centering. (Trivially fulfilled.)
  (a') At least one element from the previous utterance Un-2 is mentioned in Un, and it is more salient than any other entity mentioned in Un and Un-2, thus, it is the Cb according to Left-Right-Centering. If another pronoun in Un exists, LRCT predicts pronominalization of the Cb. In the reconstruction, distance (in utterances) is the most influential factor on salience, thus, elements from Un-1 are more salient than elements from Un-2, etc. As no elements from Un-1 appear in Un, the most salient entity is the referent with the highest grammatical role in Un-2, i.e. the backward-looking center. If a less salient referent can be pronominalized, the pronominalization threshold for the most salient element must also be exceeded. As it is the most salient one, there is no interference from ambiguity, and thus, it must be pronominalized.
  (b) At least one element from the preceding utterance is mentioned in the current utterance. As compared to any other element of Un-1 and Un, it is the most salient one, and thus, the backward-looking center. If another, less salient referent exists that is pronominalized in Un, the backward-looking center must be pronominalized. In the reconstruction, elements from the preceding utterance are more salient than elements from the earlier discourse. The most salient referent thus corresponds to the backward-looking center. If a less salient referent than the most salient one can be pronominalized, the pronominalization threshold is exceeded. As it is the most salient one, there is no intereference from ambiguity, and thus, it must be pronominalized.
  (c) More than one element from the preceding utterance is mentioned in the current utterance. However, the highest-ranking referents are equal in their salience score (e.g. both OTHER). No backward-looking center can be identified, thus, no prediction. (Trivially fulfilled.)

Sentence 4 (Compatiblity of pronominalization predictions) *No pronominalization prediction of the reconstruction is incompatible with predictions of Canonical Centering or Left-Right Centering.*

This sentence can be formulated more rigidly for the only clear prediction Centering Theory makes about pronominalization.

Lemma 1 *The reconstructions does not predict the pronominalization of a non-CB unless the CB is to be pronominalized as well.*

Proof Proof by contradiction

Given, the reconstruction predicts at least one pronoun in the current utterance, i.e. the referent does exceed the pronominalization threshold and there is no interference from ambiguity, i.e. it is the most salient from the class of morphologically compatible referents.

Assume that there is another, more salient referent which corresponds to the backward-looking center, but is not pronominalized. As the referent exceeds the pronominalization threshold, the only reason for the referent not to be pronominalized is ambiguity. As it is (by definition), the most salient referent, there cannot be another referent as salient as the backward-looking center, and thus, ambiguity cannot prevent the referent to be pronominalized. Thus, assuming that the only pronoun within an utterance is different from the backward-looking center leads to contradiction.[11]

[11]This conclusion, however, depends on a specific definition of ambiguity. If ambiguity operates only within classes of nominals with the same gender and number characteristics, it is possible that the backward-looking center is prevented from pronominalization by another, more salient referent which is not realized in the current utterance, whereas a less salient referent exceeds the pronominalization threshold but is not ambiguous because of the lack of morphologically compatible alternative antecedents for a pronoun.

Examples of this type can be constructed, but, in spite of several exceptions from Centering predictions to be discussed [BELOW], I have not been able to find this specific violation in natural occurring text.

Jim watched Peter talking to Mary.

Corollary 2 (Adequacy of pronominalization predictions) *The predictions made by the reconstruction of Canonical Centering Theory and Left-Right Centering are adequate as compared to the original theories.*

7.3 Grammatical Roles

Grammatical roles are not predicted by Centering Theory, but serve as factors for the identification of the backward-looking center.

However, pronominalization in the subsequent discourse can be regarded as a result of the planning decisions a speaker made before, and if the assumption about the centering function of grammatical roles is correct, the assignment of grammatical roles is not random, but guided by the speaker's intentions to indicate the flow of attention in the forthcoming discourse.

Therefore, pronominalization in the forthcoming discourse can be regarded an intended result of a planned action of the speaker, i.e. his intention to promote referents into the foreground of a scene (Pustet 1997), e.g. to establish a new topic as a reference point in discourse (Givón, Lambrecht). In Centering, this attention-guiding function is captured by the interpretation of grammatical roles in their effect to indicate the backward-looking center of the following utterance, thus making Givón's and Lambrecht's understanding of topic comparable to the backward-looking center, and, more specifically, establish a parallel between Lambrechts topic-announcing grammatical devices and the preferred center in Centering terminology.

