Abduction in Games for a Flexible Approach to Discourse Planning

Ralf Klabunde, Sebastian Reuße, and Björn Schlünder

Ruhr-Universität Bochum, Department of Linguistics
{ralf.klabunde,sebastian.reusse,bjoern.schluender}@rub.de

Abstract. We propose a new approach to document planning for natural language generation that considers this process as strategic interaction between the generation component and a user model. The core task of the user model is abductive reasoning about the usefulness of rhetorical relations for the document plan with respect to the user's information requirements. Since the different preferences of the generation component and the user model are defined by parametrised utility functions, we achieve a highly flexible approach to the generation of document plans for different users. We apply this approach to the generation of reports on performance data. The questionnaire-based evaluation we accomplished so far corroborates the assumptions made in the model.

1 Introduction

The basic goal of natural language generation (NLG) systems is the transformation of unwieldy information – for example, data streams or ontological knowledge – to linguistic representations that are more accessible to humans. Accessibility, in the broader sense, largely depends on the individual addressee: Different addressees possess different amounts of background knowledge which they may bring to bear upon the act of interpreting a text, and their different interests shape specific perspectives upon a discourse topic. For example, a report on meteorological data which elaborates the significance of reported climatic configurations may be indispensable to the layman, but wholly inadequate to the more meteorologically versed user.

If an NLG system has to cover a wide range of addressees, it has to incorporate a means of selecting and structuring content in a way that takes into account the communicative relevance of the generated text. In this paper, we present a general, user-oriented approach to discourse structuring that allows a flexible adaptation of document structures to individual users. The basic idea is to introduce a representation of a user's interests and prior beliefs, reflected as assumption costs on certain hypotheses, and to anticipate the relevance of different candidate messages during the generation process by means of abductive inference.

In what follows, we will first review some existing approaches to achieving text customized for different users in order to show that language generation should be considered as a joint activity between two independent agents (or players, in

A. Gelbukh (Ed.): CICLing 2012, Part II, LNCS 7182, pp. 13–24, 2012.

game-theoretic terms). Based on this, we present our model of user-customized document generation which makes explicit reference to a user's potential interests and allows user types to be defined dynamically with reference to those interests. We will give an example of how this model is realized in a specific domain, and present the first evaluation result how well the resulting documents have been accepted by readers.

2 User-Oriented Document Planning

The insight that generated texts should be tailored to the specific users in order to enhance the acceptability of the text, is not really new. However, since user models are primarily used for pragmatic decisions during the generation process, it is remarkable that the principles behind formal approaches to pragmatics – describing the joint activities of the interlocutors – did not receive much attention in NLG systems.

For example, Hovy (1988) defines rhetorical relations as planning operators with associated pre- and post-conditions that describe the interests and knowledge states of the agents involved. While this approach results in sophisticated user-oriented documents, there are no means of determining whether one of several alternative document structures might be more relevant than others with regard to a user's expectations. More recently, Mairesse and Walker (2010) describe a generation system which explicitly seeks to match output to a user based on various cognitive decisions, but does so only on the level of linguistic realization.

Dale and Reiter (2000) mention the possibility to address user variability by pre-specifying document schemas for a finite number of user types. Each user type corresponds to one alternative document structure definition. Such an approach might become inconvenient when multiple factors are involved in the mapping, such that the number of types increases exponentially with the number of factors. Furthermore, there is no explicit representation of the reasons why a document schema is defined the way it is and why it includes the given information. Ultimately one has to rely on expert judgement.

In their overview, Zukerman and Litman (2001) describe the various approaches to user modeling in NLG, but this survey demonstrates nicely that generally accepted principles of user modeling do not exist so far. Instead a whole bundle of different approaches have been proposed, mostly linked to pragmatic tasks, but without recourse to formal pragmatics. Reiter et al. (2003) also mention the fact that little is known about the acquisition of user models.

Our approach centers around the well-established idea that language production should be considered as one component of a joint activity of two agents. Speakers adapt their utterances to the linguistic and cognitive abilities of the addressee and vice versa. The speaker's adaptation is based on stereotypes and associated defaults, the established common ground, and individual features of the respective addressee (See, e.g., Brennan et al. (2010)).

If one accepts this view of communication as highly flexible, joint activity, there are no clearly defined classes of users, but rather there is a set of different needs and expectations an individual user might have. Prototypical users should be considered as mere points in a multidimensional space of user types, where points are located by assigning different priorities to a user's possible needs. Text plans are then to be derived with respect to this coordinate system and can be generated dynamically for every possible point in the user type space.

3 The Formal Model

The formal model for document planning we propose in this paper takes into account the aforementioned view of generation as a collaborative process.

3.1 Data Source

Our model relies on some data source d from which we derive an initial pool of individual, atomic messages. The messages' content depends on how we devise the parsing mechanism mapping the data source to the document planners initial message repository.

3.2 Rhetorical Relations

Our use of rhetorical relations as structuring means for document plans is more or less identical to the standard uses in NLG systems, except that we use a logical notation. We define how we may induce possible document structures over the unstructured pool of information derived from d, by giving logical definitions of how rhetorical relations may coherently be applied to complex or atomic messages. Messages are atomic if they are immediately derived from some data source, or complex if the message is made up of constituent messages, themselves either complex or atomic. The application of some type of rhetorical relation to its constituent messages results in a new, complex message of the same type and can be viewed as a partial document plan.

The definitions of the rhetorical relations are given with reference to the triggering messages' content or type, thus defining preconditions which, once satisfied, trigger the establishment of a relation. For example, a relation between two messages of specific type, with a constraint on some property of the second message, could be given as:

$$\begin{aligned} & prerequisiteType_1(X) \wedge prerequisiteType_2(Y) \\ \wedge\ & prerequisiteProperty(Y) \rightarrow compositeMessage(X, Y) \end{aligned}$$

We explicitly allow for a whole number of competing relations being triggered in some state during the generation of document plans, because we aim at singling out that relation that is most relevant to the specific user, acording to the current state of the user model.

3.3 Relevance-Relations in a User Model

To represent how the different dimensions of potential user-interests constrain document structures, we define, where appropriate, whether some interest is indicative resp. counter-indicative of a specific message type. In other words, we provide a knowledge base consisting of theorems of the form '$hypothesis \rightarrow [\neg]messageType$', where '$\neg$' marks counter-indicative rules. '$hypothesis$' models a possible dimension of interest of a user, and '$messageType$' corresponds to the type of a node in a document structure that is made relevant in the face of that dimension of interest. For example, we might use a rule '$preferDetail \rightarrow elaboration$', which introduces a dimension called $preferDetail$ into the user-type space, such that the generation of composite messages created by applying $elaboration$-relations will be encouraged for users whose model ranks prominently in the $preferDetail$-dimension. Since there could be multiple hypotheses indicating a (potentially composite) message's relevance, and since not all hypotheses are appropriate for all user types, assumption costs are introduced.

The user model assigns user coordinates in the space created by the different dimensions introduced by the relevance-relations. We interpret these coordinates as the aforementioned assumption costs of hypotheses about the user, where each dimension corresponds to one hypothesis. For example, a relatively high numerical value for $preferDetail$ would indicate some reluctance to assume that the user prefers detailed information and as such would limit the effect of $preferDetail$, encouraging the generation of certain messages for this user.

Be T a knowledge base, H the set of hypotheses given in T, ψ a document plan, and *cost* a function that assigns assumption costs to hypotheses. We call the process of selecting the most felicitous hypothesis *naïve abduction.* If $\langle T, \psi, costs \rangle$ is a cost-based abduction problem, we try to find $h^* \in H$ such that:

$$\forall h \in H : match(h, \psi) - cost(h) \leq match(h^*, \psi) - cost(h^*)$$

The function $match(h, \psi)$ measures how coherently some hypothetical user-interest h integrates with a partial or complete document plan ψ. Let $M_h = \{m | h \rightarrow m\}$ and $M_h^- = \{m | h \rightarrow \neg m\}$. We define:

$$\begin{aligned} match(h, \psi) := \ & \gamma_1 \times |M_h \in \psi| \\ & -\gamma_2 \times |M_h^- \in \psi| \end{aligned}$$

I.e., we count the number of messages related to some hypothetical interest according to the relevance definitions, increasing an initial score of zero for each expected message, and decreasing it for every counter-indicated message. This approach might be considered as a lean version of weighted abduction (Hobbs et al. (1993),Ovchinnikova et al. (2011)): To determine the compatibility between a document plan and a hypothesis, the system first assumes that the hypothesis h is true and subsequently tests if the document plan under consideration is relevant given the relevance-definitions, by counting messages related to the interests modelled by h. The parameters $\gamma_{1,2}$ allow us to assign different weights to expected vs. deprecated messages.