At the same time, Rule 2 states a preference for the continuation of a previously established backward-looking center in order to account for topic chains. Speaker salience, then, can be regarded as an expression of the speaker's intention to highlight elements of the discourse and to guide the hearer's focus of attention to them. Certainly, the development of the forthcoming discourse is only *one* dimension that affects the speaker's decision to guide the hearer's attention to a certain referent in discourse, but assuming that a cooperative speaker prepares the hearer for the development of the forthcoming discourse, in order to enhance his

Therefore, grammatical roles can be regarded as an expression of speaker salience, that is, his intention to highlight elements at a given point of discourse,

The salience of forward-looking centers, which is usually considered in Centering Theory only by contextual factors such as grammatical roles, etc., is the expression of a certain cognitive reality, and as Centering formalises a relationship between the hearer salience in the following utterance and the (indicators of) speaker salience in the current utterance, speaker salience can partly be recovered from the hearer salience in the following utterance. Accordingly, speaker salience and the assignment of grammatical role preferences can be predicted from pronominalization decisions in the following utterance.

While this interpretation does not present a claim upheld by Centering Theorists, others have explored the effect of planning decisions as reflected in the forthcoming discourse on the choice of grammatical devices, e.g. Givón with his concept of thematic importance.

Sentence 5 (Completeness of speaker salience) *Any prediction of Centering Theory on the signalling of the salience of forward-looking centers is captured by the reconstruction in terms of speaker salience.*

Instead of providing a formal proof, I will argue that the premise of this sentence is empty, and thus, that it is fulfilled trivially.

In centering Theory, the conceptual status of the salience of forward-looking centers is not very clear, and thus, it would be unreasonable to investigate the completeness of the reconstruction. To my knowledge it has not been attempted by Centering Theorists to formalize the effect of contextual factors that are signalled indirectly by the development of the subsequent discourse on the assignment of grammatical roles or other devices affecting the ordering of forward-looking centers.

(a) She really enjoyed finally gaining Jim's/Peter's/his$_p$/j attention.

In fact, the underlying assumption of Centering, that is, that attention has to be centered in discourse, and that this centering of attention has to be signalled unambiguously, entails a broader understanding of ambiguity, that is ambiguity about the current center of attention. Pronominalizing a non-CB besides a non-pronominal CB would misdirect the attention of the hearer towards a non-CB referent and confuse him. By this definition of ambiguity, however, the correspondence between the predictions of the model and the predictions of Centering becomes trivial as this is exactly the requirement of Rule 1.

Centering Theory, however, is not intended to cover any transition between utterances, but only between utterances of the same discourse segment. Thus, any violation of transition preferences as predicted from Centering Theory can be regarded as a result of interference with the global discourse structure.

Therefore, the are no clear predictions of Centering Theory on the signalling of grammatical roles, i.e. completeness is satisfied trivially.

Sentence 6 (Compatibility of speaker salience predictions) *There is no prediction of the reconstruction on the choice of grammatical roles which is not also predicted by Centering Theory.*

For the proof, I will differentiate between two aspects of speaker salience, that is, the indication of the salience of forward-looking centers by grammatical roles and the effect of the salience of forward-looking centers

Proof speaker salience indication In the reconstruction, the salience of grammatical roles is the only pragmatic determinant of the assignment of grammatical roles, whereas in Centering Theory, the salience of forward-looking centers is indicated by the assignment of grammatical roles. Hence, both the original model and the reconstruction assume a 1:1-correspondence between grammatical roles and speaker salience/salience of forward-looking centers.

effects of speaker salience In the reconstruction, pronouns are sensitive to previously established, and thus, marked, topic referents, and hence, the appearance of a pronominal form indicates that in the previous discourse, such a topic marking construction appeared. If we assume that it was the speaker's intention to use such a topic marking construction, this specific aspect of speaker salience can be recovered by the investigation of pronominalization in the subsequent discourse.

in Centering Theory, the higher ranked a referent in the current utterance is with respect to its grammatical roles, the more probable it is to be backward-looking center in the following utterance and thus, appear as a pronoun. According to the preference for continue transitions and the cheapness principle (Strube and Hahn 1999), the most likely backward-looking center is the preferred center of the previous utterance, as signalled by the most highly-ranked grammatical role. Thus, Centering entails a preference for the backward-looking center to be realized as the preferred center in the previous clause. (However, it should be noted that this is not a claim made by Centering, but only a predicted preference.)

As topic marking and the notion of the preferred center in centering are comparable conceptions, the retrospective identification of previously established discourse topics, i.e. preferred centers and the underlying speaker salience, from the use of pronouns in the following utterance is compatible with Centering Theory.

Corollary 3 (Adequacy of speaker salience scores) *The reconstruction of the salience of forward-looking centers and its indication by grammatical roles are adequate with respect to Canonical Centering and Left-Right Centering.*

Finally, from Corollaries ??? to ???, we can draw the conclusion that the reconstruction of Centering within the Mental Salience framework is *adequate* as compared to the original formulation of the theories.