3.4 A Game-Theoretic Approach

Our algorithm for user-tailored document planning uses a normal form game (cf. Leyton-Brown and Shoham (2008); Parikh (2010)) $\langle\{S,L\},\{A_S,A_L\},\{U_S,U_L\}\rangle$ which is iteratively played, with each iteration effectively establishing a single rhetorical relation over a subset of messages taken from the message pool.

The game is defined for two agents, the generation-system S and a user model L. During a single iteration, the system considers different alternative actions A_S, where each action corresponds to one rhetorical relation which might possibly be applied to a set of component messages taken from the repository. The set of possible rhetorical relations to be established is constrained by the prerequisites on component messages defined for the different relations. A_L consists of possible hypotheses an addressee might assume, in order to explain why some composite message generated by the system might be relevant to him. Finally, the utility-functions U_S and U_L determine the felicity of the resulting combination of generated composite message and underlying assumptions.

The definitions of the utility functions draw on three basic notions which model the 'cognitive burden' in establishing the resulting document structure:

1. The aforementioned function $match(h,\psi)$.
2. A metric $complexity(\psi)$ that indicates the structural complexity of a (partial or complete) document plan ψ by simply counting the number of message nodes contained in it.
3. The function $cost(h)$ gives the assumption cost of some hypothetical interest h according to the user model. The higher the cost, the less likely we are to assume that the interest represented by h does apply to the user.

Given these notions, we define the utility functions for S and L as follows:

$$U_L(m,h) := \alpha_1 \times match(h,\{m\}) - \alpha_2 \times cost(h)$$

$$U_S(m,h) := \frac{\beta_1 \times match(h,\psi \cup \{m\})}{\beta_2 \times complexity(m)}$$

Both agents, S and L, prefer documents which are coherent with regard to a user's potential interests. Furthermore, the generation system seeks to generate structurally plain documents. Both formulas are parameterized in order to control the impact of each contributing factor $(\alpha_{1,2},\beta_{1,2})$.

Figure 1 shows a partial game-tree with possible payoffs. Here, the generation system considers two possible relations it might establish. According to the relevance-relation definition, $preferDetails$ would indicate the presence of an *elaboration*-message, and as such accounts for a utility of one in the agents' payoffs, while $otherExplanation$, given that there is no relation to *elaboration*-messages in the assumed relevance-model, yields a utility of zero. To determine L's utility, we also discount the assumption cost of $preferDetails$ from his payoffs, so that the total payoff varies with the specific type of addressee we assume.

The game-theoretic mechanism as defined above is reiterated according to the algorithm shown in Table 1. We iteratively apply the most felicitous type

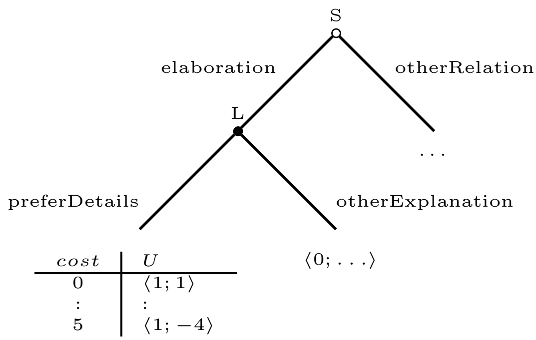

Fig. 1. An example of a single message-generation-game in extensive form. Payoffs are shown as function of *cost*.

Table 1. A game-theoretic document-generation algorithm. In each iteration, an inventory of possible new messages is generated by applying rhetorical relations to subsets of the message pool. Out of the inventory of possible new messages, the most felicitous one is selected for subsequent generation by adding it back into the message pool while removing all of its constituents from it.

$POOL \leftarrow$ all messages derived from d
$A_L \leftarrow$ {all interest types used in the relation-relevance definition}
while ($A_S \leftarrow$ {rhetorical relations which may link elements in $POOL$}) $\neq \emptyset$:
    $\langle a_S, a_L \rangle \leftarrow$ pareto-optimal pure strategy equilibirum of
        $\langle \{S, L\}, \{A_S, A_L\}, \{U_S, U_L\} \rangle$
    $POOL \leftarrow (POOL \setminus \{\text{constituents of } a_S\}) \cup a_S$
return $POOL$

of rhetorical relation until no more relations are applicable. As a result, the message pool contains one or several tree structures, depending on whether or not at some point a conjoining relation existed and was optimal in terms of both agent's goals.

3.5 An Example Game

Table 2 provides a schematic example of how our abstract model of document generation can be instantiated into a fully-fledged generation system. It lists the four input factors determining the construction of a document plan. The **message pool** contains a set of atomic messages which the system will seek to combine by means of rhetorical relations. In this case, we assume that the data source provides some basic information (*someMsg*), accompanied by related messages (*bgInfo*, *advice*) which either relate additional information regarding *someMsg*, or dispense relevant advice. The **text grammar** then defines

Table 2. A schematic instantiation of the document generation model

Message Pool: $someMsg$, $bgInfo$, $advice$	**Text Grammar:** $someMsg \wedge bgInfo \rightarrow elaboration$ $someMsg \wedge advice \rightarrow interpretation$
Relevance Theory: $isNovice \rightarrow interpretation$ $preferDetail \rightarrow elaboration$	**User Model:** $cost(isNovice) := 0$ $cost(preferDetail) := 5$

Table 3. Complete first iteration of the schematic example system

S generates $elaboration$

(a) L assumes $isNovice$ – S's utility: $0/2 = 0$ – L's utility: $0 - 0 = 0$	(b) L assumes $preferDetail$ – S's utility: $1/2 = 0.5$ – L's utility: $1 - 5 = -4$

S generates $interpretation$

(a) L assumes $isNovice$ – S's utility: $1/2 = 0.5$ – L's utility: $1 - 0 = 1$	(b) L assumes $preferDetail$ – S's utility: $0/2 = 0$ – L's utility: $0 - 5 = -5$

how these messages may be combined in a coherent way to form new complex messages. As assumed above, combining $someMsg$ with background information will form an *elaboration*, while providing advice alternatively creates an *interpretation* message.

Once the means of applying rhetorical relations are defined, the **relevance theory** indicates when each of the alternative ways of forming complex messages might be relevant, by listing hypothetical aspects of a user's interests and the message types related to those interests. Here, we assume that a novice in our schematic domain will prefer *interpretation* messages, while a preference for detail is expected to give rise to *elaboration*. Finally, the **user model** specifies the actual interests of the user currently served, by listing assumption costs for each of the predicates used in modeling user interests. In this example we assume the user is likely a novice, since assuming $isNovice$ incurs no cost, while also assuming that she has no particular interest in detail, since $preferDetail$ comes with a relatively high cost.

Once these factors are set, the system begins constructing a document plan by iterating through all relations applicable in a single turn, and determining the payoffs of both agents relative to the message in consideration and each hypothetical dimension of the listener's interests. Table 3 shows the complete first turn of the generation system set out in Table 2 and each agent's payoffs

Table 4. The outcome achieved after the first iteration

	elaboration	*interpretation*
isNovice	$\langle 0, 0 \rangle$	$\langle 0.5, 0 \rangle$ √
preferDetail	$\langle 0.5, -4 \rangle$	$\langle 0, -5 \rangle$

according to the relevant utility function. Table 4 shows the turn in normal form and marks the game's equilibrium. In this case, both agent's goals (coherence, relevance and concision) are best served by generating an *interpretation* message from the message pool. At the end of the turn, the *interpretation* will accordingly be added into the document pool, while its constituent messages are removed from it.

4 Implementation and Application

We applied this model of document planning to the generation of runner's performance data. These data, generated by a heart-rate monitor device while jogging, are transformed into different texts according to the user's needs. Relevant dimensions, defining a user's background, are the frequency of exercise, degree of experience and prior training factors such as the degree of strain caused by exercising. Our main focus regarding user variability concerned the abundance of numerical data and the presence of explanatory content, but we also incorporated the different goals of amateur vs. experienced exercisers by generating appropriate advisory messages as to how these goals might be realized in the future.