Sentence 7 (Adequacy of Centering Reconstruction) *With respect to hearer salience (backward-looking center), the effects of speaker salience (preferred center), and their indication by pronominalization and the assignment of grammatical roles, the reconstructions of Canonical Centering and Left-Right Centering are adequate with respect to their original formulations.*

It should be noted, again, that this understanding of adequacy is identical to Beaver's "equivalence". However, the modeling of speaker salience can be regarded as an extension of the original theories which was not considered in their original formulation. (However, in their application of Centering Theory in the Natural Language Generation context, Kibble and Power (2000) have also exploited Centering transitions for the assignment of grammatical roles.) With respect to this, the reconstruction of Centering within the Mental Salience framework exceeds the predictive power of the original formulation of Centering Theories.

Further, the reconstruction allows for the integration of additional salience factors besides grammatical roles and referring expressions which were originally considered in Centering Theory, and in this sense, its relationship to the original formulation of Centering Theory is comparable to Beaver's reconstruction of Centering Theory in Optimality Theory.

7.4 Extensions of Centering Theory

Above, I have claimed the adequacy of the reconstruction of Centering. However, in order to illustrate that the formulation within the Mental Salience framework also allows for the integration of additional parameters in a very natural way, I will shortly discuss two examples from natural text which are incompatible with Centering predictions and show how to handle these within the Mental Salience framework.

pronominalization can also interact with more global parameters

> Greg$_g$'s eyes flicked up from his instruments panel. He$_g$ saw them, ... "Zeros !" Todman$_t$ said excitedly, and hopefully. And he$_g$ thought Todman$_t$ might be right. (Susanne corpus, N15)

In this example, Christine is the topic of the discourse, both global and local, and I can only assume that the full name is used for stylistic reasons.

However, it is also possible that the backward-looking center is not pronominalized for stylistic reasons as in the following example in which, nevertheless, a non-Cb-referent is pronominalized.

> Christine$_c$ knew what would be waiting for her$_c$ at home and sure enough, within weeks Daniel$_d$ attacked her$_c$ again. She$_c$ managed to call the police and they came and arrested him$_d$. Christine$_c$ obtained an injunction to keep him$_d$ away from the house and filed for divorce. ("Christine's story", Sandra Horley (2004), `http://www.refuge.org.uk/page_l1-2_l2-182_l3-2592_.htm`)

Possibly, this phenomenon is related to

Stylistic reasons: effect of foregrounding on referring expressions ?

The most salient referent in the reconstruction corresponds to the backward-looking center

- The whole theory is exemplified with one example. Could you give another example where you can show that your theory is better than Centering

in einer perfekten welt würde ich jetzt den OT-Teil durch eine Beispielanalyse ersetzen ;)

References

Judith Aissen. Markedness and subject choice in Optimality Theory. In Geraldine Legendre, Jane Grimshaw, and Sten Vikner, editors, *Optimality-Theoretic Syntax*. MIT Press, Cambridge, 2001.

Mira Ariel. Accessibility theory: An overview. In *Sanders et al. (2001)*, pages 29–87. 2001.

Mira Ariel. *Accessing Noun-Phrase Antecedents*. Routledge, London, New York, 1990.

Mira Ariel. Interpreting anaphoric expressions: A cognitive versus a pragmatic approach. *Journal of Linguistics*, 30:3–42, 1994.

David I. Beaver. The optimization of discourse. *Linguistics and Philosophy*, 27(1), 2004.

Gerlof Bouma. Doing Dutch pronouns automatically in Optimality Theory. In *Proceedings of the EACL 2003 Workshop on the Computational Treatment of Anaphora*, Budapest, April 2003.

Susan E. Brennan, Marilyn W. Friedman, and Carl J. Pollard. A Centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics (ACL 1987)*, pages 155–163, Stanford, July 1987.

Adam Buchwald, Oren Schwartz, Amanda Seidl, and Paul Smolensky. Recoverability Optimality Theory: Discourse anaphora in a bidirectional framework. In *Proceedings the 6th Workshop on the Semantics and Pragmatics of Dialogue (EDILOG 2002)*, Edinburgh, September 2002.

Donna K. Byron and Whitney Gegg-Harrison. Evaluating optimality theory for pronoun resolution algorithm specification. In *Proceedings of the Discourse Anaphora and Reference Resolution Colloquium (DAARC 2004)*, pages 27–32, September 2004.