We implemented the following rhetorical relations, following their standard descriptions in Rhetorical Structure Theory: *Preparation, Conjunction, Elaboration, Background, Sequence* and *Contrast.*

Although our model is capable of generating documents for all combinations of possible users according to the user-space spanned by the relevant dimensions, we restricted ourselves to three prototypical users, i.e. amateurs, advanced and semi-professional runners represented by salient points in the space of possible users. Figure 2 shows a sample document plan generated by the system.

The system is realized as a Python module. The core consists of a parser for the data files generated by the heart-rate monitors, and an abductive reasoning module used to trigger rhetorical relations. In order to solve the game played by the generation system and the user model – and thus determine the most felicitous message to be generated – we employed the freely available Gambit tool (McKelvey et al. (2010)).

The linguistic realization of the generated document plans was performed by a schema-based approach for message types, consisting of canned text interspersed with schematic references to a message content. For example, a message of type 'RunMessage', relating general information about the user's exercise, is as follows:

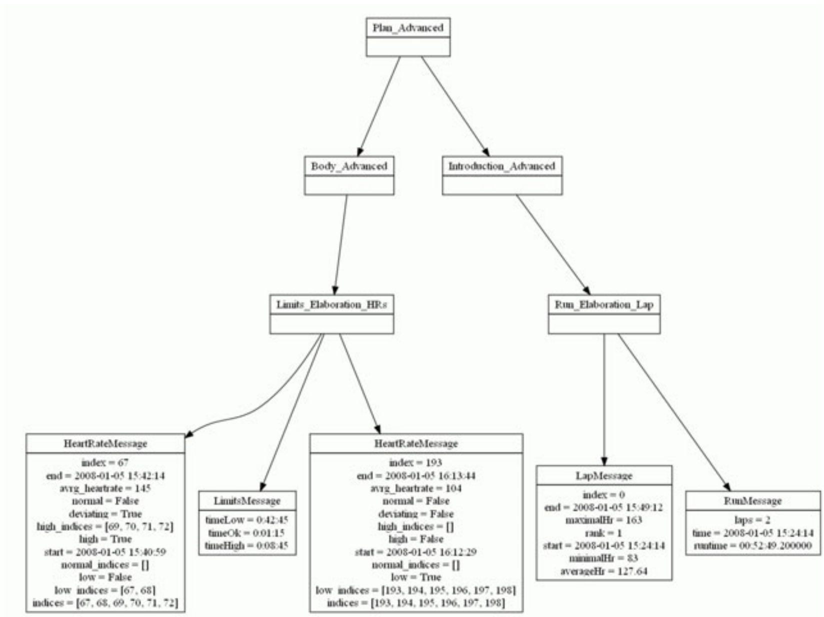

Fig. 2. An example document within the exercise data domain. Box headers indicate message types. The contents list all data contained within a message. Edges visualize component relationships between messages.

```
RunMessage: You have been running a total of <laps> laps which has
taken you <runtime>.
```

With less specific, complex message types, canned text can be used to introduce signal words into the realization. Interestingly this rather simple approach, when applied to the generated discourse structures, results in texts with apparently sufficient complexity for the users.

```
Contrast: <component[1]>. However, <component[2]>.
```

The first paragraph of a generated German text for occasional runners and its English translation are given in Table 5.

Table 5. Beginning of a generated text for occasional runners

German original:	English translation:
Am 13.10. waren Sie Joggen. Für Sie als Gelegenheitsläufer ist ein ausgewogener Lauf mit relativ niedriger Herzfrequenz wichtig. Während des Laufs lag Ihre Herzfrequenz 0:00:00 Minuten im für Sie optimalen Bereich. 0:07:45 Minuten lag sie darunter, 0:00:15 Minuten lag Sie oberhalb der optimalen Frequenz...	*On october 13th you went on a run. You as an occasional runner need to keep your run balanced and your heart rate relatively low and steady. During your run your heart rate was in the optimal interval for 0:00:00 Minutes. 0:07:45 it was lower, 0:00:15 it was higher than your optimal frequency...*

5 Evaluation Results

Our evaluation concerned the acceptability of the texts for different users. For this, 41 questionnaires were completed by students of the department of linguistics.