Karl Bühler. *Sprachtheorie. Die Darstellungsfunktion der Sprache*. Gustav Fischer, Stuttgart, 1934. translated 1990 Theory of language. Benjamins:Amsterdam, Philadelphia.

Wallace Chafe. Giveness, contrastiveness, definiteness, subjects, topics, and point of view. In *Li (1976)*, pages 25–55. 1976.

Wallace Chafe. *Discourse, Consiousness, and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing*. University of Chicago Press, Chicago and London, 1994.

C. Robin Clamons, Ann E. Mulkern, and Gerald Sanders. Salience signaling in Oromo. *Journal of Pragmatics*, 19:519–536, 1993.

James Raymond Davis and Julia Hirschberg. Assigning intonational features in synthesized spoken directions. In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics (ACL 1988)*, pages 187–193, Buffalo, June 1988.

Edward Gibson and Neal J. Pearlmutter, editors. *The Processing and Acquisition of Reference*. MIT Press, Cambridge, Mass, to appear.

Rachel Giora. On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics*, 31:919–929, 1999.

Talmy Givón. *Syntax*. John Benjamins, Amsterdam and Philadelphia, 2001. 2 volumes, revised edition of Givón (1984) and Givón (1990).

Talmy Givón. Introduction. In *Givón (1983b)*, pages 5–41. 1983a.

Talmy Givón, editor. *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. John Benjamins, Amsterdam and Philadelphia, 1983b.

Talmy Givón. *Syntax*, volume I. John Benjamins, Amsterdam and Philadelphia, 1984.

Talmy Givón. *Syntax*, volume II. John Benjamins, Amsterdam and Philadelphia, 1990.

Talmy Givón. *Functionalism and Grammar*. John Benjamins, Amsterdam and Philadelphia, 1995.

Joseph H. Greenberg. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg, editor, *Universals of language*, pages 73–113. MIT Press, Cambridge, Mass., 1963.

Barbara J. Grosz and Candace L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12:175–204, 1986.

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

Eva Hajičová, Ivana Kruijff-Korbayová, and Geert-Jan M. Kruijff. Salience in dialogues. In Svetla Cmejrková, Jana Hoffmannová, Olga Müllerová, and Jindra Svetlá, editors, *Dialogue Analysis VI: Proceedings of the 5th International Congress of the International Association of Dialogue Analysis, April 17-20 1996, Prague, Czech Republic*, pages 381–393, Prague, Czech Republic, April 17-20 1998. Max Niemeyer Verlag.

Eva Hajičová and Ivana Kruijff-Korbayová. Topics and centers: A comparison of the salience-based approach and the Centering theory. *Prague Bulletin of Mathematical Linguistics*, 67:25–50, 1997.

Eva Hajičová and Jarka Vrbova. On the role of the hierarchy of activation in the process of natural language understanding. In Jan Horecký, editor, *Proceedings of the Ninth International Conference of Computational Linguistics (COLING 1982), Prague*, pages 107–113, Prague, July 1982. Academia.

Daniel Hardt. Dynamic Centering. In *Proceedings of the Workshop on Reference Resolution and its Applications. Held in Conjunction with ACL 2004*, pages 55–62, Barcelona, 2004.

John A. Hawkins. Syntactic weight versus information structure in word order variation. In Joachim Jacobs, editor, *Informationsstruktur und Grammatik*, pages 196–219. Westdeutscher Verlag, Opladen, 1992.

Helmut Horacek. An algorithm for generating referential descriptions with flexible interfaces. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pages 206–213, Madrid, July 1997.

Elsi Kaiser and John Trueswell. Investigating the interpretation of pronouns and demonstratives in Finnish: Going beyond salience. In *Gibson and Pearlmutter (to appear)*. to appear.

John D. Kelleher and Josef van Genabith. Exploiting visual salience for the generation of referring expressions. In Valerie Barr and Zdravko Markov, editors, *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004)*, Miami, 2004. AAAI Press.

Rodger Kibble. Towards the elimination of centering theory. In Ivana Kruijff-Korbayovà and Claudia Kosny, editors, *Proceedings of the 7th Workshop on the Semantics and Pragmatics of Dialogue (DiaBruck)*, pages 51–58, University of Saarbrücken, September 2003.

Rodger Kibble and Richard Power. An integrated framework for textplanning and pronominalisation. In *Proceedings of the International Conference on Natural Language Generation (INLG)*, 2000.

Rodger Kibble and Richard Power. Optimizing referential coherence in text generation. *Computational Linguistics*, 30(4):401–416, 2004.

Christof Koch and Laurent Itti. Computational modelling of visual attention. *Nature Review Neuroscience*, 2:194–203, 2000.

Emiel Krahmer and Mariët Theune. Efficient contextsensitive generation of referring expressions. In *van Deemter and Kibble (2002)*, pages 223–264. 2002.

Geert-Jan M. Kruijff, Ivana Kruijff-Korbayová, John Bateman, and Elke Teich. Linear order as higher-level decision: Information structure in strategic and tactical generation. In Helmut Horacek, editor, *Proceedings of the 8th European Workshop on Natural Language Generation*, pages 74–83, Toulouse, France, July 5-6 2001.

Willem J.M. Levelt. *Speaking: From Intention to Articulation*. MIT Press, 1989.

Charles N. Li, editor. *Subject and Topic*. Academic Press, New York, 1976.

Ann E. Mulkern. *Cognitive Status, Discourse Salience, and Information Structure: Evidence from Irish and Oromo*. PhD thesis, University of Minnesota, 2003.

Costanza Navaretta. Combining information structure and centering-based models of salience for resolving intersentential pronominal anaphora. In Antonio Branco, Tony McEnery, and Ruslan Mitkov, editors, *Proceedings of the 4th Discourse Anaphora and Anaphora Resolution Colloquium (DAARC 2002)* , pages 135–140, Lisbon, September 18-29 2002.

Helen Fay Nissenbaum. *Emotion and Focus*. CSLI, Stanford, CA, 1985.

Andrew Ortony. Similarity in similes and metaphors. In Andrew Ortony, editor, *Metaphor and Thought*, pages 186–201. Cambridge University Press, Cambridge, 1979.

Thiyagarajasarma Pattabhiraman. *Aspects of Salience in Natural Language Generation*. PhD thesis, Simon Fraser University, August 1992.

Massimo Poesio, Hua Cheng, Renate Henschel, Janet Hitzeman, Rodger Kibble, and Rosemary Stevenson. Specifying the parameters of Centering Theory: A corpus-based evaluation using text from application-oriented domains. In *Proceedings of the 38th Meeting of the Association for Computational Linguistics (ACL 2000)*, Hong Kong, 2000.

Massimo Poesio, Barbara Di Eugenio, Rosemary Stevenson, and Janet Hitzeman. Centering: A parametric theory and its instantiations. *Computational linguistics*, 30(3):309–363, 2004.

Ellen F. Prince. Toward a taxonomy of given-new information. In P. Cole, editor, *Radical Pragmatics*, pages 223–256. Academic Press, New York, 1981.

Regina Pustet. *Diskursprominenz und Rollensemantik – Eine funktionale Typologie von Partizipantensystemen*. Lincom Europa, München, 1997.

Owen Rambow. Pragmatic aspects of scrambling and topicalization in German. In *Workshop on Centering Theory in Naturally-Occurring Discourse*. Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia, PA, 1993.

Ted Sanders, Joost Schilperoord, and Wilbert Spooren, editors. *Text Representation. Linguistic and Psycholinguistic Aspects*. John Benjamins, Amsterdam and Philadelphia, 2001.

Petr Sgall, Eva Hajićová, and Jarmila Panevova. *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*. Reidel, Dordrecht, 1986.

Michael Strube and Udo Hahn. Functional Centering. In *Proceedings of 34th Annual Meeting of the Association for Computational Linguistics (ACL 1996)*, pages 270–277, Santa Cruz, June 1996.

Michael Strube and Udo Hahn. Functional Centering - grounding referential coherence in information structure. *Computational Linguistics*, 25(3):309–344, 1999.

Joel R. Tetreault. Analysis of syntax based pronoun resolution methods. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL 1999)*, Maryland/MD, 1999.

Russel S. Tomlin. Focal attention, voice, and word order. An experimental, cross-linguistic study. In Mickey Noonan and Pamela Downing, editors, *Word Order in Discourse*, pages 517–554. John Benjamins, Amsterdam and Philadelphia, 1995.

Amos Tversky. Features of similarity. *Psychological Review*, 84(4):327–352, 1977.

Kees van Deemter and Rodger Kibble, editors. *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*. CSLI, Stanford, 2002.

Ielka van der Sluis and Emiel Krahmer. Generating referring expressions in a multimodal context: An empirical approach. In Walter Daelemans et al., editor, *Selected Papers from the 11th CLIN Meeting*. Rodopi, Amsterdam and Atlanta, 2001.

Andrea Weber and Karin Müller. Word order variation in German main clauses: A corpus analysis. In *Proceedings of the 5th International Workshop on Linguistically Interpreted Corpora. Held in Conjunction with the 20th International Conference on Computational Linguistics (COLING 2004)*, pages 71–78, Geneva, August 2004.