5.1 Method of Evaluation

The questionnaires at first presented three short texts generated by the system. After reading, the test persons were asked to assess the stereotype of runner they belong to, and assess themselves in a set of attributes which correlate with attributes used in the generating system. The ratings were to be assessed on a scale from 1 (False) to 5 (True). Propositions to be assessed were:

- I train regularly.
- I find training easy.
- Sometimes I train to intensively.
- I am in good physical condition.

After their self-assessment the test persons were asked to rate each of the texts previously read. For every text in question a set of three propositions was given, each to be rated on the same scale as before:

- The data presented in the text is explained sufficiently for my concerns.
- The amount of data and numbers are after my fancy.
- The information given is useful for my further training.

The test persons then were asked to choose one of the given texts as the one they would prefer.

5.2 First Results

From a subjective point of view, the overall ratings seem encouraging, since in every instance there is at least one text that is rated as acceptable, and the

Table 6. Average score of the document types in propositions checked over all instances

Document type	**Explanation sufficient?**	**Amount of Data okay?**	**Information helpful?**
Beginner	3.7	3.7	3.5
Advanced	3.3	3.4	3.0
Professional	3.7	3.0	3.0

Table 7. Chosen text per runner type over all instances

Runner type	**No. of instances**	**Chose Beginner**	**Chose Advanced**	**Chose Pro**
Beginner	29	15	9	5
Advanced	9	5	2	2
Professional	3	1	1	1

average rating of all questions regarding all three texts over all 41 instances is at least 3.0. Therefore we may conclude that the texts generated using the game-theoretic planning algorithm seem to be of sufficient quality.

The results, however, must be interpreted with caution, since only a small number of advanced runners and professionals were available. Since the test persons are only able to evaluate the resulting text and not the underlying abstract discourse structure, the results give us just a tentative hint that the resulting discourse structures are really tailored to different listener types.

Anyhow, the data gathered in the survey clearly indicate that the texts generated by the system are of sufficient quality and are accepted by most test persons. Therefore, the assumptions represented in the knowledge base seem to be mostly correct in principle. However, a more sophisticated evaluation is certainly needed as future work.

6 Summary and Outlook

We presented an account of document planning based on the rigorous definition of a user's needs and expectations and their relation to a document's content and structure. As matters stand, it will be necessary to gather further experimental evidence that takes into account the continuous representation of users in our model. So far, we worked with three prototypical user types only. However, we believe that the evaluation of our system already demonstrates a methodological gain, compared with some established approaches to user modeling in NLG.

References

Brennan, S.E., Galati, A., Kuhlen, A.K.: Two Minds, One Dialog: Coordinating Speaking and Understanding. In: Ross, B.H. (ed.) The Psychology of Learning and Motivation, vol. 53, pp. 301–344. Academic Press, Burlington (2010)

Reiter, E., Dale, R.: Building Natural Language Generation Systems. Cambridge University Press, Cambridge (2000)

Hobbs, J.R., Stickel, M., Appelt, D., Martin, P.: Interpretation as Abduction. Artificial Intelligence 63, 69–142 (1993)

Hovy, E.: Generating Natural Language under Pragmatic Constraints. Lawrence Erlbaum, Hillsdale (1988)

Leyton-Brown, K., Shoham, Y.: Essentials of Game Theory. Morgan & Claypool Publishers (2008)

McKelvey, R.D., McLennan, A.M., Turocy, T.L.: Gambit: Software Tools for Game Theory, Version 0.2010.09.01 (2010), `http://www.gambit-project.org`

Mairesse, F., Walker, M.A.: Towards Personality-Based User Adaptation: Psychologically Informed Stylistic Language Generation. User Modeling and User-Adapted Interaction 20(3), 227–278 (2010)

Ovchinnikova, E., Montazeri, N., Alexandrov, T., Hobbs, J., McCord, M.C., Mulkar-Mehta, R.: Abductive Reasoning with a Large Knowledge Base for Discourse Processing. In: Proceedings of IWCS 2011, Oxford, pp. 225–234 (2011)

Parikh, P.: Language and Equilibrium. The MIT Press, Cambridge (2010)

Reiter, E., Sripada, S., Williams, S.: Acquiring and Using Limited User Models in NLG. In: Proceedings of ENLGW, Budapest, pp. 87–94 (2003)

Zukerman, I., Litman, D.: Natural Language Processing and User Modeling: Synergies and Limitations. User Modeling and User-Adapted Interaction 11, 129–158 (2001)